รายงานวิจัยฉบับสมบูรณ์

โครงการการพัฒนาต่อยอดทฤษฎีเซตด้วยปริพันธ์รัฟเชิงฐานความจุและขอบเขตรัฟเนต

โดย ผศ.ดร.พรรณทิพย์ ภัทรอินทากร

15 พฤษภาคม 2553

สัญญาเลขที่ MRG5180071

รายงานวิจัยฉบับสมบูรณ์

โครงการการพัฒนาต่อยอดทฤษฎีเซตด้วยปริพันธ์รัฟเชิงฐานความจุและขอบเขตรัฟเนต

ผศ.ดร.พรรณทิพย์ ภัทรอินทากร สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา และสำนักงานกองทุนสนับสนุนการวิจัย

บทคัดย่อ

รหัสโครงการ **MRG5180071**

ชื่อโครงการ **การพัฒนาต่อยอดทฤษฎีเซตด้วยปริพันธ์รัฟเชิงฐานความจุและขอบเขตรัฟเนต**

ชื่อนักวิจัย **ผศ.ดร.พรรณทิพย์ ภัทรอินทากร**
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

E-mail Address : kppuntip@kmitl.ac.th

ระยะเวลาโครงการ **2 ปี**

บทคัดย่อ

ทฤษฎีรัฟเซตเป็นคณิตศาสตร์แขนงใหม่ที่กำลังได้รับความสนใจ เนื่องจากสามารถประหยัดทรัพยากรในการคำนวณ ใช้เวลาน้อยกว่า รับขนาดในการคำนวณได้ใหญ่และวิเคราะห์ข้อมูลที่มีความไม่แน่นอนได้ เป็นผลให้ได้รับการยอมรับในวงการวิจัยและถูกนำไปประยุกต์ใช้งานอย่างแพร่หลาย ผู้วิจัยจึงบูรณาการความรู้จากคณิตศาสตร์แขนงต่าง ๆ เพื่อต่อยอดรัฟเซตดังนี้ 1. พิสูจน์หาขอบเขตรัฟเนตบนและล่างสำหรับเซตวิภัชนัยและสมบัติทางคณิตศาสตร์ที่เกี่ยวข้อง 2. คิดค้นปริพันธ์ใหม่ที่ชื่อ Capacity-based definite rough integral (DRI) พร้อมเสนอสมบัติทางคณิตศาสตร์ต่าง ๆ และทดลองประยุกต์ใช้ในปัญหาการคัดเลือกปัจจัยที่มีผลต่อการสร้างกฎสำหรับพยากรณ์ พบว่า DRI สามารถคัดเลือกพร้อมบอกความสัมพันธ์ที่สำคัญได้อย่างถูกต้อง และ 3. คิดค้นสัมประสิทธิ์และดีกรีของกราฟสายงานและกราฟสายงานผกผันบนพื้นฐานของรัฟเซตขึ้นใหม่ พร้อมทำการทดลองวิเคราะห์ข้อมูลและแปลผลลัพธ์ของกราฟสายงาน ท้ายที่สุดยังได้เสนอเอนโทรปีของกราฟสายงานและเอนโทรปีที่เกี่ยวข้อง เพื่อใช้บอกคุณภาพของตัวแบบและความแม่นยำของแต่ละปัจจัยในกราฟสายงานได้ และยังได้คิดค้นการแปลงกราฟสายงานเป็นต้นไม้การตัดสินใจขึ้นใหม่อย่างมีประสิทธิภาพ

คำหลัก

ทฤษฎีรัฟเซต ทฤษฎีเซตวิภัชนัย ขอบเขตรัฟเนต ปริพันธ์รัฟเชิงฐานความจุ กราฟสายงาน เอนโทรปี ต้นไม้การตัดสินใจ

Abstract

Project Code : **MRG5180071**

Project Title : **Two Extensions of Classical Sets: Capacity-Based Rough Integrals and Roughness Bounds**

Investigator : **Asst. Prof. Dr. Puntip Pattaraintakorn**
King Mongkut's Institute of Technology Ladkrabang

E-mail Address : **kppuntip@kmitl.ac.th**

Project Period : **2 years**

Abstract:

Rough set theory, a modern mathematical tool, has recently attracted researcher attention. Rough set theory can be viewed as a soft computing technique. Thus, less resources, computational costs and time are required, when compared to traditional data analysis techniques. It is also tolerant to uncertainty and large scale data. For these reasons, we propose to build new bridges between rough set theory and various branches of other disciplines as follows. (i) We establish several important properties and roughness bounds for fuzzy set operations. (ii) We invent a capacity-based definite rough integral (DRI) and prove it satisfies many natural properties. By experiment, the integral provides a means of measuring the relevance of functions representing features useful in the classification of sets of objects. (iii) Based on flow graph contexts, we define several fundamental measurements for rough sets. To illustrate, we apply these measures to analyze data successfully with supportive interpretations. Furthermore, we propose entropy measures of flow graphs, used to identify predictablilty and quality of a flow graph. More practically, we create an efficient new construction for decision trees from a flow graph, using our new information theoretic measure.

Keywords :

Rough set theory, Fuzzy set theory, Roughness measure, Capacity-Based Rough Integrals, Flow graphs, Entropy, Decision trees

2. Executive summary

ในช่วงระยะเวลา 2 ปีของโครงการวิจัย ผู้วิจัยได้รวบรวมและศึกษาเอกสาร วารสารวิชาการ หนังสือ และทฤษฎีที่เกี่ยวข้องของ ทฤษฎีรัฟเซต (rough set theory) โดย Pawlak และทฤษฎีเซตวิภัชนัย (fuzzy set theory) โดย Zadeh พบว่าทฤษฎีทางคณิตศาสตร์นี้ต่างก็ถูกไปประยุกต์ใช้ในหลากหลายสายงาน เช่น Bioengineer, Medical data analysis, Computer graphic, Robotics, Statistics และอื่น ๆ โดยมีข้อแตกต่างไปจากทฤษฎีเซตที่ถูกใช้มาโดยตลอดกว่าหลายร้อยปีตรงที่ มุ่งเน้นวิเคราะห์เซตที่มีลักษณะไม่แน่นอนหรือมีบางสมาชิกที่ไม่สามารถระบุสภาพสมาชิกได้ (uncertainty and vagueness) และที่สำคัญยังสามารถประยุกต์ใช้กับวิทยาการคอมพิวเตอร์สมัยใหม่แบบ soft computing จึงประหยัดทรัพยากรในการคำนวณ ใช้เวลาน้อยกว่า รับขนาดเซตในการคำนวณได้ใหญ่กว่าและวิเคราะห์ข้อมูลที่มีความไม่แน่นอนได้ เป็นผลให้ได้รับการยอมรับในวงการวิจัยและประยุกต์ใช้งานอย่างแพร่หลาย

รัฟเซตได้ถูกคิดค้นขึ้นมาประมาณ 30 ปี การวิจัยเกี่ยวกับทฤษฎีรัฟเซตจึงจำกัดอยู่ในวงแคบ ๆ (ในประเทศโปแลนด์ต้นกำเนิดเป็นส่วนใหญ่) ผู้วิจัยจึงได้ค้นคว้าเพิ่มเติม พบว่าความสัมพันธ์ระหว่าง roughness measure ของ fuzzy sets ยังมิได้มีการค้นพบหรือพิสูจน์เกี่ยวกับขอบเขต ผู้วิจัยจึงได้สร้างสมมติฐานว่าสามารถใช้ certain increment operator และ uncertain decrement operator พิสูจน์สมบัติทางคณิตศาสตร์ใหม่ที่สำคัญของ roughness measure ได้ จากผลการวิจัย ผู้วิจัยได้พิสูจน์หาขอบเขตบนและล่างของ roughness measure of fuzzy sets ได้ ขอบเขตดังกล่าวมีประโยชน์สำหรับการดำเนินการทางคณิตศาสตร์ของ fuzzy sets โดยเฉพาะอย่างยิ่งสำหรับเซตที่มีขนาดใหญ่ เช่น microarray, graphic หรือข้อมูลทางด้านการแพทย์ในฐานข้อมูลต่าง ๆ

นอกจากประเภทของข้อมูลที่ต้องวิเคราะห์ทั้งเซตข้อมูลหรือรูปภาพทั้งรูปแล้ว ในงานวิจัยบางประเภทอาจต้องการเพียงบางส่วนของข้อมูลหรือบางช่วงของข้อมูล ซึ่งรัฟเซตสามารถตอบโจทย์นี้ได้เช่นกัน ผู้วิจัยจึงได้ศึกษาเกี่ยวกับ Lebesgue integral โดย Choquet และ discrete Choquet integral ชื่อ Rough integral โดย Pawlak และ Skowron ซึ่งเป็นปริพันธ์ที่สร้างจาก indiscernibility relation ของรัฟเซต เพื่อคิดค้นปริพันธ์แบบใหม่ที่เรียกว่า Capacity-based definite rough integral (DRI) และพิสูจน์สมบัติทางคณิตศาสตร์ที่เกี่ยวข้องเบื้องต้น ทั้งนี้ผู้วิจัยยังได้ต่อยอดด้วยการนำปริพันธ์ที่คิดค้นขึ้นไปทดลองประยุกต์ใช้งานจริง ในการคัดเลือกปัจจัยที่มีผลต่อการสร้างตัวแบบหรือกฎเพื่อใช้พยากรณ์ปรากฏการณ์ต่าง ๆ และพลอตกราฟแสดงผลลัพธ์ พบว่า DRI ที่คิดค้นขึ้นสามารถคัดเลือกพร้อมบอกความสัมพันธ์ได้อย่างถูกต้อง โดยสามารถเลือกฟังก์ชันหรือปัจจัยที่ต้องการเปรียบเทียบ เลือก class ที่ต้องการและทำการทดลองเปรียบเทียบที่ช่วงขอบเขตต่าง ๆ ได้อย่างมีประสิทธิภาพและรวดเร็ว

12 ปีหลังจาก Pawlak ได้คิดค้นรัฟเซตขึ้นมา Pawlak ได้นำเสนออีกโครงสร้างหนึ่งคือกราฟสายงาน (flow graph) ที่มีความสัมพันธ์กับรัฟเซตระหว่างค่า lower approximation กับ upper approximation และ ค่า certainty กับ coverage ในกราฟสายงาน แต่งานวิจัยและการ

สำหรับข้อด้อยที่ไม่สามารถบอกคุณภาพของตัวแบบที่สร้างขึ้นจากกราฟสายงานได้ ผู้วิจัยได้คิดค้นการนำเอนโทนปีมาวิเคราะห์กราฟสายงาน ทำให้สามารถคำนวณเอนโทรปีของกราฟ เอนโทรปีร่วม (joint entropy) และ เอนโทรปีมีเงื่อนไข (conditional entropy) ที่ใช้บอกคุณภาพของตัวแบบและความแม่นยำของแต่ละปัจจัยในกราฟได้ นอกจากนี้ผู้วิจัยได้เสนอใช้ information gain ที่คิดค้นขึ้นใหม่แปลงกราฟสายงานเป็นต้นไม้การตัดสินใจ (decision trees) ที่เป็นตัวแบบอีกชนิดหนึ่งที่แปลผลได้ง่ายกว่า และเนื่องจากเป็นต้นไม้จึงสามารถพัฒนาต่อยอดการแวะให้มีประสิทธิภาพขึ้นได้ต่อไป ซึ่งระเบียบวิธีแตกต่างจากงานวิจัยเบื้องต้นที่ไม่สนใจลำดับการสร้างต้นไม้ว่าควรใช้ปมใดก่อนและหลัง ทำให้ต้นไม้ที่ได้จาก information gain มีประสิทธิภาพมากขึ้นในการประยุกต์ใช้ในงานด้าน decision analysis ดังผลที่ได้รายงานไว้

3. วัตถุประสงค์

งานวิจัยนี้ใช้รัฟเซตบูรณาการกับสาขาอื่น ๆ เพื่อค้นพบองค์ความรู้ใหม่ที่สามารถนำมาต่อยอดประยุกต์ใช้กับการแก้ปัญหาจริงได้ โดยมีวัตถุประสงค์หลักดังต่อไปนี้

พิสูจน์สมบัติทางคณิตศาสตร์ใหม่เกี่ยวกับ roughness measure รวมถึงพิสูจน์หาขอบเขตบนและล่างสำหรับการดำเนินการทางคณิตศาสตร์ของ roughness measure of fuzzy sets ที่เหมาะสำหรับข้อมูลที่มีขนาดใหญ่และไม่สามารถระบุการเป็นสมาชิกได้ชัดเจน เช่น microarray รูปภาพ หรือข้อมูลทางด้านการแพทย์ในฐานข้อมูลต่าง ๆ

คิดค้นและพัฒนาการหาปริพันธ์แบบใหม่ที่เรียกว่า Capacity-based definite rough integral (DRI) และพิสูจน์สมบัติทางคณิตศาสตร์ที่เกี่ยวข้อง รวมถึงนำปริพันธ์ที่คิดค้นขึ้นไปทดลองประยุกต์ใช้งานจริง ในการคัดเลือกปัจจัยที่มีผลต่อการสร้างตัวแบบหรือกฎเพื่อใช้พยากรณ์ปรากฏการณ์ต่าง ๆ และสามารถพลอตกราฟแสดงผลลัพธ์ได้ โดยสามารถเลือกฟังก์ชันหรือปัจจัยที่ต้องการเปรียบเทียบ เลือก class ที่ต้องการ และทำการทดลองเปรียบเทียบที่ช่วงขอบเขตต่าง ๆ ได้อย่างมีประสิทธิภาพและรวดเร็ว

คิดค้น categories of vagueness, accuracy of approximation, roughness approximation และ dependency degree สำหรับกราฟสายงานบนพื้นฐานของรัฟเซตขึ้น และหาสูตรคำนวณสำหรับกราฟสายงานผกผัน พร้อมทำการทดลองวิเคราะห์ข้อมูลและแปลผลลัพธ์ของกราฟสายงานได้

เพื่อเสนอเอนโทรปีของกราฟ เอนโทรปีร่วม และเอนโทรปีมีเงื่อนไข ที่ใช้บอกคุณภาพของตัวแบบและความแม่นยำของแต่ละปัจจัยในกราฟสายงาน และใช้ information gain ที่คิดค้นขึ้นใหม่แปลงกราฟสายงานเป็นต้นไม้การตัดสินใจซึ่งแปลผลได้ง่ายกว่าและ มีประสิทธิภาพมากขึ้นสำหรับงานด้าน decision analysis

สร้างความร่วมมือและการแลกเปลี่ยนความรู้ ในการวิจัยร่วมกับนักวิจัยระดับชาติ นั่นคือนักวิจัยที่ปรึกษา รศ. ดร. กัลยา นฤดมกุล และระดับนานาชาตินั่นคือ นั่นคือ Prof. Dr. James F. Peters และ Prof. Dr. Sheela Ramanna จากประเทศแคนาดาที่จะนำไปสู่ความเป็นเลิศทางวิชาการ

เผยแพร่และแลกเปลี่ยนผลงานวิจัยในที่ประชุมวิชาการระดับนานาชาติ รวมถึงตีพิมพ์ผลงานวิจัยในวารสารวิชาการระดับนานาชาติ

4. Research Methodology, Results and Discussion

We discuss roughness measure of fuzzy sets in Section 4.1, capacity-based rough definite integral in Section 4.2 and rough set flow graphs in Section 4.3, respectively.

4.1 Roughness Measures of Fuzzy Sets

Introduction

Rough set theory (RST or RS) was invented by Pawlak [1,2] and fuzzy set theory (FST or FS) by Zadeh [3,4]. Both are extensions of classical set theory. In practice, classical set theory permits the membership of elements in relation to a set with precise condition: an element either belongs or does not belong to the given set. Thus, this concept requires a sharp boundary. In fact, real situations involve complex problems which cannot be represented and solved using simple models. Several traditional mathematical tools we found unsatisfactory for this purpose [5]. Contrarily, rough sets and fuzzy sets approximate the given sets and consider non-sharp boundaries. Rough sets approximation is carried out in terms of two sets that are lower and upper approximations [1,2]. Alternatively, fuzzy sets approximation is carried out by a predefined membership funcion [4–6].

Rough sets approaches have recently attracted researcher attention [8–13,16,51]. The given set (data set or information system), which is the target of the study, typically contains vagueness, data uncertainty, imprecision and incompleteness that all require tuning and adjustments. In order to accommodate such difficulties, use of approximation is required and thus rough set theory is expedient. According to Pawlak [2], the power of rough sets *"... is that it does not need any preliminary or additional information about the data, such as probability distributions in statistics, basic probability assignment in the Dempster-Shafer theory, or grade of membership or the value of possibility in fuzzy set theory"*. The rough sets approach can be viewed as a soft computing technique rather than hard computing technique in general mathematics [19]. Research in soft computing demonstrated successes because it worked synergistically with other methods to provide flexible analytical tools in real situations. Medsker [19] stated that soft computing differs from traditional computing in that it is tolerant of imprecision, uncertainty and partial truth. For this reason, mathematical rough sets approaches are effective to fields of data analysis, machine learning, artificial intelligent, pattern recognition, information retrieval, survival analysis etc. [8–14,27,16,51,18]. The bridges be-

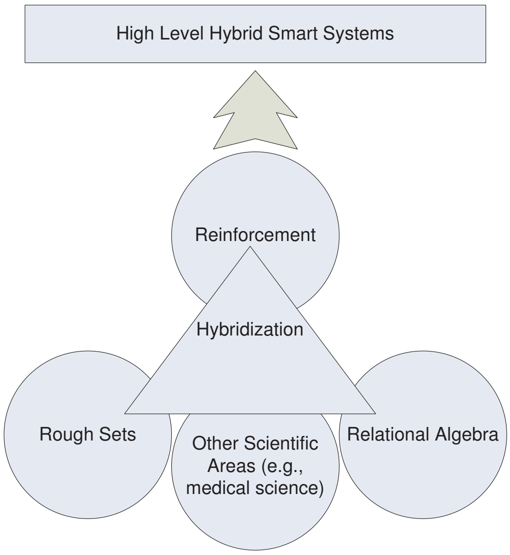


Fig. 1. A perspective of rough sets hybrid intelligent system.

tween rough set theory and various branches of other disciplines have been built [1,2,7–11]. Fig. 1 depicts the sample components of a hybrid intelligent system in soft computing i.e., rough sets, relational algebra and other scientific areas. The reinforcement step increases the intelligence to a high level hybrid intelligent system.

Proposal of fuzzy set theory by Zadeh attracted researchers since the 1965 and one most crucial discovery is the emergence of fuzzy control theory. The first fuzzy controllers were built in the 1980s and play important role regarding natural language, pattern recognition, clustering, image processing, robotics etc. [5,19]. Dubois and Prade (1990) proposed rough fuzzy sets and fuzzy rough sets, theoretical rough sets and fuzzy sets integrations were then established [20]. The study by Banerjee et al. (1996) strengthens the connection between rough set theory and fuzzy set theory with the roughness measure of fuzzy sets [21]. Yang and John (2006) have investigated the roughness bounds under different set operations for rough sets [24].

Not only were the theoretical development of these theories expanded but practical applications of such discoveries were studied as well. Measures of the roughness of fuzzy sets were useful for applications in pattern recognition field [21] and more recently (2004) by Zhang et al. in [25]. Janardhan Rao et al. (2005) proposed to use the rough fuzzy set as the model for images [26]. Roughness measures of fuzzy sets can be optimized for object extraction to determine the intensity threshold of images. The experimental results show that objects are extracted with their approach with higher accuracy compared to Shannon's probabilistic entropy. In [23] (2005), the relational database was analyzed by Huynh and Nakamori's roughness measure.

Nevertheless, when we focus on data mining, machine learning, bioinformatics, computer security, natural language processing, etc., data sets are usually huge. For this reason, we investigate the bounds of roughness measures for fuzzy set operations, namely, union and intersection.

Rough Sets and Fuzzy Sets

The following definitions are taken from the studies of Pawlak [1,2], Zadeh [4], Dubois et al. [20] and Banerjee et al. [21] and will form the basis of this research. We begin with the introduction to basic definitions of rough set theory.

Definition 1 (Approximation Space) *Let U be a non-empty finite set called universe (or domain that we interested in) and R be an equivalence relation on U (R is reflexive, symmetric and transitive), we define $\langle U, R\rangle$ as an approximation space.*

A set A can be approximated in two ways that are *lower approximation* and *upper approximation* in the following definition.

Definition 2 (Lower and Upper Approximations) *Let $\langle U, R\rangle$ be an approximation space, $A \subseteq U$ and $X_1, X_2, ..., X_n$ denote the equivalence classes in U with respect to R ($X_1, X_2, ..., X_n$ partition U into a family of disjoint subsets U/R). The lower approximation $\underline{A}$ and upper approximation $\overline{A}$ are defined as:*

$$\underline{A} = \bigcup\{X_i : X_i \subseteq A\},$$

$$\overline{A} = \bigcup\{X_i : X_i \cap A \neq \emptyset\},$$

where $i \in \{1, 2, ..., n\}$, respectively.

Pawlak's roughness measure of rough sets follows.

Definition 3 (Roughness Measure of Rough Sets) *Let $\langle U, R\rangle$ be an approximation space and $A \subseteq U$, a measure of roughness of A in $\langle U, R\rangle$, ρ_A, defined as:*

$$\rho_A = 1 - \frac{|\underline{A}|}{|\overline{A}|},$$

where $|X|$ denotes the cardinality of a the set X.

Yao [27] interpreted Pawlak's accuracy measure [2] in terms of *Marczewski-Steinhaus metric*. Consequently, he stated that the roughness measure can be understood as the distance between the lower and upper approximations. We next come to some basic concepts of fuzzy sets which will be necessary fo our present bounds.

Let $\mathbf{A}$: $U \to [0,1]$ be a fuzzy set in U, $\mathbf{A}(x)$, $x \in U$, giving the degree of membership of x in $\mathbf{A}$.

Definition 4 (Lower and Upper Approximations of Fuzzy Sets) *The lower and upper approximations of fuzzy set* $\mathbf{A}$ *in* U*, denoted* $\underline{\mathbf{A}}$ *and* $\overline{\mathbf{A}}$*, respectively, are defined as fuzzy sets in* $U/R \to [0,1]$*, such that*

$$\underline{\mathbf{A}}(X_i) = inf_{x \in X_i}\mathbf{A}(x) \qquad and$$

$$\overline{\mathbf{A}}(X_i) = sup_{x \in X_i}\mathbf{A}(x), \quad i = 1, \ldots, n,$$

where inf (sup) denotes minimum (maximum).

When $\mathbf{A}$ is a crisp set, $\underline{\mathbf{A}}$ and $\overline{\mathbf{A}}$ reduce to the collection of equivalence classes constituting its lower and upper approximation in $\langle U, R \rangle$, respectively [21].

Definition 5 (Fuzzy Set) *Fuzzy sets* $\underline{\mathscr{A}}$, $\overline{\mathscr{A}} : U \to [0,1]$ *are defined as follows:*

$$\underline{\mathscr{A}}(x) = \underline{\mathbf{A}}(\mathbf{X_i}) \quad and$$

$$\overline{\mathscr{A}}(x) = \overline{\mathbf{A}}(\mathbf{X_i}),$$

if $x \in X_i, i \in \{1, \ldots, n\}$.

Roughness Measure of Fuzzy Sets

In this section, we consider the roughness measure of a fuzzy set associated with the parameters α and β taken from [21] .

Definition 6 *[21] Suppose* α, β *are two given parameters, where* $0 < \beta \leq \alpha \leq 1$*. The* α*-cut set,* β*-cut set of fuzzy sets* $\underline{\mathscr{A}}$*,* $\overline{\mathscr{A}}$ *are defined as* $\underline{\mathscr{A}}_\alpha = \{x : \underline{\mathscr{A}}(x) \geq \alpha\}$ *and* $\overline{\mathscr{A}}_\beta = \{x : \overline{\mathscr{A}}(x) \geq \beta\}$*, respectively.*

$\underline{\mathscr{A}}_\alpha$ and $\overline{\mathscr{A}}_\beta$ can be considered as the collection of objects with α as the minimum degree of definite membership, and β as the minimum degree of possible membership in the fuzzy set $\mathscr{A}$ [21].

We next come to definition of roughness measure of fuzzy set.

Definition 7 *[21] A roughness measure $\rho_{\mathbf{A}}^{\alpha,\beta}$ of fuzzy set $\mathbf{A}$ in U with respect to parameters α, β, where $0 < \beta \leq \alpha \leq 1$, and the approximation space $\langle U, R\rangle$, is defined as*

$$\rho_{\mathbf{A}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A}}_{\alpha}|}{|\overline{\mathscr{A}}_{\beta}|}.$$

This roughness measure depends strongly on parameters α and β. The studies of parameter free roughness measure of fuzzy set can be found in [23]. There are several crucial properties of $\rho_{\mathbf{A}}^{\alpha,\beta}$ introduced in [21] as follows.

Proposition 1 *Let α, β: $U \to [0,1]$ be two fuzzy sets in universe U and $0 < \beta \leq \alpha \leq 1$, we have*

(a) $\overline{\mathscr{A} \cup \mathscr{B}}_{\beta} = \overline{\mathscr{A}}_{\beta} \cup \overline{\mathscr{B}}_{\beta}$,

(b) $\underline{\mathscr{A} \cap \mathscr{B}}_{\alpha} = \underline{\mathscr{A}}_{\alpha} \cap \underline{\mathscr{B}}_{\alpha}$,

(c) $\underline{\mathscr{A}}_{\alpha} \cup \underline{\mathscr{B}}_{\alpha} \subseteq \underline{\mathscr{A} \cup \mathscr{B}}_{\alpha}$,

(d) $\overline{\mathscr{A} \cap \mathscr{B}}_{\beta} \subseteq \overline{\mathscr{A}}_{\beta} \cap \overline{\mathscr{B}}_{\beta}$.

Property 1 *For fuzzy sets $\mathbf{A}$ and $\mathbf{B}$, it holds that*

(a) $$\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A} \cup \mathscr{B}}_{\beta}|} = 1 - \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A}}_{\beta} \cup \overline{\mathscr{B}}_{\beta}|} \leq 1 - \frac{|\underline{\mathscr{A}}_{\alpha} \cup \underline{\mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A}}_{\beta} \cup \overline{\mathscr{B}}_{\beta}|},$$

(b) $$\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A} \cap \mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A} \cap \mathscr{B}}_{\beta}|} = 1 - \frac{|\underline{\mathscr{A}}_{\alpha} \cap \underline{\mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A} \cap \mathscr{B}}_{\beta}|} \leq 1 - \frac{|\underline{\mathscr{A}}_{\alpha} \cap \underline{\mathscr{B}}_{\alpha}|}{|\overline{\mathscr{A}}_{\beta} \cap \overline{\mathscr{B}}_{\beta}|}.$$

For more relations between roughness measures of fuzzy sets $\mathbf{A}, \mathbf{B}, \mathbf{A} \cup \mathbf{B}$ and $\mathbf{A} \cap \mathbf{B}$ please refer to [21].

The properties of the roughness measure in Definition 7, Proposition 1, Property 1 and etc. will be used to derive the bounds in this research.

Certain Increment and Uncertain Decrement Operators

The pioneering studies of fuzzy rough sets [2,21] derived Propositions 1(c) and (d). They perform *subset* instead of *set equal*. This cannot be analyzed quantitatively. Successive roughness measures also depend on the parameters

α, β, and these two difficulties restrict some computations [25]. Thus, two new parameter-free operators of fuzzy sets were devised [25].

Definition 8 *[25] Let U be the universe and let R be an equivalence relation on U. Let X, $Y \subseteq U$. When X is extended by Y (i.e., $X \cup Y$), $\underline{Z}_{(\cdot)}(\cdot) : U \times U \to U$ defined by $\underline{Z}_{(X)}(Y) = \bigcup\{[x]_R \,|\, x \in L(X),\, l_X(x) \not\subseteq Y \text{ and } h_X(x) \subseteq Y\}$, is called the certain increment operator of X, where $L(X) = \bigcup\{l_X(x)|x \in BN_R(X) \cap X\}$, $h_X(x) = [x]_R - X$, and $l_X(x) = [x]_R - h_X(x)$.*

Definition 9 *[25] Let U be the universe and let R be an equivalence relation on U. Let X, $Y \subseteq U$. When X is cut by Y (i.e., $X \cap Y$), $\overline{Z}_{(\cdot)}(\cdot) : U \times U \to U$ defined by $\overline{Z}_{(X)}(Y) = \bigcup\{[x]_R \,|\, x \in L(X),\, l_X(x) \cap Y = \emptyset \text{ and } h_X(x) \cap Y \neq \emptyset\}$, is called the uncertain decrement operator of X, where $L(X) = \bigcup\{l_X(x)|x \in BN_R(X) \cap X\}$, $h_X(x) = [x]_R - X$, and $l_X(x) = [x]_R - h_X(x)$.*

The certain increment operator is the collection of objects for which the certain information of $X \cup Y$ is larger than the union of the certain information of X and Y [25]. The uncertain decrement operator is the collection of objects for which the uncertain information of $X \cap Y$ is less than the intersection of the uncertain information of X and Y [25].

Property 2 *[25] For crisp sets X, $Y \subseteq U$, it holds that*
(a) $\underline{Z}_{(X)}(Y) = \underline{Z}_{(Y)}(X)$,
(b) $\overline{Z}_{(X)}(Y) = \overline{Z}_{(Y)}(X)$.

Theorem 1 *[25] Let α, β: $U \to [0,1]$ be two fuzzy sets in universe U and $0 < \beta \leq \alpha \leq 1$, where $\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)$, $\underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha)$, $\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)$ and $\overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta)$ are certain increment operators of $\mathscr{A}_\alpha$, $\mathscr{B}_\alpha$ and the uncertain decrement operators of $\mathscr{A}_\beta$, $\mathscr{B}_\beta$, respectively. We have*
(a) $\underline{\mathscr{A} \cup \mathscr{B}}_\alpha = \underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha) = \underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha)$,
(b) $\overline{\mathscr{A} \cap \mathscr{B}}_\beta = \overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta) = \overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta)$.

Property 3 *[25] For fuzzy sets $\mathbf{A}$ and $\mathbf{B}$, it holds that*
(a) $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha)|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|}$,
(b) $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta)|}$.

Some New Bounds of Roughness Measures of Fuzzy Set

The roughness measure is an important indicator of the uncertainty and accuracy associated with a given set [24]. Since data sets are usually huge in most applications, operations on these data sets are time and space consuming. Thus, before completing large volume operations involving two fuzzy sets, we must know the bounds of such results. In this section, we propose new theorems on roughness lower and upper bounds of the fuzzy set operations as the following.

Theorem 2 *An upper bound of the roughness measure* $\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta}$ *of fuzzy sets* $\mathbf{A},\mathbf{B}$ *in* U *with respect to* α, β, *is given by*

$$\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} \leq \frac{1-\rho_{\mathbf{A}}^{\alpha,\beta}\rho_{\mathbf{B}}^{\alpha,\beta}}{2-(\rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta})},$$

where $0<\beta\leq\alpha\leq 1$.

Proof From Property 1(a) and a basic set property, we have $\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} \leq 1-\frac{\max\{|\underline{\mathscr{A}}_\alpha|,|\underline{\mathscr{B}}_\alpha|\}}{|\overline{\mathscr{A}}_\beta|+|\overline{\mathscr{B}}_\beta|}$. If $|\underline{\mathscr{A}}_\alpha| \geq |\underline{\mathscr{B}}_\alpha|$ (or vice versa), then we obtain $\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} \leq 1-\frac{1}{(|\overline{\mathscr{A}}_\beta|/|\underline{\mathscr{A}}_\alpha|)+(|\overline{\mathscr{B}}_\beta|/|\underline{\mathscr{A}}_\alpha|)}$.

Thus, $\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} \leq \frac{1-\rho_{\mathbf{A}}^{\alpha,\beta}\rho_{\mathbf{B}}^{\alpha,\beta}}{2-(\rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta})}$ by Definition 7. $\square$

Theorem 2 illustrates that upper bound of the union of two fuzzy sets depends solely on the roughness measures of fuzzy sets **A** and **B**.

Theorem 3 *An upper bound of the roughness measure* $\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta}$ *of fuzzy sets* $\mathbf{A},\mathbf{B}$ *in* U *with respect to* α, β, *is given by*

$$\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta}-1+U^*,$$

where $0<\beta\leq\alpha\leq 1$ *and* $U^*=\frac{|\underline{\mathscr{A}\cup\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$.

Proof From Property 1(b) and a basic set property, $\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} = 1-\frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}-\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}+\frac{|\underline{\mathscr{A}}_\alpha\cup\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$.

With respect to Proposition 1(d), $\overline{\mathscr{A}\cap\mathscr{B}}_\beta \subseteq \overline{\mathscr{A}}_\beta$ (and $\subseteq \overline{\mathscr{B}}_\beta$) implies $|\overline{\mathscr{A}\cap\mathscr{B}}_\beta| \leq |\overline{\mathscr{A}}_\beta|$ (and $\leq |\overline{\mathscr{B}}_\beta|$), we therefore have $\frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} \leq \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$ and $\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|} \leq \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$. Then $\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} \leq 1-\frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|}-\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|}+\frac{|\underline{\mathscr{A}\cup\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$ by Proposition 1(c).

According to Definition 7, $\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta}-1+\frac{|\underline{\mathscr{A}\cup\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$.

We finally have $\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta}-1+U^*$, where $U^*=\frac{|\underline{\mathscr{A}\cup\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|}$. $\square$

The bound in Theorem 2 depends on the roughness measures of fuzzy sets **A** and **B** while the bound in Theorem 3 depends on the roughness measures of fuzzy sets **A** and **B** in addition to $|\underline{\mathscr{A}\cup\mathscr{B}}_\alpha|$ and $|\overline{\mathscr{A}\cap\mathscr{B}}_\beta|$.

Theorem 4 *A lower bound on the roughness measure $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}}$ of fuzzy sets $\mathbf{A},\mathbf{B}$ in U with respect to α, β, is given by*

$$\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq \rho^{\alpha,\beta}_{\mathbf{A}} + \rho^{\alpha,\beta}_{\mathbf{B}} - 1 - L_*,$$

where $0 < \beta \leq \alpha \leq 1$ *and* $U_* = \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$.

Proof From Property 3(a) and a basic set property, we have that $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha|+|\underline{\mathscr{B}}_\alpha|+|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$.

For $|\overline{\mathscr{A}}_\beta| \geq |\overline{\mathscr{B}}_\beta|$, $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha|+|\underline{\mathscr{B}}_\alpha|+|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$.

In accordance with Definition 7 and $\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} \leq \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|}$, we have $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|} - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|} = 1 - (1 - \rho^{\alpha,\beta}_{\mathbf{A}}) - (1 - \rho^{\alpha,\beta}_{\mathbf{B}}) - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$.

Therefore, $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq \rho^{\alpha,\beta}_{\mathbf{A}} + \rho^{\alpha,\beta}_{\mathbf{B}} - 1 - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$. Similarly, for $|\overline{\mathscr{A}}_\beta| < |\overline{\mathscr{B}}_\beta|$, $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha|+|\underline{\mathscr{B}}_\alpha|+|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{B}}_\beta|}$. Therefore, $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq \rho^{\alpha,\beta}_{\mathbf{A}} + \rho^{\alpha,\beta}_{\mathbf{B}} - 1 - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{B}}_\beta|}$.

From (1) and (2), we finally have $\rho^{\alpha,\beta}_{\mathbf{A}\cup\mathbf{B}} \geq \rho^{\alpha,\beta}_{\mathbf{A}} + \rho^{\alpha,\beta}_{\mathbf{B}} - 1 - L_*$, where $L_* = \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$. $\square$

Theorem 5 *A lower bound of the roughness measure $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$ of fuzzy sets $\mathbf{A},\mathbf{B}$ in U with respect to α, β, is given by*

$$\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1 - \rho^{\alpha,\beta}_{\mathbf{A}} - \rho^{\alpha,\beta}_{\mathbf{B}} + \rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2 - \rho^{\alpha,\beta}_{\mathbf{A}} - \rho^{\alpha,\beta}_{\mathbf{B}} - I_*(1 - \rho^{\alpha,\beta}_{\mathbf{A}})(1 - \rho^{\alpha,\beta}_{\mathbf{B}})},$$

where $0 < \beta \leq \alpha \leq 1$ *and* $I_* = \frac{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta| + |\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{\min\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}$.

Proof From Property 3(b), and basic set properties, then we have that $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$ $1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta| - |\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|} \geq 1 - \frac{\min\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}{|\overline{\mathscr{A}}_\beta| + |\overline{\mathscr{B}}_\beta| - |\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta| - |\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}$.

For $|\underline{\mathscr{A}}_\alpha| \leq |\underline{\mathscr{B}}_\alpha|$, we have that $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1}{\frac{|\overline{\mathscr{A}}_\beta|}{|\underline{\mathscr{A}}_\alpha|} + \frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|} - \frac{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|} - \frac{|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}}$.

From Definition 7 and $\frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|} \geq \frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{B}}_\alpha|}$, $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1}{\frac{|\overline{\mathscr{A}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}+\frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{B}}_\alpha|}-\frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}-\frac{|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}}$ $= 1 - \frac{1}{\frac{1}{1-\rho^{\alpha,\beta}_{\mathrm{A}}}+\frac{1}{1-\rho^{\alpha,\beta}_{\mathrm{B}}}-\frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}}$. We define $I_{|\underline{\mathscr{A}}_\alpha|} = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}$, thus $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_{|\underline{\mathscr{A}}_\alpha|}(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$.

Similarly, for $|\underline{\mathscr{A}}_\alpha| > |\underline{\mathscr{B}}_\alpha|$, if we define $I_{|\underline{\mathscr{B}}_\alpha|} = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{B}}_\alpha|}$, therefore we have $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_{|\underline{\mathscr{B}}_\alpha|}(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$.

Thus, $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_*(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$, where $I_* = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{\min\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}$. $\square$

The lower bounds for $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$ differ from the upper bounds on $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$, in that they depend on the roughness measures of the fuzzy sets **A** and **B** and also $|\overline{\mathscr{A}}_\beta|$, $|\overline{\mathscr{B}}_\beta|$, $|\underline{\mathscr{A}}_\alpha|$, $|\underline{\mathscr{B}}_\alpha|$ and $|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|$.

Concluding Remarks

The roughness measure is an important indicator of uncertainty and accuracy associated with a given fuzzy set. More importantly are the propagations of these roughness measures through various set operations. Thus, before completing large volume operations involving two large fuzzy sets, we should have bounds on the roughness measure of the result as were proven in this work.

4.2 Capacity-Based Definite Rough Integral

This section introduces an extension of the original capacity-based rough integral defined over a specific interval. The approach hearkens back to the pioneering work on capacities and the generalization of the Lebesgue integral by Gustav Choquet during the 1950s. Variations in the definition of the capacity function has led to various forms of the discrete Choquet integral. In particular, it is the rough capacity function (also called a rough membership function) introduced by Zdzisław Pawlak and Andrzej Skowron during the 1990s that led to the introduction of a capacity-based rough integral.

By extension of the original work on the rough integral introduced in 2000, a discrete form of a capacity-based definite rough integral is introduced in this section. This new form of the rough integral provides a means of measuring the relevance of functions representing features useful in the classification of sets of sample objects.

Introduction

During the early 1990s, Zdzisław Pawlak introduced a discrete form of the definite Riemann integral of a continuous, real function defined over intervals representing equivalence classes [30]. Since that time, work on various forms of rough integrals has continued fairly steadily (see, *e.g.*, [39,41,37,31]). The original integral was called a *rough integral* because it was defined over intervals represented equivalence classes in a partition of sequences of reals defined by an equivalence relation.

Based on work on the Choquet integral $(C)\int f d\mu$ considered in the context of rough sets [32], a new form of discrete rough integral $\int f d\mu_x^B$ was introduced [39,41,41,37] and elaborated in [37]. The Choquet integral is a generalization of the Lebesgue integral defined relative to a capacity μ [34,33].

A *capacity* is a function μ that assigns a non-negative real number to every subset of a finite set X and satisfies $f(\emptyset) = 0$ [35]. When the discrete form of the Choquet integral $(C)\int f d\mu$ is defined relative to a finite universe, the Lebesgue integral reduces to a (convex) linear combination, where each individual integrand function value is weighted with a capacity function value [34].

This section will introduces a new form of the Choquet integral called a *capacity-based definite rough integral* because its capacity is a function defined relative to equivalence classes. The extension of the capacity-based rough integral has a number applications. In particular, we show later how this integral can be used in feature selection with the DRI tool implemented in MATLAB®.

Rough Capacity Function

This section gives a brief introduction to one form of additive set functions in measure theory from [37]. Let X be a finite set, we denote $card(X)$ as the cardinality of X.

Definition 10 (Set Function) *Let X be a finite, non-empty set. A function $\lambda : \wp \rightarrow \Re$ where $\Re$ is the set of all real numbers is called a set function on X.*

Definition 11 (Additive) *Let X be a finite, non-empty set and let λ be a set function on X. The function λ is called to be additive on X if and only if $\lambda(A \cup B = \lambda(A) + \lambda(B)$ for every A, $B \in \wp(X)$ such that $A \cap B = \emptyset$.*

Definition 12 (Non-negative) *Let X be a finite, non-empty set and let λ be a set function on X. A function λ is called to be non-negative on X if and only if $\lambda(Y) \geq 0$ for any $Y \in \wp(X)$.*

Definition 13 (Monotonic) *Let X be a set and let λ be a set function on X. A function λ is called to be monotonic on X if and only if $A \subseteq B$ implies that $\lambda(A) \leq \lambda(B)$ for every $A, B \in \wp(X)$.*

Rough capacity functions were introduced during the mid-90s [38]. A rough capacity function returns the degree of overlap between a fixed set containing objects of interest and a set of sample objects.

Definition 14 (Rough Capacity Function) *Let $S = (\mathcal{O}, \mathcal{F})$ denote an information system. Assume $X \subseteq \wp(\mathcal{O})$, $B \subseteq \mathcal{F}$, $x \in X$ and $[x]_B \subseteq X/\sim_B$. The capacity $\mu_x^B : \wp(\mathcal{O}) \longrightarrow [0,1]$ is defined (1).*

$$\mu_x^B(X) = \begin{cases} \frac{|X \ \cap \ [x]_B|}{|[x]_B|}, & \textit{if } X \neq \emptyset, \\ 0, & \textit{otherwise}, \end{cases} \tag{1}$$

The capacity μ_x^B is an example of a set function, *i.e.*, a function whose domain is a family of sets [36]. This set function measures the degree of overlap between the set X and the class $[x]_B$, *i.e.*, the extent that X is covered by $[x]_B$.

Recall that $[x]_B$ is a set of objects having matching descriptions. This is important in evaluating the set X, since $\mu_x^B(X)$ is a measure of the extent that the objects in X are part of the classification represented by $[x]_B$. In the context of capacity-based integrals, the function μ_x^B is a source of weights, *i.e.*, degree of importance of each set X in a weighted sum for the discrete form of the integral.

It can also be observed that μ_x^B is a real-valued set function that is additive. That is, for X, X' in $\wp(\mathcal{O})$, it can be shown that

$$\mu_x^B(X \ \cup \ X') = \mu_x^B(X) + \mu_x^B(X').$$

Rough Set Theory

In the early 1980's, Pawlak [2] introduced rough set theory. Mathematical rough set theory provides system designers with the capability to handle the uncertainty that commonly exists in data, especially in real world data. Rough sets make balanced approaches between theory and practice possible, and also form a framework for building hybrid systems with pertinent soft computing techniques. In this research rough sets will be theoretically combined to a certain type of discrete integral and illustrate their performance in the example problem.

Let $S = (U, A)$ be an information system where U is a finite, non-empty set of objects and A is a finite, non-empty set of attributes, where $a : U \to V_a$ for every $a \in A$. For each $B \subseteq A$, let there is associated an equivalence relation $ind_A(B)$ such that

$$ind_A(B) = \{(x, x') \in U^2 | \forall a \in B, a(x) = a(x')\},$$

$[x]_B$ refers equivalence classes of $ind_A(B)$, $U/ind_A(B)$ refers the family of all equivalence classes of relation $ind_A(B)$ on U [37,2].

Definition 15 (Rough Membership Function) *(Pawlak et al. [38]) Let $S = (U, A)$ be an information system, $\wp(U)$ be the powerset of U, $B \subseteq A$, $u \in U$ and let $[x]_B$ be an equivalence class of an object $u \in U$ of $ind_A(B)$. The set function, rough membership function is defined by*

$$\mu_u^B : \wp(U) \to [0, 1], \textit{ where } \mu_u^B(X) = \tfrac{|X \cap [u]_B|}{|[u]_B|}.$$

The form of rough membership function in Definition 15 is different from classical definition in which the argument of the rough membership function is an object x and the set X is fixed [2].

Definition 16 (Rough Measure) *(Pawlak et al. [38]) Let $u \in U$. A non-negative, additive set function $\rho_u : \wp(X) \to [0, \infty)$ defined by $\rho_u(Y) = \rho'(Y \cap$*

$[u]_B$) *for* $Y \in \wp(X)$*, where* $\rho' : \wp(X) \to [0, \infty)$ *is called a rough measure relative to* $U/ind_A(B)$ *and* u *on the indiscernibility space* $(X, \wp(X), U/ind_A(B))$*.*

Furthermore, the rough membership function $\mu_u^B : \wp(X) \to [0, 1]$ is a non-negative set function [38].

Proposition 2 *(Pawlak et al. [38]) The rough membership function* μ_u^B *as defined in Definition 16 is additive on* U*.*

Proposition 3 *(Pawlak et al. [39])* $(X, \wp(X), U/ind_A(B), \{\mu_u^B\}_{u\in U})$ *is a rough measure space over* X *and* B*.*

Capacity-Based Discrete Integrals

This section gives a brief introduction to the Choquet integral, which eventually led to the introduction of a capacity-based rough integral.

Discrete Choquet Integral

Recall that the Choquet integral $(C)\int f d\mu$ is a generalization of the Lebesgue integral defined with respect to a non-classical measure μ called a capacity. Also recall that a capacity is a real-valued set function

$$\mu : \wp(X) \longrightarrow \Re^+,$$

such that $\mu(\emptyset) = 0$ and $X' \subset X'' \subset \wp(X)$ implies $\mu(X') \leq \mu(X'')$ (monotonicity). When the Choquet integral is defined relative to a finite sets, then the Choquet integral reduces to a weighted sum that has a variety of applications, especially in multi-criteria decision-making (see, *e.g.*, [34,33]).

Definition 17 (Discrete Choquet Integral [34]) *Let* μ *be a capacity defined on a finite set* X*. The discrete Choquet integral of a function* $f : X \longrightarrow \Re^+$ *with respect to capacity* μ *is defined by*

$$(C)\int f d\mu = \sum_{i=1}^{n} \left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right)\right] \cdot \mu(X_{(i)}),$$

where $\cdot_{(i)}$ *denotes a permuted index so that* $0 \leq f\left(x_{(1)}\right) \leq f\left(x_{(2)}\right) \leq, \ldots, \leq f\left(x_{(i)}\right) \leq, \ldots, \leq f\left(x_{(n)}\right) \leq 1$*. Also,* $X_{(i)} = \left\{x_{(i)}, \ldots, x_{(n)}\right\}$*, and* $f(x_{(0)}) = 0$*.*

The introduction of the rough capacity function μ_x^B paved the way for a discrete rough integral $(P)\int f d\mu_x^B$ named after Zdzisław Pawlak. This rough integral is a variation of the discrete Choquet integral [34,33,39,41,43].

Definition 18 (Discrete Rough Integral) *Let μ_x^B be a rough capacity function defined on a finite set X. The discrete rough integral of a function $f : X \longrightarrow \Re^+$ with respect to capacity μ_x^B is defined by*

$$(P)\int f d\mu_x^B = \sum_{i=1}^{n}\left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right)\right] \cdot \mu_x^B(X_{(i)}),$$

where $\cdot_{(i)}$ denotes a permuted index so that $0 \leq f\left(x_{(1)}\right) \leq f\left(x_{(2)}\right) \leq, \ldots, \leq f\left(x_{(i)}\right) \leq, \ldots, \leq f\left(x_{(n)}\right) \leq 1$. Also, $X_{(i)} = \left\{x_{(i)}, \ldots, x_{(n)}\right\}$, and $f(x_{(0)}) = 0$.

If f is non-negative, then $(P)\int f d\mu_x^B$ represents the lower approximation of the area under the graph of f.

Proposition 4 *If $Min\mu \leq \mu_x^B(X_{(i)}) \leq Max\mu$, $1 \leq i \leq n$, then $0 \leq (P)\int f\ d\mu_x^B \leq Max\mu$.*

PROOF.

$$\begin{aligned}(P)\int f d\mu_x^B &= \sum_{i=1}^{n}\left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right)\right] \cdot \mu_x^B(X_{(i)})\\ &\leq \sum_{i=1}^{n}\left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right)\right] \cdot Max\mu\\ &= Max\mu \cdot \left[\left(f\left(x_{(1)}\right) - f\left(x_{(0)}\right)\right) + \cdots + \left(f\left(x_{(n)}\right) - f\left(x_{(n-1)}\right)\right)\right]\\ &= Max\mu \cdot \left[-f\left(x_{(0)}\right) + f\left(x_{(n)}\right)\right]\\ &\leq Max\mu \qquad \text{(because } f(x_{(0)}) = 0 \text{ and } f\left(x_{(n)}\right) \leq 1).\end{aligned}$$

It can be proven in a straightforward way that $(P)\int f\ d\mu_x^B \geq 0$.

Consider a specialized capacity $\mu_x^{\{f\}}$ defined in terms of a single function $f \in B$, where B is a set of functions representing features of objects in a finite set X.

Proposition 5 *[41] Let $0 < s \leq r$ and $f \in B$. If $f(x) \in [s, r]$ for all $x \in X$, then $(P)\int f\ d\mu_x^{\{B\}} \in (0, r]$.*

Definite Discrete Rough Integral

In this section we introduce a discrete form of definite rough integral (DRI) of a function f denoted by $\int_a^b f\ d\mu_x^B$, where $\cdot_{(i)}$ is a permuted index and a, b such that $x_{(1)} \leq a \leq b \leq x_{(n)}$ are the lower and upper integral limits, respectively. The limits on the rough integral specify the interval over which a function f is integrated. This integral is defined in terms of an upper integral $\overline{\int_a^b} f\ d\mu_x^B$ and a lower integral $\underline{\int_a^b} f\ d\mu_x^B$.

$$\int_a^b f\ d\mu_x^B = \overline{\int_a^b} f\ d\mu_x^B - \underline{\int_a^b} f\ d\mu_x^B.$$

The discrete forms of the lower and upper integral are defined w.r.t $[x_1]_B$ in (2) and (3), respectively.

$$\underline{\int_a^b} f\ d\mu_{x_1}^B = \sum_{i=a}^{b} \left[f(x_{(i-1)}) \right] \Delta_{(i)} \mu_x^B, \tag{2}$$

$$\overline{\int_a^b} f\ d\mu_{x_1}^B = \sum_{i=a}^{b} \left[f(x_{(i)}) \right] \Delta_{(i)} \mu_x^B, \tag{3}$$

where $\Delta_{(i)} \mu_x^B$ is defined in (4).

$$\Delta_{(i)} \mu_x^B = \left| \mu_x^B(X_{(i)}) - \mu_x^B(X_{(i-1)}) \right| \tag{4}$$

Definition 19 (Definite Discrete Rough Integral) *Let μ_x^B denote a rough capacity function with domain $X_{(i)}$ that is the set*

$$X_{(i)} = \left\{ x_{(i)}, x_{(i+1)}, \ldots, x_{(n)} \right\},$$

where $\cdot_{(i)}$ is a permuted index, $X_{(0)} = \emptyset$, and a, b such that $x_{(1)} \leq a \leq b \leq x_{(n)}$ are the lower and upper integral limits, respectively. The difference $\Delta_{(i)} \mu_x^B$ is defined in (4) relative to the set $X_{(i)}$. The discrete definite rough integral of $f : X \to \Re^+$ is defined by

$$\int_a^b f\ d\mu_x^B = \overline{\int_a^b} f\ d\mu_x^B - \underline{\int_a^b} f\ d\mu_x^B,$$

where the lower and upper integrals are defined in (2) and (3), respectively.

Interpretation

Observe that the capacity function μ_x^B is defined in terms of a set of functions B representing the features of sample objects of interest. The set B provides a basis for the relation $\sim_B$ that defines a partition of the sample objects X (source of integral limits). Then a function f is integrated with respect to μ_x^B. In the discrete form of the DRI, μ_x^B provides a weight on each summand $X_{\{i\}}$, a set of sample objects μ_x^B computes the degree of overlap between a class $[x]_B$ representing objects that have been classified relative to the features represented by B. In effect, B is a source of criteria for grouping together objects matching the criteria represented by B. Hence, the definite rough integral indicates the importance and relevance of a function integrated with respect to μ_x^B. Hence, if a function representing an object feature is integrated with respect to μ_x^B, the DRI provide effective means of selecting features that used to discriminate objects. In effect, the definite rough integral is useful for feature selection within the prescribed limits of the integral.

Feature Selection

In this section, we briefly illustrate an application of the discrete definite rough integral (DRI). For simplicity, we assume that each vector of function values used to describe a sample object is evaluated, *e.g.*, acceptable ($d = 1$) vs. unacceptable ($d = 0$). Put $B = \{d\}$, where $d \in \{0, 1\}$ in defining the capacity μ_x^B. Then any set of objects X can be partitioned using the set B.

Illustration

Next, consider, for example, a set of objects described with two functions, namely, ϕ_1, ϕ_2 representing features of objects in a sample X. For the purposes of illustration, we treat ϕ_1, ϕ_2 abstractly *i.e.*, without considering specific functions. Here is a partial, sample description table with the evaluation (decision) column d included. Note that $x_{(i)}$ indicates a permuted object with values in ascending order.

We now compute the DRI relative to $\left[x_{(1)}\right]_{\{d=0\}}$ starting with integration of ϕ_1 where $B = \{d = 0\}$. The lower integral $\underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \ d\mu_{x_1}^{\{d=0\}}$ is

$$\underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \ d\mu_{x_1}^{\{d=0\}} = \phi_1(x_{(0)}) \cdot \Delta_{(1)}\mu_{x_1}^{\{d=0\}} + \phi_1(x_{(1)}) \cdot \Delta_{(2)}\mu_{x_1}^{\{d=0\}} + \ldots = 0.7,$$

Table 1
Sample data

X	ϕ_1	ϕ_2	d	X	ϕ_1	ϕ_2	d
$x_{(1)} = x_4$	0.79224	29.988	0	$x_{(11)} = x_{11}$	0.92282	13.787	1
$x_{(2)} = x_3$	0.79467	30.114	0	$x_{(12)} = x_{12}$	0.93357	11.387	1
$x_{(3)} = x_2$	0.80596	27.402	0	$x_{(13)} = x_{13}$	0.94553	9.7302	1
$x_{(4)} = x_5$	0.81286	27.633	0	$x_{(14)} = x_{14}$	0.90996	13.979	1
$x_{(5)} = x_1$	0.85808	21.754	0	$x_{(15)} = x_{15}$	0.95608	7.1776	1
$x_{(6)} = x_6$	0.86020	22.866	0	$x_{(16)} = x_{16}$	0.94722	11.387	1
$x_{(7)} = x_7$	0.84569	24.43	0	$x_{(17)} = x_{17}$	0.94467	9.0222	1
$x_{(8)} = x_8$	0.87886	20.16	0	$x_{(18)} = x_{18}$	0.96424	6.3259	1
$x_{(9)} = x_9$	0.88235	19.945	0	$x_{(19)} = x_{19}$	0.92804	12.398	1
$x_{(10)} = x_{10}$	0.88094	19.227	0	$x_{(20)} = x_{20}$	0.93925	9.9671	1
				$x_{(21)} = x_{21}$	0.99435	2.3081	1

and the upper integral $\overline{\int_{x_{(1)}}^{x_{(10)}}}\phi_1 \; d\mu_{x_1}^{\{d=0\}}$ is

$$\overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \; d\mu_{x_1}^{\{d=0\}} = \phi_1(x_{(1)}) \cdot \Delta_{(1)}\mu_{x_1}^{\{d\}} + \phi_1(x_{(2)}) \cdot \Delta_{(2)}\mu_{x_1}^{\{d=0\}} + \ldots = 1.45.$$

Using these results, the definite integral $\int_{x_{(1)}}^{x_{(10)}} \phi_1 d\mu_{x_1}^{\{d=0\}}$ is

$$\int_{x_{(1)}}^{x_{(10)}} \phi_1 d\mu_{x_1}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \; d\mu_{x_1}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \; d\mu_{x_1}^{\{d=0\}} = 0.75.$$

Next, integrate ϕ_2 with respect to $\mu_{x_1}^{\{d=0\}}$ using class $\left[x_{(1)}\right]_{\{d=0\}}$, and obtain

$$\int_{x_{(1)}}^{x_{(10)}} \phi_2 d\mu_{x_1}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_2 \; d\mu_{x_1}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_2 \; d\mu_{x_1}^{\{d=0\}} = 2.3.$$

This means that for class $\left[x_{(1)}\right]_{\{d=0\}}$, the feature represented by ϕ_2 is more important than the feature represented by ϕ_1. The computations have been performed using the DRI tool (see Fig. 2) implemented in MATLAB ®.

From the plots in Fig. 3, notice that there is less dispersion of the values for the plot for $\int_{x_{(1)}}^{x_{(10)}} \phi_2 d\mu_{x_1}^{\{d=0\}}$ *i.e.*, the values for the successive lower and upper values are tightly groups around the $\int_{x_{(1)}}^{x_{(10)}} \phi_1 d\mu_{x_1}^{\{d=0\}}$ values compared with the $\int_{x_{(1)}}^{x_{(10)}} \phi_2 d\mu_{x_1}^{\{d=0\}}$ values.

Similarly, consider the case for class $\left[x_{(1)}\right]_{\{d=1\}}$ and use the DRI to discover

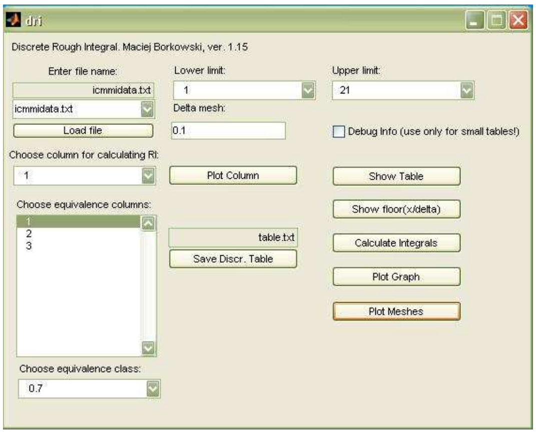
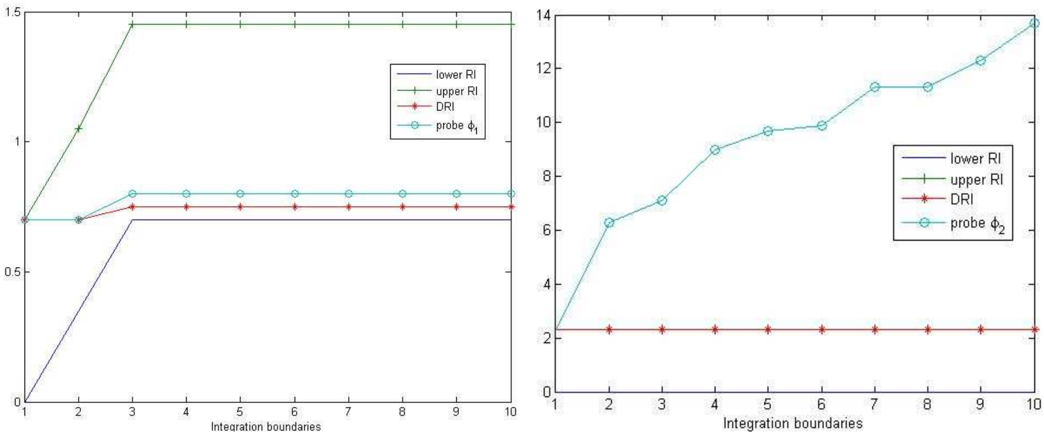


Fig. 2. DRI Tool interface.



(a) DRI $\phi_1, d = 0$ (b) DRI $\phi_2, d = 0$

Fig. 3. DRI for ϕ_1, ϕ_2 for class $[x_1]_{\{d=0\}}$.

which function has greater importance, *i.e.*, carries more weight. Then the DRI computed relative to ϕ_1 is

$$\int_{x_{(11)}}^{x_{(21)}} \phi_1 d\mu_{x_1}^{\{d=1\}} = \int_{x_{(11)}}^{\overline{x_{(21)}}} \phi_1 \; d\mu_{x_1}^{\{d=1\}} - \int_{\underline{x_{(11)}}}^{x_{(21)}} \phi_1 \; d\mu_{x_1}^{\{d=1\}} = 0.$$

Next, integrate ϕ_2 for the same class, and obtain

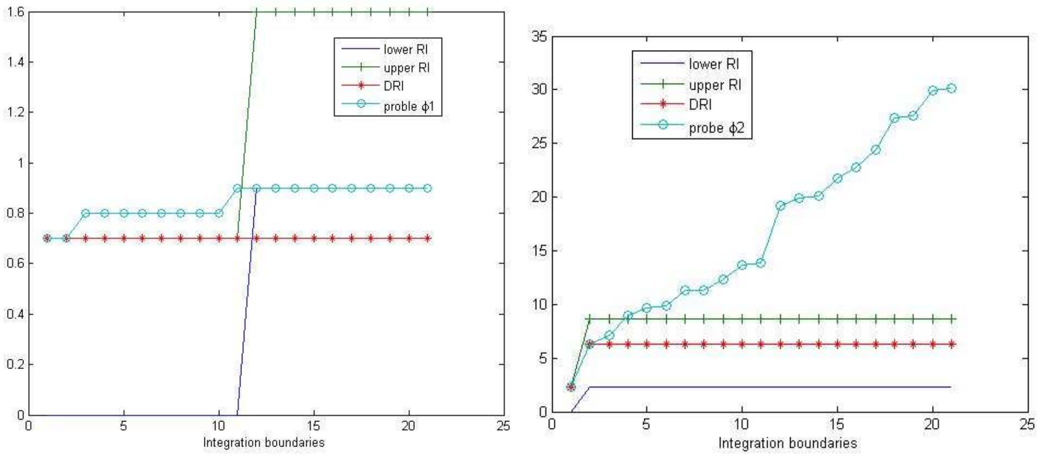
$$\int_{x_{(11)}}^{x_{(21)}} \phi_2 d\mu_{x_1}^{\{d=1\}} = \int_{x_{(11)}}^{\overline{x_{(21)}}} \phi_2 \; d\mu_{x_1}^{\{d=1\}} - \int_{\underline{x_{(11)}}}^{x_{(21)}} \phi_2 \; d\mu_{x_1}^{\{d=1\}} = 0.1.$$

We now calculate the DRI over the entire data set, for both classes $\left[x_{(2)}\right]_{\{d=0\}}$

and $\left[x_{(2)}\right]_{\{d=1\}}$ for the two features ϕ_1 and ϕ_2 respectively:

$$\int_{x_{(1)}}^{x_{(21)}} \phi_1 d\mu_{x_2}^{\{d=0\}} = \int_{x_{(1)}}^{\overline{x_{(21)}}} \phi_1 \; d\mu_{x_2}^{\{d=0\}} - \int_{\underline{x_{(1)}}}^{x_{(21)}} \phi_1 \; d\mu_{x_2}^{\{d=0\}} = 0.7.$$

$$\int_{x_{(1)}}^{x_{(21)}} \phi_2 d\mu_{x_2}^{\{d=0\}} = \int_{x_{(1)}}^{\overline{x_{(21)}}} \phi_2 \; d\mu_{x_2}^{\{d=0\}} - \int_{\underline{x_{(1)}}}^{x_{(21)}} \phi_2 \; d\mu_{x_2}^{\{d=0\}} = 6.3.$$



(a) DRI ϕ_1 (b) DRI ϕ_2

Fig. 4. DRI for ϕ_1, ϕ_2 for class $[x_2]$.

From this, we can conclude that the feature represented by ϕ_2 is more important than the feature represented by ϕ_1 with respect to both classes and for different equivalence classes. The plots in Figures 3 and 4 reveal an interesting feature of the sample objects that are a source of limits on the DRI, *i.e.*, at $x_{(12)}$ there is a sharp change in DRI values (this is especially evident in the plot). This suggests a place to begin experimenting with different limits on the integral. This break in the DRI values also indicates a change in the evaluation of the functions representing object features.

Conclusion

This section introduces the definite rough integral defined relative to a rough capacity function. It is the capacity μ_x^B that distinguishes the rough integral from the original capacity-based integral introduced by Choquet during the 1950s. In addition to the the geometric interpretation of the definite rough integral already mentioned, this integral is highly significant because it offers a measure of the importance and relevance of features of sample objects. This measurement of the relevance of a feature is a byproduct of the rough capacity function that provides the basis for the rough integral. Now, with the introduction of rough integral limits, it possible to measure the relevance of features relative to selected ranges of objects. In effect, the definite rough integral provides a new basis for feature selection in the classification of objects. Future work includes a study of the properties of the definite rough integral and its applications.

4.3 Rough Sets and Flow Graphs Integrations

4.3.1 Rough Sets Approximations in Terms of a Flow Graph

A mathematical flow graph, invented by Pawlak in 2002, is an extension of *rough set theory* [2]. A flow graph represents the information flow from the given data set [52–54,68,56]. The branches of a flow graph can be constructed as decision rules, with every decision rule, there are three associated coefficients: *strength*, *certainty* and *coverage* [56]. These coefficients satisfy Bayes' theorem. Inference in flow graphs has polynomial time and flow conservation comes with probabilistic conditional independencies in the problem domain [44]. Flow graphs have led to many interesting applications and extensions such as preference analysis [53], decision tree [68], survival analysis [63], association rule [46], data mining [56], search engines [45], fuzzy set [47,49], entropy measures [51] and granular computing [48]. More studies involving rough sets are discussed and provided in [57].

Flow distribution in flow graphs can be exploited for approximation and reasoning. Based on flow graph contexts, we define fundamental definitions for rough sets: four categories of vagueness, accuracy of approximation, roughness of approximation and dependency degree. In addition, we state formulas to conveniently compute these measures for inverse flow graphs. To illustrate, a possible car dealer preference analysis is provided to support our propositions. New categories and measures assist and alleviate some limitations in flow graphs to discover new patterns and explanations.

Rough Set Theory

The following rough sets preliminary is taken from [2] . Rough sets are based on an *information system*. An information system is a decision table, whose columns are labeled by attributes, rows are labeled by objects of interest and entries of the table are attribute values. Formally, it is a pair $S = (U, A)$, where U is a nonempty finite set of objects called the *universe* and A is a nonempty finite set of attributes such that a: $U \rightarrow V_a$ for every $a \in A$. The set V_a is called the *domain* of a.

If we partition an information system into two disjoint classes of attributes, called *condition* and *decision attributes*, then the information system will be called a *decision system*, denoted by $S = (U, C, D)$, where $C \cap D = \emptyset$. Any subset B of A determines a binary relation $I(B)$ on U called an *indiscernibility relation*. It is defined as $(x, y) \in I(B)$ if and only if $a(x) = a(y)$ for every $a \in A$, where $a(x)$ denotes the attribute value of element x. Equivalence classes of the relation $I(B)$ are referred to as *B-elementary sets* or *B-elementary granules* denote by $B(X)$, i.e., $B(X)$ describes X in the terms of attribute values from B [53]. Below, we recall key feature definitions of approximations in rough sets.

Definition 20 *[57] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$. The B-lower approximations, B-upper approximations and B-boundary region of X are defined as $\underline{B}(X) = \bigcup_{x\in U} \{B(X)|\, B(X) \subseteq X\}$, $\overline{B}(X) = \bigcup_{x\in U} \{B(X)|\; B(X) \cap X \neq \emptyset\}$ and $BN_B(X) = \overline{B}(X) - \underline{B}(X)$, respectively.*

If the boundary region of X is the empty set (i.e., $BN_B(x) = \emptyset$), then X is *crisp*. On the contrary, if $BN_B(X) \neq \emptyset$, then X is *rough*. In what follows we recall four basic classes of rough sets, i.e., four categories of vagueness. If $BN_B(x) = \emptyset$, then X is *crisp*. On the contrary, if $BN_B(X) \neq \emptyset$, then X is *rough*. In what follows we recall four basic classes of rough sets.

Definition 21 *[57] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$, the four categories of vagueness are defined as*

- $\underline{B}(X) \neq \emptyset$ *and* $\overline{B}(X) \neq U$ *iff* X *is roughly B-definable,*
- $\underline{B}(X) = \emptyset$ *and* $\overline{B}(X) \neq U$ *iff* X *is internally B-indefinable,*
- $\underline{B}(X) \neq \emptyset$ *and* $\overline{B}(X) = U$ *iff* X *is externally B-definable,*
- $\underline{B}(X) = \emptyset$ *and* $\overline{B}(X) = U$ *iff* X *is totally B-indefinable.*

Approximation of a rough set can be characterized numerically by some measurements as follows.

Definition 22 *[57] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$, the accuracy of approximation, $\alpha_B(X)$, and roughness of approximation, $\gamma_B(X)$, are defined respectively as $\alpha_B(X) = \frac{card(\underline{B}(X))}{card(\overline{B}(X))}$ and $\gamma_B(X) = 1 - \alpha_B(X) = 1 - \frac{card(\underline{B}(X))}{card(\overline{B}(X))}$, where $card(X)$ denotes the cardinality of X.*

Let us observe that, $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, then X is *crisp* with respect to B and otherwise, if $\alpha_B(X) < 1$, then X is *rough* with respect to B.

Definition 23 *Let $S = (U, A)$ be an information system and $F = \{X_1, X_2, \ldots, X_n\}$ be a partition of the universe U. For $B \subseteq A$, F depends on B to a degree $k_B(F) = \frac{\sum_{i=1}^{n} card(\underline{B}(X_i))}{card(U)}$.*

Definitions $2-4$ will be stated in the context of flow graphs later.

Flow Graphs

Flow graphs were introduced by Pawlak in 2002 [52]. In this section, we recall some concepts of flow graphs which were introduced by Pawlak in [52–54,68,56].

A *flow graph* is a *directed, acyclic, finite graph* $G = (\mathcal{N}, \mathcal{B}, \varphi)$, where $\mathcal{N}$ is a set of nodes, $\mathcal{B} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of *directed branches*, φ: $\mathcal{B} \to R^+$ is a *flow function* and R^+ is the set of non-negative real numbers. If $(x, y) \in \mathcal{B}$ then x is an *input* of node y denoted by $I(y)$ and y is an *output* of node x denoted by $O(x)$. The *input* and *output* of a flow graph G are defined by $I(G) = \{x \in \mathcal{N} \mid I(x) = \emptyset\}$ and $O(G) = \{x \in \mathcal{N} \mid O(x) = \emptyset\}$. These inputs and outputs of G are called *external nodes* of G whereas other nodes are called *internal nodes* of G. If $(x, y) \in \mathcal{B}$ then we call (x, y) a *throughflow* from x to y. We will assume in what follows that $\varphi(x, y) \neq 0$ for every $(x, y) \in \mathcal{B}$. With every node x of a flow graph G, we have its associated *inflow* and *outflow* respectively as: $\varphi_+(x) = \sum_{y \in I(x)} \varphi(y, x)$ and $\varphi_-(x)=\sum_{y \in O(x)} \varphi(x, y)$.

Similarly, an *inflow* and an *outflow* for the flow graph G are defined as: $\varphi_+(G) = \sum_{x \in I(G)} \varphi_-(x)$ and $\varphi_-(G) = \sum_{x \in O(G)} \varphi_+(x)$. We assume that for any internal node x, $\varphi_-(x) = \varphi_+(x) = \varphi(x)$, where $\varphi(x)$ is a *throughflow* of node x. Similarly then, $\varphi_-(G) = \varphi_+(G) = \varphi(G)$ is a *throughflow* of graph G. As discussed by Pawlak [53], the above equations can be considered as *flow conservation equations* (or *pairwise consistent* [44]).

Normalized Flow Graphs, Paths and Connections

In order to demonstrate interesting relationships between flow graphs and other disciplines (e.g., statistics), we come to the normalized version of flow graphs.

A *normalized flow graph* is a *directed, acyclic, finite graph* $G = (\mathcal{N}, \mathcal{B}, \sigma)$, where $\mathcal{N}$ is a set of *nodes*, $\mathcal{B} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of *directed branches* and σ: $\mathcal{B} \to [0, 1]$ is a *normalized flow function* of (x, y). The *strength* of (x, y) where $0 \leq \sigma(x, y) \leq 1$ is $\sigma(x, y) = \frac{\varphi(x,y)}{\varphi(G)}$.

With every node x of a flow graph G, the associated *normalized inflow* and *outflow* are defined as: $\sigma_+(x) = \frac{\varphi_+(x)}{\varphi(G)} = \sum_{y \in I(x)} \sigma(y, x)$, $\sigma_-(x) = \frac{\varphi_-(x)}{\varphi(G)} = \sum_{y \in O(x)} \sigma(x, y)$. For any internal node x, it holds that $\sigma_+(x) = \sigma_-(x) = \sigma(x)$, where $\sigma(x)$ is a *normalized throughflow* of x. Similarly, *normalized inflow* and *outflow* for the flow graph G are defined as: $\sigma_+(G) = \frac{\varphi_+(G)}{\varphi(G)} = \sum_{x \in I(G)} \sigma_-(x)$, $\sigma_-(G) = \frac{\varphi_-(G)}{\varphi(G)} = \sum_{x \in O(x)} \sigma_+(x)$. It also holds that $\sigma_+(G) = \sigma_-(G) = \sigma(G) =$

1. With every branch (x, y) of a flow graph G, the *certainty* and the *coverage* of (x, y) are defined respectively as: $cer(x, y) = \frac{\sigma(x,y)}{\sigma(x)}$, $cov(x, y) = \frac{\sigma(x,y)}{\sigma(y)}$, where $\sigma(x), \sigma(y) \neq 0$. Properties of these coefficients were studied by Pawlak in [52–54,68,56].

Next, if we focus on sequence of nodes in a flow graph, we can find them by using the concept of a directed simple path. A (directed) *path* from x to y $(x \neq y)$ in G, denoted by $[x \ldots y]$, is a sequence of nodes $x_1, \ldots, x_n$ such that $x_1 = x$ and $x_n = y$ and $(x_i, x_{i+1}) \in B$ for every i, $1 \leq i \leq n-1$. The *certainty, coverage* and *strength of the path* $[x_1 \ldots x_n]$ are defined respectively as: $cer[x_1 \ldots x_n] = \prod_{i=1}^{n-1} cer(x_i, x_{i+1})$, $cov[x_1 \ldots x_n] = \prod_{i=1}^{n-1} cov(x_i, x_{i+1})$, $\sigma[x \ldots y] = \sigma(x) cer[x...y] = \sigma(y) cov[x \ldots y]$.

The set of all paths from x to y $(x \neq y)$ in G, denoted by $\langle x, y \rangle$, is a *connection* of G determined by nodes x and y. For every connection $\langle x, y \rangle$, the associated *certainty, coverage* and *strength of the connection* $\langle x, y \rangle$ are defined as: $cer\,\langle x, y \rangle = \sum_{[x \ldots y] \in \langle x,y \rangle} cer[x \ldots y]$, $cov\,\langle x, y \rangle = \sum_{[x \ldots y] \in \langle x,y \rangle} cov[x \ldots y]$, $\sigma\,\langle x, y \rangle = \sum_{[x \ldots y] \in \langle x,y \rangle} \sigma[x \ldots y] = \sigma(x) cer\,\langle x, y \rangle = \sigma(y) cov\,\langle x, y \rangle$.

If $[x \ldots y]$ is a path such that x and y are the input and output of G, then $[x \ldots y]$ will be referred to as a *complete path*. The set of complete paths from x to y will be called a *complete connection* from x to y in G.

If we substitute every complete connection $\langle x, y \rangle$ in G, where x and y are an input and an output of a graph G with a single branch (x, y) such that $\sigma(x, y) = \sigma\,\langle x, y \rangle$, $cer(x, y) = cer\,\langle x, y \rangle$ and $cov(x, y) = cov\,\langle x, y \rangle$ then we have a new flow graph G' with the property: $\sigma(G) = \sigma(G')$. G' is called a *combined flow graph*. A combined flow graph represents the relationship between its inputs and outputs more precisely.

Starting from a flow graph, if we invert the direction of all branches in G, then the resulting graph G^{-1} will be called the *inverted graph of* G (or the *inverse flow graph of* G) [56]. Essentially, three coefficients of an inverse flow graph can be computed from its flow graph as follows: $\sigma_{G^{-1}}(y, x) = \sigma_G(x, y)$, $cer_{G^{-1}}(y, x) = cov_G(x, y)$ and $cov_{G^{-1}}(y, x) = cer_G(x, y)$.

Rough Set Approximations and Flow Graphs

In this section, we provide a bridge between flow graphs and rough approximation. From standard definitions of approximations made by rough sets, we give these definitions in the context of flow graphs below.

Suppose we are given a normalized flow graph $G = (A, \mathcal{B}, \sigma)$, where $A =$

$\{A_{l_1}, A_{l_2}\ \ , \ldots, A_{l_n}\}$ is a set of attributes [1], $\mathcal{B}$ is a set of directed branches and σ is a normalized flow function. A set of nodes in a flow graph G corresponding to A_{l_i} is referred to as a *layer i*. For $A = C \cup D$, we have that every layer corresponding to C will be called a *condition layer* whereas every layer corresponding to D will be called a *decision layer*. If an attribute A_{l_i} contains n_{l_i} values, we say that it contains n_{l_i} nodes.

We now consider how to approximate an attribute value $Y \in A_{l_{i+1}}$ from attribute values of A_{l_i} where $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, to indicate lower approximation, upper approximation and boundary region of Y. In Definition 24, we recall Pawlak's definitions of lower approximation, upper approximation and boundary region for flow graphs.

Definition 24 *[53] Let $G = (A, \mathcal{B}, \sigma)$ be a normalized flow graph, $A_{l_i} = \left\{X_1, X_2, \ \ldots, X_{n_{l_i}}\right\}$, $1 \le i \le k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of the flow graph G, the union of all inputs X_j of Y is the upper approximation of Y (denoted $\overline{A_{l_i}}(Y)$), the union of all inputs X_j of Y, such that $cer(X_j, Y) = 1$, is the lower approximation of Y (denoted $\underline{A_{l_i}}(Y)$). Moreover, the union of all inputs X_j of Y, such that $cer(X_j, Y) < 1$, is the boundary region of Y (denoted $A_{l_i}N_{A_{l_i}}(Y)$).*

In Definition 25, we state four categories of rough sets mentioned in Definition 21 in terms of flow graph.

Definition 25 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \le i \le k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of G, we define four categories of vagueness as*

- $\exists X_j\,[cer(X_j, Y) = 1]$ *and* $\exists X_j\,[X_j \notin I(Y)]$ *iff* Y *is roughly* A_{l_i}*-definable,*
- $\forall X_j\,[cer(X_j, Y) \neq 1]$ *and* $\exists X_j\,[X_j \notin I(Y)]$ *iff* Y *is internally* A_{l_i}*-indefinable,*
- $\exists X_j\,[cer(X_j, Y) = 1]$ *and* $\forall X_j\,[X_j \in I(Y)]$ *iff* Y *is externally* A_{l_i}*-definable,*
- $\forall X_j\,[cer(X_j, Y) \neq 1]$ *and* $\forall X_j\,[X_j \in I(Y)]$ *iff* Y *is totally* A_{l_i}*-indefinable.*

From the definition we obtain the following interpretation:

- if Y is roughly A_{l_i}*-definable*, this means that we are able to decide for some elements of U whether they belong to Y or $-Y$ [2], using A_{l_i},
- if Y is internally A_{l_i}*-indefinable*, this means that we are able to decide whether some elements of U belong to $-Y$, but we are unable to decide for any element of U, whether it belongs to Y or not, using A_{l_i},

[1] In what follows, we regard $\mathcal{N}$ as A for simplicity.
[2] Where $-Y = U - Y$.

- if Y is externally A_{l_i}*-indefinable*, this means that we are able to decide for some elements of U whether they belong to Y, but we are unable to decide, for any element of U whether it belongs to $-Y$ or not, using A_{l_i},
- if Y is totally A_{l_i}*-indefinable*, we are unable to decide for any element of U whether it belongs to Y or $-Y$, using A_{l_i}.

Property 4 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $2 \leq i \leq k$, be an attribute in layer i and W be a node in $A_{l_{i-1}}$. For any branch (X_j, W), $j \in \{1, \ldots, n_{l_i}\}$ in the inverse flow graph of G, the union of all output X_j of W in flow graph G is the upper approximation of W, the union of all outputs X_j of W in a flow graph G, such that $cov(W, X_j) = 1$, is the lower approximation of W. Moreover, the union of all outputs X_i of W, such that $cov(W, X_j) < 1$, is the boundary region of Y.*

PROOF. It can be proved in a straightforward way according to definition and property of inverse flow graph and Definition 24. □

The following example illustrates the four basic categories.

Example Suppose we are given the flow graph for the preference analysis problem depicted in Fig. 5, that describes four disjoint models of cars $X = \{X_1, X_2, X_3, X_4\}$. They are sold to four disjoint groups of customers $Z = \{Z_1, Z_2, Z_3, Z_4\}$ through three dealers $Y = \{Y_1, Y_2, Y_3\}$.

By Definition 24, when we consider customer Z_1: the lower approximation of Z_1 is an empty set, the upper approximation of Z_1 is $Y_1 \cup Y_2$ and the boundary region Z_1 is $Y_1 \cup Y_2$. Hence, by Definition 25, we conclude that Z_1 is internally Y-indefinable. In Fig. 5 (only limited information is available), by using the set of dealers (Y) to approximate the customer group Z_1 together with the flow distribution visualized in layers two and three, our results can be summarized as the following.

- Since no branch connects Y_3 and Z_1, there is no customer Z_1 buys a car from dealer Y_3. As a result if dealer Y_3 plans to run new promotional campaigns, they do not need to pay attention to customer group Z_1 in these campaigns.
- If a customer buys a car through dealer Y_1 or Y_2, then we cannot conclude whether this is a customer in group Z_1 or not. Thus, if dealers Y_1 and Y_2 plan to run promotional campaigns, then they should, at least, target at customer group Z_1 in their campaigns.

Similarly, we can approximate all attribute values (node) in the inverse flow graph of G by using Property 4.

However, the flow graph perspective on rough sets' categories in Definition 25

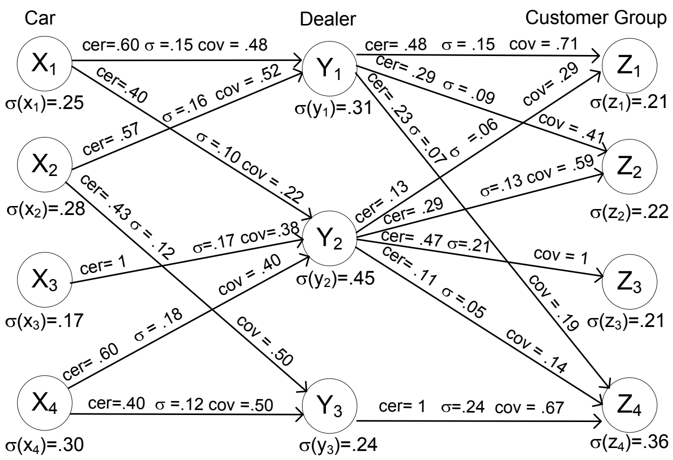


Fig. 5. A normalized flow graph.

do not provide approximations quantitively. Hence, in Definitions 26 and 27, we define two measures for flow graphs, the accuracy of approximation and the roughness of approximation.

Definition 26 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \le i \le k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of G, the accuracy of approximation, $\alpha_{A_{l_i}}(Y)$, is defined as: $\alpha_{A_{l_i}}(Y) = \frac{card(\underline{A_{l_i}}(Y))}{card(\overline{A_{l_i}}(Y))}$.*

We can use the accuracy of approximation to specify the quality of an approximation. Obviously, $0 \le \alpha_B(X) \le 1$. If $\alpha_{A_{l_i}}(Y) = 1$, then Y is crisp with respect to A_{l_i}, and otherwise, if $\alpha_{A_{l_i}}(Y) < 1$, then Y is rough with respect to A_{l_i}.

Definition 27 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \le i \le k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_i, Y), $i \in \{1, \ldots, n_{l_i}\}$, of G, the roughness of approximation, $\gamma_{A_{l_i}}(Y)$, is defined as: $\gamma_{A_{l_i}}(Y) = 1 - \alpha_{A_{l_i}}(Y) = \frac{card(\overline{A_{l_i}}(Y)) - card(\underline{A_{l_i}}(Y))}{card(\overline{A_{l_i}}(Y))}$.*

We have that $0 \le \gamma_{A_{l_i}}(Y) \le 1$. If $\gamma_{A_{l_i}}(Y) = 0$, then Y is crisp with respect to A_{l_i}, and otherwise, if $\gamma_{A_{l_i}}(Y) < 1$, then Y is rough with respect to A_{l_i}.

Property 5 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \le i \le k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of G, we have*

(1) $\alpha_{A_{l_i}}(Y) = \frac{\sum_{cer(X_j,Y)=1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_j)}$ *and (2)* $\gamma_{A_{l_i}}(Y) = \frac{\sum_{cer(X_j,Y)<1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_i)}$.

PROOF. (1) From Definition 24, we have $card(\underline{A_{l_i}}(Y)) = \sum_{cer(X_j,Y)=1} card(X_j)$ and $card(\overline{A_{l_i}}(Y)) = \sum_{X_j \in I(Y)} card(X_i)$. Since $card(X_j) = \varphi(X_j) = \sigma(X_j)\varphi(G) = \sigma(X_j)\varphi(U)$ and by Definition 26, then $\alpha_B(Y) = \frac{\sum_{cer(X_j,Y)=1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_j)}$.

(2) It can be proved similarly to (1). □

Let us briefly comment on Property 2(1) that the greater the boundary of Y, the lower is the accuracy. If $\alpha_{A_{l_i}}(Y) = 1$, the boundary region of Y is empty.

Property 6 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $2 \leq i \leq k$, be an attribute in layer i and W be a node in $A_{l_{i-1}}$. For any branch (X_j, W), $j \in \left\{1, \ldots, n_{l_j}\right\}$ in the inverse flow graph of G, we have*

(1) $\alpha_{A_{l_j}}(W) = \frac{\sum_{cov(W,X_j)=1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$ *and (2)* $\gamma_{A_{l_j}}(W) = \frac{\sum_{cov(W,X_j)<1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$.

PROOF. (1) From Property 4, we have $card(\underline{A_{l_j}}(W)) = \sum_{cov(X_j,W)=1} card(X_j)$ and $card(\overline{A_{l_j}}(Y)) = \sum_{X_j \in O(W)} card(X_j)$. Since $card(X_j) = \varphi(X_j) = \sigma(X_j)\varphi(G) = \sigma(X_j)\varphi(U)$ and by Definition 26, then $\alpha_{A_{l_j}}(W) = \frac{\sum_{cer(X_j,W)=1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$.

(2) It can be proved similarly to (1). □

Example (Cont.) Consider the branches between dealer and customer group in Fig. 5. We can read from our flow graph that 24% of all customers buy cars through dealer Y_3 ($\sigma(Y_3) = 0.24$) and all of them are in customer group Z_3 ($cer(Y_3, Z_4) = 1$). There is only one branch (Y_3, Z_4) with $cer(Y_3, Z_4) = 1$ [3]. Thus, by Property 5(1), we have $\alpha_Y(Z_1) = \alpha_Y(Z_2) = \alpha_Y(Z_3) = 0$ and $\alpha_Y(Z_4) = \frac{\sum_{cer(Y_i,Z_4)=1} \sigma(Y_i)}{\sum_{Y_i \in I(Z_4)} \sigma(Y_i)} = \frac{\sigma(Y_3)}{\sigma(Y_1)+\sigma(Y_2)+\sigma(Y_3)} = 0.24$.

These results imply that we should not make decisions involving customer groups Z_1, Z_2 and Z_3 solely by using dealers due to high imprecision. Nevertheless, we can partly check that it will be customer group Z_4 with low accuracy by using dealers. Similarly, if we consider the roughness of approximation between dealer and customer group, then by Property 5(2), we have $\gamma_Y(Z_1) = \gamma_Y(Z_2) = \gamma_Y(Z_3) = 1$ and $\gamma_Y(Z_4) = 0.76$. We can draw a conclusion in a similar manner as we did for the roughness measure.

---

[3] By employing the approach presented in our previous study [46], we can extract some interesting association rules. If the model of car X_2 (or X_4) is bought through dealer Y_3 then the customer group is Z_4 with support 0.12 and confidence 1.

Please note that we can calculate the accuracy and the roughness of approximation between attributes in the inverse flow graph by using Property 6. Another important topic in data analysis is dependency between attributes. We introduce dependency degree between any two attributes in Definition 28.

Definition 28 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i}= \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$ and $A_{l_{i+1}}= \left\{Y_1, Y_2, \ldots, Y_{n_{l_{i+1}}}\right\}$, $1 \leq i \leq k$, be any two adjacent layers. $A_{l_{i+1}}$ depends on A_{l_i} to a degree $k_{A_{l_i}}(A_{l_{i+1}}) = \frac{\sum_{l=1}^{n_{l_{i+1}}} card(\underline{A_{l_i}}(Y_l))}{card(U)}$.*

If $k_{A_{l_i}}(A_{l_{i+1}}) = 1$, we say that $A_{l_{i+1}}$ depends totally on A_{l_i}, and if $k_{A_{l_i}}(A_{l_{i+1}}) < 1$, we say that $A_{l_{i+1}}$ depends partially in a degree $k_{A_{l_i}}(A_{l_{i+1}})$ on A_{l_i}. It is worth pointing out that our dependency measure is different to the one given by Pawlak [56]. The former gives dependency degree between two adjacent attributes (layers) while the latter gives dependency degree between two nodes connected by directed branch.

Property 7 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i}= \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$ and $A_{l_{i+1}}=\left\{X_1, X_2, \ldots, X_{n_{l_{i+1}}}\right\}$, $1 \leq i \leq k$, be any two adjacent layers. $A_{l_{i+1}}$ depends on A_{l_i} to a degree $k_{A_{l_i}}(A_{l_{i+1}}) = \sum_{cer(X_i,X_j)=1} \sigma(X_i)$.*

PROOF. From Definition 24, $\sum_{j=1}^{n_{l_{i+1}}} card(\underline{A_{l_i}}(X_j))= \sum_{j=1}^{n_{l_{i+1}}} \sum_{cer(X_i,Y_j)=1} card(X_i)$. Since $X_n \cap X_m = \emptyset$, $1 \leq n \neq m \leq n_{l_i}$, then $\underline{A_{l_i}}(X_n) \cap \underline{A_{l_i}}(X_m) = \emptyset$. Thus $\sum_{j=1}^{n} card(\underline{A_{l_i}}(X_j)) = \sum_{cer(X_i,Y_j)=1} card(X_i)$. Since $\varphi(X_i) = \sigma(X_i)\varphi(G) = \sigma(X_i)\varphi(U)$ and by Definition 28, we can write $\gamma_B(D) = \sum_{cer(X_i,X_j)=1} \sigma(X_i)$. □

Property 8 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_j} = \left\{X_1, X_2, \ldots, X_{n_{l_j}}\right\}$ and $A_{l_{j-1}}= \left\{X_1, X_2, \ldots, X_{n_{l_{j-1}}}\right\}$, $1 \leq j \leq k+1$, be any two adjacent layers in the inverse flow graph of G. $A_{l_{j-1}}$ depends on A_{l_j} to a degree $k_{A_{l_j}}(A_{l_{j-1}}) = \sum_{cov(X_i,X_j)=1} \sigma(X_i)$.*

PROOF. It can be proved similarly as Property 7 □

Example (Cont.) Consider model of car and dealer in the flow graph G in Fig. 5. By Property 7, dealer depends on model of car to a degree $\gamma_X(Y) = \sum_{cer(X_i,Y_j)=1} \sigma(X_i) = \sigma(X_3) = 0.17$. On the other hand, if we consider customer and dealer in the inverse flow graph of G, then by Property 8, we obtain that dealer depends on customer group to a degree $\gamma_Z(Y) = \sum_{cov(Y_i,Z_j)=1} \sigma(Z_i) = \sigma(Z_3) = 0.21$. These results give a conclusion that dealers depend on customer groups more than models of cars.

In what follows, we aim to approximate a specific attribute value by some attribute values such that they are not in adjacent layers. We can use the concept of a *connection* to do this. More specifically, if we aim to approximate an attribute value in an output layer by attribute values in an input layer, then we will use the concept of *complete connection*.

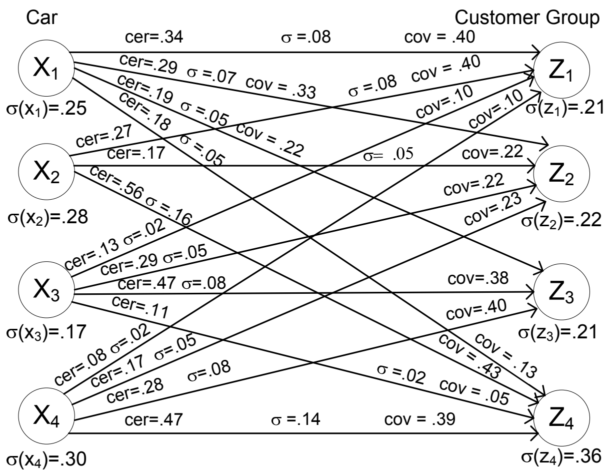


Fig. 6. A combined flow graph.

Example (Cont.) For model of car and customer group in Fig. 5, we give a combined flow graph in Fig. 6. By Definition 24, for Z_4, the lower approximation of Z_4 is an empty set, the upper approximation and the boundary region of Z_4 are $X_1 \cup X_2 \cup X_3 \cup X_4$. Hence, by Definition 25, Z_4 is totally X-indefinable.

By Property 5, we have the accuracy and the roughness approximation of customer Z_4 by model of car as: $\alpha_X(Z_4) = 0$ and $\gamma_X(Z_4) = 1$. Additionally, we can use Property 7 to compute the dependency between model of car and customer group, and the result is 0. From these results due to the imprecision and dependency, we should not make decisions involving customer group Z_4 by using only model of car. As before, we can approximate and measure them for the inverse flow graph in the same way. Comparing the obtained accuracy and roughness measures, we can draw a conclusion that from this population *dealer* is a better indicator for analyzing customer group Z_4 than *model of car*.

Conclusion

In this research, we introduce definitions and properties of rough set approximations, accuracy and roughness of approximation which are defined in terms of a flow graph. They can be useful when the initial data is in the form of flow graph and contains some limitations. We illustrate a car dealer preference analysis to support our propositions. Our future work is to explore relationships between flow graphs and three-way decision rules.

4.3.2 Flow Graphs and Decision Trees

Flow graphs, invented by Pawlak as an extention of *rough set theory* [64], model the information flow of a given data set [65–68]. When starting from a large data set (as in databases around the world), reasoning is referred to as *inductive reasoning.* Reasoning using flow graphs is included in inductive reasoning. This is in contrast to *deductive reasoning*, where axioms expressing some universal truths are used as a departure point of reasoning [65]. We can discover dependencies, correlations and decision rules within a data set without reference to its probabilistic nature by using flow graphs [65]. It is an efficient method for uncertainty management, partly because the branches of a flow graph are interpreted as decision rules. Flow graphs play an important role in reasoning from uncertain data and have been successfully applied in many areas e.g., fuzzy sets [60], search engines [61], rule analysis [63], conflict analysis [67] and data mining [68].

We look at two developments here. One concerns the quality of an individual flow graph. A promising measure considered in this paper is entropy. A decision tree can be constructed as a unique flow graph by removing the root while its nodes are labeled by the same attribute [68]. We further investigate decision tree generation from flow graphs, which is the inverse problem. Thus, creation of decision trees can be accomplished without referring to decision tables but using the information flow about the problem we are interested in.

Flow Graphs

In this section we breifly review and discuss basic definitions and some mathematical properties of flow graphs from the studies of Pawlak [66,68].

Flow graphs have traditionally been used for managing uncertainty [58,60,61,63,65–68]. In order to demonstrate interesting relationships between flow graphs and other disciplines, we consider the normalized version of flow graphs.

A *normalized flow graph* is a *directed, acyclic, finite graph* $G = (N,\ B,\ \sigma)$, where N is a set of *nodes*, $B \subseteq N \times N$ is a set of *directed branches* φ: $B \to R^+$ is a flow function, $\varphi(G)$ is a *throughflow* of flow graph G, σ: $B \to [0, 1]$ is a *normalized flow* of (x, y) and $\sigma(x)$ is a *normalized throughflow* of x.

With every decision rule, there are three associated coefficients: *strength*, *certainty* and *coverage*. The *strength* of (x, y) is given by

$$\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)}. \tag{5}$$

For every node x of a flow graph G, the associated *normalized inflow* and *outflow* are defined respectively as $\varphi_+(x) = \sum_{y \in I(x)} \sigma(y,x)$, $\varphi_-(x) = \sum_{y \in O(x)} \sigma(x,y)$. For every branch (x,y) of a flow graph G, the *certainty* and the *coverage* of (x,y) are defined respectively as:

$$cer(x,y) = \frac{\sigma(x,y)}{\sigma(x)}, \tag{6}$$

$$cov(x,y) = \frac{\sigma(x,y)}{\sigma(y)} \tag{7}$$

where $\sigma(x), \sigma(y) \neq 0$. As a consequence of the previous definitions, the following properties hold: $\sum_{y \in O(x)} cer(x,y) = \sum_{y \in I(x)} cov(x,y) = 1$, $\sigma(x) = \sum_{y \in O(x)} cer(x,y)\sigma(x) = \sum_{y \in O(x)} \sigma(x,y) = \sum_{x \in I(x)} cov(x,y)\sigma(y) = \sum_{x \in I(x)} \sigma(x,y)$ and $cer(x,y) = \frac{cov(x,y)\sigma(y)}{\sigma(x)}$ and $cov(x,y) = \frac{cer(x,y)\sigma(x)}{\sigma(y)}$ [66,68]. The two last equations are Bayes' rules [65] which simplifies computation. Furthermore, *flow conservations* of flow graphs are discussed in [66].

Example 1 Consider the well known weather data set for data mining, given in Table 2 [62], where *Outlook*, *Temperature*, *Humidity*, *Wind* are condition attributes and *PlayTennis* is the decision attribute. Fig. 7 depicts the flow graph corresponding to this data set. Each group of nodes (vertically) in the flow graph is referred to as a *layer*, for example, Fig. 7 has five layers. Every layer corresponds to a particular attribute, and all nodes of a layer correspond to possible values of the attribute. We can interpret some patterns e.g., the database shows 36% sunny outlook, 29% overcast outlook and 36% rain outlook. We also know that 40% of sunny days are high, 40% are mild and 20% are low, etc. Briefly, the flow graph visualizes the information of the data given in Table 2.

If we focus on *Wind* and the decision *PlayTennis*, then we can construct a flow graph to analyze its information flow and decision rules as shown in Fig. 8(a) [4]. Nodes in the first layer are the possible values of *Wind* labelled by Weak and Strong with their normalized throughflow, $\sigma(x)$, (calculated from $\frac{8}{14}$ and $\frac{6}{14}$) indicate there are 57% and 43% of days which have weak and strong wind, respectively. The nodes in the second layer are the possible values of *PlayTennis*, labelled by No and Yes. These nodes indicate that 64% and 36% of the days they do and do not play tennis, respectively.

All branches are interpreted as decision rules with certainty, strength and coverage coefficients computed by (5)–(7). In the branches starting from Weak, $\sigma = 0.14$ and $\sigma = 0.43$, there are 14% and 43% of days that wind is weak but

[4] The computations may contain roundoff errors.

Table 2
Weather data set [62].

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

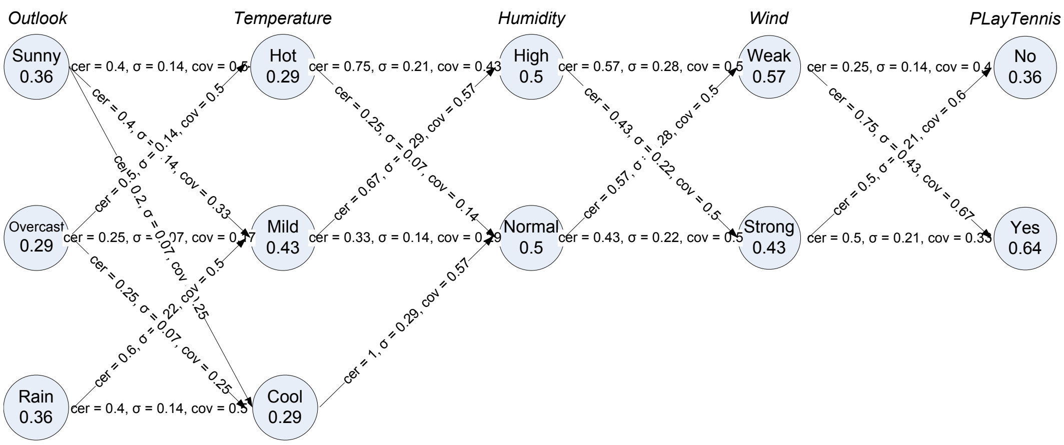
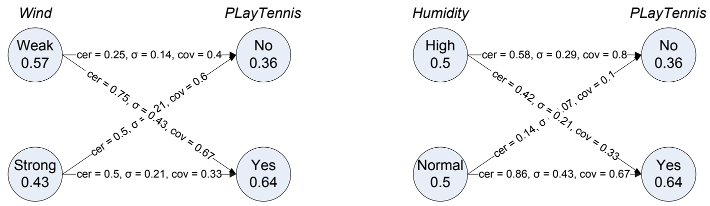


Fig. 7. Flow graph weather data.

they do not play tennis and play tennis, respectively. Accordingly, $cer = 0.25$ and $cer = 0.75$ indicate that there are 25% and 75% of the weak wind days that they do not play tennis and play tennis, respectively. Finally, for the branches ending at No, $cov = 0.4$ and $cov = 0.6$ indicate that there are 40% and 60% of the do not play tennis days which are weak and strong wind, respectively. Similarly, Fig. 8(b) illustrates the flow graph of *Humidity* and *PlayTennis*. Traditionally, decision rules from flow graphs with large values of certainty are included in the classifier system e.g., IF *Wind* = Weak THEN *PlayTennis* = Yes or IF *Humidity* = High THEN *PlayTennis* = No. The respective values of coverage are useful to give explanations (reasons) for these decision rules.



(a) *Wind* and *PlayTennis* (b) *Humidity* and *PlayTennis*

Fig. 8. Flow graphs of weather data.

Entropy

Entropy-based measurements of uncertainty and predictability of flow graphs are considered in this paper. The entropy of a random variable measures the uncertainty associated with that variable (sometimes called the *Shannon entropy* [69]).

Definition 29 *If X is a discrete random variable and $p(x)$ is the value of probability distribution, then the entropy of X is $H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$.*

Definition 30 *If X and Y are discrete random variables and $p(x, y)$ is the value of their joint probability distribution at (x, y), then the joint entropy of X and Y is $H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y)$.*

Joint entropy is the amount of information in two (or more) random variables whereas conditional entropy is the amount of information in one random variable given we already know the other.

Definition 31 *If X and Y are discrete random variables and $p(x, y)$ and $p(y|x)$ are the values of their joint and conditional probability distributions, then the conditional entropy of Y given X is*

$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x)$.

Entropy Measures of Flow Graphs

Similar to the standard definition of entropy, we give definitions of the entropy of a flow graph in this section (measured in *bits*).

Traditionally, given a collection (training data) S, the entropy is defined relative to its classification to characterizes the purity of this collection of examples. This is essentially the entropy of the probability distribution defined by the data set for the decision attribute Y. We define the entropy of a flow graph by replacing this probability with the normalized throughflow of value y, which is natural, since $\sigma(y) = p(Y = y)$.

Definition 32 *The entropy, $H(G)$, of a flow graph G is defined as*

$$H(G) = -\sum_{y \in Y} \sigma(y) \log_2 \sigma(y)$$

where $\sigma(y)$ is the normalized throughflow of the decision attribute value $Y = y$.

The joint entropy and conditional entropy between attributes X and Y are also defined similar to existing definitions.

Definition 33 *If X and Y are attributes in the flow graph G then the joint entropy of X and Y is*

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} \sigma(x, y) \log_2 \sigma(x, y)$$

where $\sigma(x, y)$ is the strength coefficient of the attribute values $X = x$ and $Y = y$.

The task of inference in knowledge discovery is related to computing $p(Y = y|X = x)$ where x and y are values of condition and decision attributes. This how the entropy measures of a flow graph describe the predictive performance of an attribute. Recall that $cer(x, y) = p(Y = y|X = x)$. We define the conditional entropy of attributes in a flow graph below.

Definition 34 *If X and Y are attributes in the flow graph G then the conditional entropy of Y given X is defined as*

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} \sigma(x, y) \log_2 cer(x, y).$$

where $\sigma(x, y)$ and $cer(x, y)$ are the strength and certainty coefficients of the attribute values $X = x$ and $Y = y$, respectively.

Next, we give a formula to compute the information gain, $Gain(G, X)$ of a particular condition attribute X. It is the original entropy of the flow graph (Definition 32) minus the conditional entropy of Y given X.

Definition 35 *If X and Y are attributes in the flow graph G then the information gain, $Gain(G, X)$ of an attribute X, relative to a flow graph G, is defined as*

$$Gain(G, X) = H(G) - H(Y|X).$$

Information gain can serve as a tool for predictive performance discovery. It measures the effectiveness of an attribute to classify the given (training) flow graph. In other words, it indicates the best prediction attribute (having highest $Gain(G, X)$) of decision attributes in flow graph. Using these new definitions, we will analyze flow graphs not in probabilistic flavor but in deterministic flavor in detail in the next sections.

Illustrative Example

The attributes of the data[5] given in Fig. 8, can be regarded as discrete random variables where *Wind* takes values $w = \{\text{Weak, Strong}\}$, *Humidity*—$h = \{\text{High, Normal}\}$ and *PlayTennis*—$t = \{\text{Yes, No}\}$.

First let us focus on *Wind* and *PlayTennis* in the flow graph in Fig. 8(a), their normalized throughflows are:

- $\sigma(x)$ of *Wind* are $\sigma(Weak) = 0.57$ and $\sigma(Strong) = 0.43$, and
- $\sigma(y)$ of *PlayTennis* are $\sigma(No) = 0.36$ and $\sigma(Yes) = 0.64$.

The strength coefficients $\sigma(x, y)$ of *Wind* and *PlayTennis* are:

- $\sigma(Weak, No) = 0.14$,
- $\sigma(Strong, No) = 0.21$,
- $\sigma(Weak, Yes) = 0.43$,
- $\sigma(Strong, Yes) = 0.21$.

First, let us calculate the entropy of the flow graph G in Fig. 8(a) by using Definition 32.

$$\begin{aligned} H(G) &= -\sum_{y \in PlayTennis} \sigma(y) \log_2 \sigma(y) \\ &= -(0.36 \log_2 0.36 + 0.64 \log_2 0.64) \\ &= 0.94. \end{aligned}$$

[5] $Wind$ and $PlayTennis$ are not independent since $p(Wind = w, PlayTennis = t) \neq p(Wind = w) \times p(PlayTennis = t)$, c.f. [58] for dependency issue.

We calculate the certainty coefficients $cer(x, y)$ of $Wind$ and $PlayTennis$ as

- $cer(Weak, No) = \frac{\sigma(Weak,No)}{\sigma(Weak)} = 0.25$,
- $cer(Weak, Yes) = 0.75$,
- $cer(Strong, No) = 0.5$,
- $cer(Strong, Yes) = 0.5$.

According to Definition 34, conditional entropy of *PlayTennis* given *Wind* is

$$\begin{aligned} H(PlayTennis|Wind) &= -\sum_{x \in Wind} \sum_{y \in PlayTennis} \sigma(x, y) \log_2 cer(x, y) \\ &= -(0.14 \log_2 0.25 + 0.21 \log_2 0.5 \\ &\quad +0.43 \log_2 0.75 + 0.21 \log_2 0.5) \\ &= 0.89. \end{aligned}$$

Thus, the information gain of attribute *Wind*, relative to the flow graph G given in Fig. 8(a) is

$$\begin{aligned} Gain(G, Wind) &= H(G) - H(PlayTennis|Wind) \\ &= 0.94 - 0.89 \\ &= 0.05. \end{aligned}$$

Similarly, the information gain of *Humidity*, relative to a flow graph G in Fig. 8(b) is $Gain(G, Humidity) = 0.15$

According to the discussion provided in [62], our definition of information gain of a particular attribute relative to a flow graph G is similar to the information gain they used to build decision trees. $Gain(G, Wind)$ is less than $Gain(G, Humidity)$ indicates condition attribute $Wind$ has less predictive power than $Humidity$. In other words, knowing the $Humidity$ value helps to predict outcome better than knowing the $Wind$ value. This is thus referring to the quality of the flow graph in Fig. 8(a) as a classifier is not as predictive as the flow graph in Fig. 8(b).

An Application to Decision Trees

Entropy in machine learning has been successful in computing information gain in decision trees [62]. In further discussion, we adopt standard terminology concerning decision trees like root, branches, etc. Starting from a decision tree, it can be constructed as a unique flow graph by removing the root while its nodes are labeled by the same attribute [68]. Theoretically, Butz et al. showed that a flow graph is a special case of chain Bayesian network [58]. On

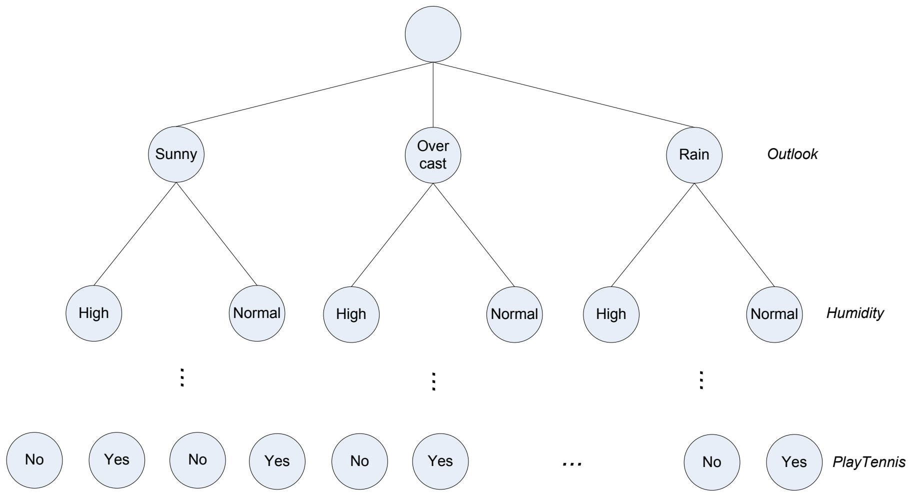


Fig. 9. Predictive decision tree from weather data set.

the contrary, in some situations, such as voting analysis and supply–demand problems (discussed in [65]), the available format is not a decision table or a database but is a flow graph. Hence, a classifier system constructed directly from a flow graph instead of a decision table is required. It is shown that classification of objects in flow graph representation boils down to finding the maximal output flow [65]. For these reasons, flow graph information gain is applied to solve this problem.

In this section, a new construction for decision trees from a flow graph is established. It is not sufficient to just precede in the inverse order as a (unique) flow graph construction from a decision tree, since a flow graph can be rearranged its layer order. Thus, it can construct several decision trees with distinct predictabilities.

Example 2 We are given the initial flow graph in Fig. 7 and we aim to construct a decision tree classifier. We can construct the decision tree by adding its root. All corresponding nodes and branches are inherited from the flow graph. Although, a decision tree constituted directly from this graph will have the level of nodes as appeared in the flow graph which has less predictive performance as discussed in [62]. Alternatively, from Section 4, the information gain (predictive) order of condtional attributes is *Outlook*, *Humidity*, *Wind* and *Temperature*. Then we can generate a decision tree classifier (Fig. 9). Its levels are determined according to their flow graph information gains. As one can see, this flow graph can be used as a classifier for *PlayTennis*. Hence, a more predictive decision tree can be constructed by using the proposed flow graph entropy and information gain.

Concluding Remarks

We propose flow graphs' analysis based on entropy computation. We have shown a new mathematical relationship between flow graphs and entropy, which can be used for data analysis. In particular, an information gain derived from entropy is suitable for creating and analyzing decision trees from flow graphs when starting from a specified format. The entropy of flow graphs may have applications not necessarily associated with decision trees, but these require further study. Future works are an exploration of such problems and a more formal scheme for decision tree generation.

References

[1] Z. Pawlak, Rough sets, International Journal of Computer & Information Sciences 11 (5) (1982) 341–356.

[2] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.

[3] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–356.

[4] L.A. Zadeh, Calculus of fuzzy restrictions, in: L.A. Zadeh et al. (Eds.), Fuzzy Sets and Their Applications to Cognitive and Decision Process, Academic Press, London (1975) 1–39.

[5] G.J. Klir, U.H. St. Clair, B. Yuan, Fuzzy Set Theory Foundations and Applications, Prentice-Hall, London, 1997.

[6] D. Dubois, H. Prade, Possibillity Theory an Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.

[7] J. Komorowski, L. Polkowski, A. Skowron, Rough sets: a tutorial, in: S.K. Pal, A., Showorn (Eds.), Rough Fuzzy Hybridization: A New Trend in Decision-Making, Springer, Berlin (1999) 3–98.

[8] Z. Pawlak, A treatise on rough sets, in: J.F. Peters, A. Skowron (Eds.), Transactions on Rough Sets IV, LNCS 3700, Springer-Verlag, Berlin, Heidelberg (2005) 1–17.

[9] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.

[10] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (1) (2007) 28–40.

[11] Z. Pawlak, Some issues on rough sets, in: J.F. Peters, A. Skowron, J.W. Grzymala-Busse, B. Kostek, R.W. Swiniarski, M.S. Szczuka (Eds.), Transactions on Rough Sets I, LNCS 3100, Springer-Verlag, Berlin, Heidelberg (2004) 1–58.

[12] X. Hu, N. Cercone, Data mining via discretization, generalization and rough set feature selection, Knowledge and Information Systems 1 (1) (1999) 33–60.

[13] A. An, N. Cercone, Rule quality measures improve the accuracy of rule induction: an experimental approach computational intelligence, Computational Intelligence 17 (3) (2001) 409–424.

[14] W. Zirako, Variable precision rough set model, Journal of Computer and System Sciences 46 (1) (1993) 39–59.

[15] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (1) (2001) 87–104.

[16] P. Pattaraintakorn, N., Cercone, K., Naruedomkul, Rule learning: ordinal prediction based on rough set and soft–computing, Applied Mathematics Letters 19 (12) (2006) 1300-1307.

[17] P. Pattaraintakorn, N., Cercone, K., Naruedomkul, Selecting attributes for soft-computing analysis in hybrid intelligent systems, in: D. Slezak, J.T. Yao, J.F. Peters, W. Ziarko, X. Hu (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Part II, LNAI 3642, Springer-Verlag, Heidelberg (2005) 698-708.

[18] J. Li, P. Pattaraintakorn, N. Cercone, Rule evaluations, attributes, and rough sets: extension and a case study, in: J.F. Peters, A. Skowron, I. Duntsch, J. Grzymala-Busse, E. Orlowska, L. Polkowski (Eds.), Transactions on Rough Sets VI: Commemorating Life and Work of Zdislaw Pawlak, Part I, LNCS 4374, Springer-Verlag, Berlin, Heidelberg (2007) 152–171.

[19] L.R. Medsker, Hybrid Intelligent System, Kluwer Academic Publishers, Dordrecht, 1995.

[20] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–208.

[21] M. Banerjee, S.K. Pal, Roughness of a fuzzy set, Information Sciences 93 (1996) 235–246.

[22] S. Bodjanova, Alpha-bounds of fuzzy numbers, Information Sciences 152 (2003) 237–266.

[23] V. Huynh, Y. Nakamori, A roughness measure for fuzzy sets, Information Sciences 173 (2005) 255–275.

[24] Y. Yang, R. John, Roughness bounds in rough set operations, Information Sciences 176 (2006) 3256–3267.

[25] H. Zhang, H. Liang, D. Liu, Two new operations in rough set theory with applications to fuzzy sets, Information Sciences 166 (2004) 147–165.

[26] D.V. Janardhan Rao, M. Banerjee, P. Mitra, Object extraction in gray-scale images by optimizing roughness measure of a fuzzy set, in: S.K. Pal et al. (Eds.), Pattern Recognition and Machine Intelligence, LNCS 3776, Springer-Verlag, Heidelberg (2005) 744–749.

[27] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (1) (2001) 87–104.

[28] J. Wang, S. Liu, J. Zhang, Roughness of a vague set, International Journal of Computational Cognition 3 (3) (2005) 83–87.

[29] E. Marczewski, H. Steinhaus, On a certain distance of sets and the corresponding distance of functions, Colloquium Mathematicum 6 (1958) 319–327.

[30] Pawlak, Z.: On rough derivatives, rough integrals and rough differential equations. Tech. Rep. 41/95, Institute of Computer Science, Warsaw Institute of Technology (1995)

[31] Nguyen, H.S., Jankowski, A., Peters, J.F., Skowron, A., Szczuka, M., Stepaniuk, J.: Discovery of process models from data and domain knowledge: A rough-granular approach. In: Yao, J.T. (ed.) Novel Developements in Granular Computing: Applications of Advanced Human Reasoning and Soft Computation. Information Science Reference, Hersey (2009) (in press)

[32] Peters, J.F., Han, L., Ramanna, S.: The Choquet integral in a rough software cost decision system. In: Grabisch,M., Murofushi, T., Sugeno, M., Kacprzyk, J. (eds.) Fuzzy Measures and Integrals. Theory and Applications, pp. 392414. Physica Verlag, Heidelberg (2000)

[33] Grabisch, M.: Modelling data by the Choquet integral. In: Torra, V. (ed.) Information Fusion in Data Mining, pp. 135148. Physica Verlag, Heidelberg (2001)

[34] Grabisch, M., Murofushi, T., Sugeno, M., Kacprzyk, J.: Fuzzy Measures and Integrals. Theory and Applications. Physica Verlag, Heidelberg (2000)

[35] Lehrer, E.: A new integral for capacities. Tech. rep., School of Mathematical Sciences, Tel Aviv University (2005)

[36] Halmos, P.R.: Measure Theory. Van Nostrand and Co., London (1950)

[37] Pawlak, Z., Peters, J.F., Skowron, A., Suraj, Z., Ramanna, S., Borkowski, M.: Rough Measures, Rough Integrals and Sensor Fusion. In: Inuiguchi, M., Tsumoto, S., Hirano, S. (Eds.): Rough Set Theory and Granular Computing. Spriger-Verlag, Berlin, Heidelberg (2003) 263–272

[38] Pawlak, Z., Skowron, A.: Rough Membership Functions. In: Yager, Y., Fedrizzi, M., Kacprzyk, J. (Eds.): Advances in the Dempster-Shafer Theory of Evidence. John Wiley & Sons, New York (1994), 251–271

[39] Pawlak, Z., Peters, J.F., Skowron, A., Suraj, Z., Ramanna, S., Borkowski, M.: Rough Measures and Integrals: A Brief Introduction. In: Terano, T. et al. (Eds.): New Frontiers in Artificial Intelligence, Joint JSAI 2001 Workshop Post-Proceedings, LNAI 2253, Springer-Verlag, Berlin, Heidelberg (2001) 375–379

[40] Choquet, G.: Theory of Capacities. Annales de l'Institut Fourier 5, (1953) 131-295

[41] Pawlak, Z., Peters J.F., Skowron, A., Surai, Z., Ramanna, S., Borkowski, M.: Rough Measures: Theory and Applications. In: Hirano, S., Inuiguchi, M., Tsumoto, S. (Eds.): Bulletin of the International Rough Set Society, Vol. 5 (1/2) (2001) 177-184

[42] Peters, J.F., Ramanna, S., Skowron, A., Borkowski, M.: Wireless Agent Guidance of Remote Mobile Robots: Rough Integral Approach to Sensor Signal Analysis. In Zhong, N. et. al. (Eds.): WI 2001, LNAI 2198, Springer-Verlag, Berlin, Heidelberg (2001) 413–422

[43] Pawlak, Z., Peters, J., Skowron, A.: Approximating functions using rough sets. In: Proceedings of the 2004 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 04), vol. 2, pp. 785790 (2004)

[44] Butz, C.J., Yen, W., Yang, B.: The Computational Complexity of Inference Using Rough Set Flow Graphs. In: Slezak, D. et al. (eds), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. LNAI, vol. 3641, pp. 335–344. Springer–Verlag, Berlin Heidelberg (2005)

[45] Czyzewski, A., Kostek, B.: Musical Metadata Retrieval with Flow Graphs. In: Tsumoto, S. et al. (eds.), Rough Sets and Current Trends in Computing. LNAI, vol. 3066, pp. 691–698. Springer–Verlag, Berlin Heidelberg (2004)

[46] Chitcharone, D., Pattaraintakorn, P.: Knowledge Discovery by Rough Sets Mathematical Flow Graphs and its Extension. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, pp. 340-345. (2008)

[47] Chitcharone, D., Pattaraintakorn, P.: Towards Theories of Fuzzy Set and Rough Set to Flow Graphs. In: the 2008 IEEE World Congress on Computational Intelligence, pp. 1675–1682. IEEE Press, Hong Kong (2008)

[48] Liu, H., Sun, J., Zhang, H.: Interpretation of Extended Pawlak Flow Graphs Using Granular Computing, In: Peters, J.F., Skowron, A. (eds.), T. Rough Sets VIII. LNCS, vol. 5084, pp. 93–115. Springer–Verlag, Heidelberg (2008)

[49] Matusiewicz, Z., Pancerz, K.: Rough Set Flow Graphs and Max $-*$ Fuzzy Relation Equations in State Prediction Problems. In: Chan, C.C. et al. (eds), Rough Sets and Current Trends in Computing. LNAI, vol. 5306, pp. 359–368. Springer-Verlag, Berlin (2008)

[50] Pattaraintakorn, P., Cercone, N., Naruedomkul, K.: Rule Learning: Ordinal Prediction Based on Rough Set and Soft-Computing. Appl. Math. Lett. 19(12), 1300-1307 (2006)

[51] Pattaraintakorn, P.: Entropy Measures of Flow Graphs with Applications to Decision Trees. In: Wen, P. et al. (eds.), Rough Sets and Knowledge Technology. LNCS, vol. 5589, pp. 618–625. Springer–Verlag, Berlin (2009)

[52] Pawlak, Z.: Rough Sets, Decision Algorithms and Bayes' Theorem. European J. of Oper. Res. 136, 181–189 (2002)

[53] Pawlak, Z.: Rough Set and Flow Graphs. In: Slezak, D. et al. (eds), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. LNAI, vol. 3641, pp. 1–11. Springer–Verlag, Berlin Heidelberg (2005)

[54] Pawlak, Z.: Inference Rules and Decision Rules. In: Rutkowski, L. et al. (eds.) International Conference on Artificial Intelligence and Soft Computing. LNAI, vol. 3070, pp. 102–108. Springer–Verlag, Berlin Heidelberg (2004)

[55] Pawlak, Z.: Decision Trees and Flow Graphs. In Greco, S. et al. (eds.) Rough Sets and Current Trends in Computing. LNAI, vol. 4259, pp. 1-11. Springer–Verlag, Heidelberg (2006)

[56] Pawlak, Z.: Flow Graphs and Data Mining. In: Peters, J.F., Skowron, A. (eds.) Transcations on Rough Sets III. LNCS, vol. 3400, pp. 1–36. Springer–Verlag, Berlin Heidelberg (2005)

[57] Pawlak, Z., Skowron, A.: Rudiments of Rough Sets. Inform. Sciences. 177, 3–20 (2007)

[58] Butz, C.J., Yan, W., Yang, B.: An Efficient Algorithm for Inference in Rough Set Flow Graphs. In: Peters, J.F. et al. (eds.) T. Rough Sets V. LNAI, vol. 4100, pp. 102–122. Springer-Verlag, Heidelberg (2006)

[59] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Inc., New York (1991)

[60] Chitcharone, D., Pattaraintakorn, P.: Towards Theories of Fuzzy Set and Rough Set to Flow Graphs. In: the 2008 IEEE World Congress on Computational Intelligence, pp. 1675-1682. IEEE Press, Hong Kong (2008)

[61] Kostek, B., Czyzewski, A.: Processing of Musical Metadata Employing Pawlak's Flow Graphs. In: Peters, J.F. et al. (eds.) T. on Rough Sets I. LNAI, vol. 3100, pp. 279–298. Springer-Verlag, Heidelberg (2004)

[62] Mitchell, T.M.: Machine Learning. McGraw-Hill Higher Education (1997)

[63] Pattaraintakorn, P., Cercone, N., Naruedomkul, K.: Rule Learning: Ordinal Prediction Based on Rough Set and Soft-Computing. Appl. Math. Lett. 19(12), 1300–1307 (2006)

[64] Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)

[65] Pawlak, Z.: In Pursuit of Patterns in Data Reasoning from Data–the Rough Set Way. In Alpigini, J.J. et al. (eds.) Rough Sets and Current Trends in Computing. LNAI, vol. 2475, pp. 1–9. Springer-Verlag, Heidelberg (2002)

[66] Pawlak, Z.: Decision Algorithms and Flow Graphs: a Rough Set Approach. J. of Telecom. and Inform. Tech. 3, 98–101 (2003)

[67] Pawlak, Z.: Some Remarks on Conflict Analysis. European J. of Oper. Res. 166, 649–654 (2005)

[68] Pawlak, Z.: Decision Trees and Flow Graphs. In Greco, S. et al. (eds.) Rough Sets and Current Trends in Computing. LNAI, vol. 4259, 1–11. Springer-Verlag, Heidelberg (2006)

[69] Shannon, C.E.: A Mathematical Theory of Communication. AT&T Tech. J. 27, 379-423 and 623-656 (1948)

5. ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

ผู้วิจัยได้ค้นพบสมบัติทางคณิตศาสตร์ที่สำคัญของ roughness measure รวมถึงคิดค้นและพิสูจน์ขอบเขตบนและล่างของ roughness measure of fuzzy sets สำหรับยูเนียนและอินเตอร์เซกชันสำหรับสองเซต ถือเป็นเพียงการพิสูจน์บางส่วนเพื่อระบุขอบเขตการดำเนินการทางคณิตศาสตร์ของ roughness measure of fuzzy sets เท่านั้น และยังต้องทำการศึกษาและพิสูจน์สำหรับรูปแบบทั่วไปมากขึ้นต่อไปตามลำดับ และพัฒนาเพื่อประยุกต์ใช้ roughness measure และขอบเขตที่ได้กับข้อมูลต่าง ๆ ต่อไป

จากผลการทดลองของ DRI ผู้วิจัยสังเกตข้อบ่งชี้ที่สำคัญอีก 2 ประการ นั่นคือ เมื่อพบตำแหน่งในกราฟแสดงผลลัพธ์ที่มีการเปลี่ยนมุมหรือมีจุดหักเปลี่ยนค่า DRI เมื่อเปรียบเทียบกับค่าฟังก์ชันแล้ว ควรเลือกค่าขอบเขตของปริพันธ์ในการทดลองที่ต่างออกไปใหม่ และจากความถูกต้อง ประสิทธิภาพ ความรวดเร็ว และความยืดหยุ่นที่สามารถเลือกฟังก์ชัน ช่วงและ class ที่ต้องการได้ จึงควรต่อยอดนำ DRI ไปวิเคราะห์ข้อมูลที่เก็บจากตัวเซนเซอร์ทางวิศวกรรมศาสตร์ต่อไป

ในแง่ของกราฟสายงาน นิยาม สูตร และระเบียบวิธีจากผลการวิจัยที่ได้นั้นคิดค้นสำหรับโหนดที่ประชิดกัน จึงควรต่อยอดสำหรับกรณีทั่วไป และสำหรับวิถีและการเชื่อมต่อด้วย นอกจากนี้ เอนโทรปี information gain, categories of vagueness, accuracy of approximation, roughness approximation และ dependency degree ที่ได้เสนอขึ้นอาจมีประโยชน์ในแง่อื่น ๆ ที่ต้องศึกษาต่อไป รวมถึงการพัฒนาระเบียบวิธีการแปลงกราฟสายงานเป็นต้นไม้การตัดสินใจโดยพิจารณาวิถีและการเชื่อมต่อ เพื่อยกระดับต้นไม้การตัดสินใจผลลัพธ์ให้มีความถูกต้องและสร้างได้รวดเร็วมากยิ่งขึ้น

Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (ระบุชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร ปี เล่มที่ เลขที่ และหน้า) หรือผลงานตามที่คาดไว้ในสัญญาโครงการ

    1. Pattaraintakorn, P., Naruedomkul, K., Palasit, K., *A note on roughness measure of fuzzy set,* Applied Mathematics Letters **2009**; 22: 1170-1173.
    2. Pattaraintakorn, P., *Entropy measures of flow graphs with applications to decision trees,* Lecture Notes in Artificial Intelligence **2009**; 5589: 618-625.
    3. Pattaraintakorn, P., Peters, J.F., Ramanna, S., *Capacity-based definite rough integral and its application,* Advances in Intelligent and Soft Computing **2009**; 59: 299-308.
    4. Chitcharone, D., Pattaraintakorn, P., *An Extension of Rough Set Approximation to Flow Graph Based Data Analysis*. Lecture Notes in Artificial Intelligence **2010**; 6086: 418-427.

2. การนำผลงานวิจัยไปใช้ประโยชน์

    - เชิงพาณิชย์ (มีการนำไปผลิต/ขาย/ก่อให้เกิดรายได้ หรือมีการนำไปประยุกต์ใช้โดยภาคธุรกิจ/บุคคลทั่วไป)

    สำหรับผลงาน Pattaraintakorn, P., Naruedomkul, K., Palasit, K., *A note on roughness measure of fuzzy set,* Applied Mathematics Letters **2009**; 22: 1170-1173. ได้มีวารสารวิชาการที่อ้างอิงบทความนี้ใน Scopus แล้วคือ

    Tripathy, B.C., Baruah, A., Nörlund and Riesz mean of sequences of fuzzy real numbers, Applied Mathematics Letters **2009**; 23 (5): 651-655.

    นอกจากนี้บทความนี้ยังติดอยู่ใน ScienceDirect Top 25 Hottest Articles ช่วงเดือน Apr - Jun 2009 อีกด้วย

    - เชิงสาธารณะ (มีเครือข่ายความร่วมมือ/สร้างกระแสความสนใจในวงกว้าง)

ร่วมวิจัยกับนักวิจัยที่ปรึกษา รศ.ดร.กัลยา นฤดมกุล และนักวิจัยจากประเทศแคนาดานั่นคือ Prof. Dr. James F. Peters และ Prof. Dr. Sheela Ramanna

- เชิงวิชาการ (มีการพัฒนาการเรียนการสอน/สร้างนักวิจัยใหม่)

ได้ฝึกฝนนักวิจัยใหม่คือ น.ส. ดวงรัตน์ จิตต์เจริญ

3. อื่นๆ (เช่น ผลงานตีพิมพ์ในวารสารวิชาการในประเทศ การเสนอผลงานในที่ประชุมวิชาการ หนังสือ การจดสิทธิบัตร)

ผลงานวิจัยในหัวข้อ 1. ได้ถูกนำเสนอในที่ประชุมวิชาการระดับนานาชาติดังต่อไปนี้

1. Pattaraintakorn, P., "Entropy Measures of Flow Graphs with Applications to Decision Trees". The Fourth International Conference on Rough Set and Knowledge Technology (RSKT 2009), July 14-16, Gold Coast, Australia, 618-625.
2. Pattaraintakorn, P., Peters, J.F., Ramanna, S., "Capacity-Based Definite Rough Integral and Its Application". International Conference on Man-Machine Interactions (ICMMI 2009), September 25-27, Kocierz Pass, Poland, 299-308.
3. Chitcharone, D., Pattaraintakorn, P., "An Extension of Rough Set Approximation to Flow Graph Based Data Analysis." The Seventh International Conference on Rough Sets and Current Trends in Computing (RSCTC 2010), June 28-30, Warsaw, Poland, 418-427.

ภาคผนวก

Applied Mathematics Letters 22 (2009) 1170–1173

Contents lists available at ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

A note on the roughness measure of fuzzy sets

P. Pattaraintakorn [a,*], K. Naruedomkul [b], K. Palasit [a]

[a] Department of Mathematics and Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand

[b] Department of Mathematics, Mahidol University, Bangkok, 10400, Thailand

ARTICLE INFO

Article history:
Received 1 June 2008
Accepted 12 September 2008

Keywords:
Fuzzy set theory
Roughness measure
Rough set theory

ABSTRACT

In this work, we study the roughness measure of fuzzy sets. New properties and roughness bounds for fuzzy set operations are established. Knowing these bounds of the operations results helps one to avoid unnecessary space in computation.

1. Introduction

Rough set theory was invented by Pawlak [1] and fuzzy set theory by Zadeh [2]. The given set, which is the target of the study, typically contains vagueness and data uncertainty that all require adjustments. In order to accommodate such difficulties, approximation is required and thus rough sets and fuzzy sets are expedient. Rough sets approximation is carried out in terms of two sets: the lower and upper approximations [1]. Alternatively, fuzzy sets approximation is carried out using a membership function [2,3]. According to Pawlak [1], *"A rough set represents a new mathematical approach to vagueness and uncertainty, the emphasis being on the discovery of patterns in data"*. The rough sets approach is viewed as a soft, rather than hard, computing technique. Thus, rough sets approaches are effective in the fields of data analysis, machine learning, information retrieval, survival analysis [4–8]. Fuzzy set theory has attracted researchers and played an important role in natural language, image processing, robotics, etc. [3]. Dubois et al. proposed rough fuzzy sets and fuzzy rough sets, and then established theoretical rough sets and fuzzy sets integrations [9]. The study by Banerjee et al. strengthens the connections between rough sets and fuzzy sets with the roughness measure of fuzzy sets [10]. Huynh et al. introduced a new roughness measure of fuzzy sets based on mass assignment [11]. The roughness of a fuzzy set was then interpreted as the weighted sum of the roughness measures of nested focal subsets. Yang et al. have investigated roughness bounds under different set operations for rough sets [12].

Not only has the theoretical development expanded, but practical applications were studied as well. Banerjee et al. and Zhang et al. found measures of the roughness of fuzzy sets useful for applications in pattern recognition [10,13]. The authors proposed using rough fuzzy sets as the model for images [14]. The results show that objects are extracted with higher accuracy using their approach, compared to Shannon's probabilistic entropy. In the work of Huynh et al. [11], a new roughness measure was used to analyze relational databases. Yang et al. reported that roughness measures are important indicators for decision making applications [12]. When we focus on data mining, machine learning, bioinformatics, network security, natural language processing, etc., data sets are usually huge. Thus, we investigate bounds on roughness measures for the fuzzy set operations. This work is organized as follows. Sections 2 and 3 contain requisite notions of rough sets, fuzzy sets and roughness measures. The theoretical background for important operators and new bounds on the roughness measure for fuzzy set operations are provided in Sections 4 and 5.

* Corresponding author. Tel.: +66 011 662 326 4341; fax: +66 011 662 326 4354.
E-mail address: kppuntip@kmitl.ac.th (P. Pattaraintakorn).

0893-9659/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2008.09.002

2. Rough sets and fuzzy sets

Definition 1 (*[1]*). Let U be a non-empty finite set called the universe and R be an equivalence relation on U. We call $\langle U, R\rangle$ an approximation space.

Definition 2 (*[1]*). Let $\langle U, R\rangle$ be an approximation space, $A \subseteq U$ and $X_1, X_2, \ldots, X_n$ denote the equivalence classes in U with respect to R. The lower approximation $\underline{A}$, upper approximation $\overline{A}$ and boundary region BN_R are defined as $\underline{A} = \bigcup\{X_i : X_i \subseteq A\}$, $\overline{A} = \bigcup\{X_i : X_i \cap A \neq \emptyset\}$, and $BN_R = \overline{A} - \underline{A}$ where $i \in \{1, 2, \ldots, n\}$, respectively.

Definition 3 (*[1]*). Let $\langle U, R\rangle$ be an approximation space and $A \subseteq U$. A measure of roughness of A in $\langle U, R\rangle$ is defined as $\rho_A = 1 - \frac{|\underline{A}|}{|\overline{A}|}$, where $|X|$ denotes the cardinality of the set X.

Yao [7] stated that the roughness measure can be understood as the distance between the lower and upper approximations. Next, let $\mathbf{A} : U \to [0, 1]$ be a fuzzy set in U, $\mathbf{A}(x)$, $x \in U$, giving the degree of membership of x in $\mathbf{A}$ [10].

Definition 4 (*[10]*). The lower and upper approximations of fuzzy set $\mathbf{A}$ in U, $\underline{\mathbf{A}}$ and $\overline{\mathbf{A}}$, are defined as fuzzy sets in $U/R \to [0, 1]$ and $\underline{\mathbf{A}}(X_i) = \inf_{x\in X_i} \mathbf{A}(x)$, $\overline{\mathbf{A}}(X_i) = \sup_{x\in X_i} \mathbf{A}(x)$, $i = 1, \ldots, n$, where inf (sup) denotes infimum (supremum).

Definition 5 (*[10]*). Fuzzy sets $\underline{\mathscr{A}}, \overline{\mathscr{A}} : U \to [0, 1]$ are defined as follows: $\underline{\mathscr{A}}(x) = \underline{\mathbf{A}}(X_i)$ and $\overline{\mathscr{A}}(x) = \overline{\mathbf{A}}(X_i)$, if $x \in X_i$, $i \in \{1, \ldots, n\}$.

3. Roughness measure of fuzzy sets

Definition 6 (*[10]*). Suppose α, β are two given parameters, where $0 < \beta \le \alpha \le 1$. The α-cut set and the β-cut set of fuzzy sets $\underline{\mathscr{A}}, \overline{\mathscr{A}}$ are defined as $\underline{\mathscr{A}}_\alpha = \{x : \underline{\mathscr{A}}(x) \ge \alpha\}$ and $\overline{\mathscr{A}}_\beta = \{x : \overline{\mathscr{A}}(x) \ge \beta\}$, respectively.

$\underline{\mathscr{A}}_\alpha$ and $\overline{\mathscr{A}}_\beta$ can be considered as the collection of objects with α as the minimum degree of definite membership, and β as the minimum degree of possible membership in the fuzzy set $\mathscr{A}$ [10].

Definition 7 (*[10]*). A roughness measure $\rho_{\mathbf{A}}^{\alpha,\beta}$ of fuzzy set $\mathbf{A}$ in U with respect to parameters α, β, where $0 < \beta \le \alpha \le 1$, and the approximation space $\langle U, R\rangle$, is defined as $\rho_{\mathbf{A}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|}$.

This roughness measure depends strongly on parameters α and β. There are several crucial properties of $\rho_{\mathbf{A}}^{\alpha,\beta}$ introduced in [10] as follows.

Proposition 1 (*[10]*). *Let α, β be two given parameters, where $0 < \beta \le \alpha \le 1$ and let $\underline{\mathscr{A}}_\alpha$ and $\overline{\mathscr{A}}_\beta$ be the α-cut set and β-cut set of fuzzy sets $\underline{\mathscr{A}}, \overline{\mathscr{A}}$; we have*

(a) $\overline{\mathscr{A} \cup \mathscr{B}}_\beta = \overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta$, (b) $\underline{\mathscr{A} \cap \mathscr{B}}_\alpha = \underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha$,

(c) $\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \subseteq \underline{\mathscr{A} \cup \mathscr{B}}_\alpha$, (d) $\overline{\mathscr{A} \cap \mathscr{B}}_\beta \subseteq \overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta$.

Property 1 (*[10]*). *For fuzzy sets* $\mathbf{A}$ *and* $\mathbf{B}$, *it holds that*

(a) $$\rho_{\mathbf{A}\cup\mathbf{B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cup \mathscr{B}}_\beta|} = 1 - \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|} \le 1 - \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|},$$

(b) $$\rho_{\mathbf{A}\cap\mathbf{B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A} \cap \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|} \le 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta|}.$$

4. Certain increment and uncertain decrement operators

The pioneering studies of fuzzy rough sets [1,10] derived Proposition 1(c) and (d). They perform *subset* instead of *set equal*. This cannot be analyzed quantitatively. Successive roughness measures also depend on the parameters α, β, and these two difficulties restrict some computations [13]. Thus, two new parameter-free operators of fuzzy sets were devised [13].

Definition 8 (*[13]*). Let U be the universe and let R be an equivalence relation on U. Let $X, Y \subseteq U$. When X is extended by Y (i.e., $X \cup Y$), $\underline{Z}_{(\cdot)}(\cdot) : U \times U \to U$, defined by $\underline{Z}_{(X)}(Y) = \bigcup\{[x]_R \mid x \in L(X), l_X(x) \not\subseteq Y \text{ and } h_X(x) \subseteq Y\}$, is called the certain increment operator of X, where $L(X) = \bigcup\{l_X(x) | x \in BN_R(X) \cap X\}$, $h_X(x) = [x]_R - X$, and $l_X(x) = [x]_R - h_X(x)$.

Definition 9 (*[13]*). Let U be the universe and let R be an equivalence relation on U. Let $X, Y \subseteq U$. When X is cut by Y (i.e., $X \cap Y$), $\overline{Z}_{(\cdot)}(\cdot) : U \times U \to U$, defined by $\overline{Z}_{(X)}(Y) = \bigcup\{[x]_R \mid x \in L(X), l_X(x) \cap Y = \emptyset \text{ and } h_X(x) \cap Y \neq \emptyset\}$, is called the uncertain decrement operator of X, where $L(X) = \bigcup\{l_X(x) | x \in BN_R(X) \cap X\}$, $h_X(x) = [x]_R - X$, and $l_X(x) = [x]_R - h_X(x)$.

Property 2 *([13]). For crisp sets $X, Y \subseteq U$, it holds that*

$$\text{(a)}\quad \underline{Z}_{(X)}(Y) = \underline{Z}_{(Y)}(X), \qquad \text{(b)}\quad \overline{Z}_{(X)}(Y) = \overline{Z}_{(Y)}(X).$$

Theorem 1 *([13]). Let $\alpha, \beta : U \to [0, 1]$ be two fuzzy sets in universe U and $0 < \beta \leq \alpha \leq 1$, where $\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)$, $\underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha)$, $\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)$ and $\overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta)$ are certain increment operators of $\mathscr{A}_\alpha$, $\mathscr{B}_\alpha$ and the uncertain decrement operators of $\mathscr{A}_\beta$, $\mathscr{B}_\beta$, respectively. We have*

$$\text{(a)}\quad \underline{\mathscr{A} \cup \mathscr{B}}_\alpha = \underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha) = \underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha),$$

$$\text{(b)}\quad \overline{\mathscr{A} \cap \mathscr{B}}_\beta = \overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta) = \overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta).$$

Property 3 *([13]). For fuzzy sets **A** and **B**, it holds that*

$$\text{(a)}\quad \rho_{\mathbf{A\cup B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha \cup \underline{Z}_{\mathscr{B}_\alpha}(\mathscr{A}_\alpha)|}{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta|},$$

$$\text{(b)}\quad \rho_{\mathbf{A\cap B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha \cap \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta \cap \overline{\mathscr{B}}_\beta - \overline{Z}_{\mathscr{B}_\beta}(\mathscr{A}_\beta)|}.$$

5. Some new bounds of roughness measures of fuzzy sets

The roughness measure is an important indicator of the uncertainty and accuracy associated with a given set [12]. Since data sets in most applications are usually huge, operations on these data sets are time- and space-consuming. Thus, before completing large volume operations involving two fuzzy sets, we must know the bounds of such results.

Theorem 2. *An upper bound of the roughness measure $\rho_{\mathbf{A\cup B}}^{\alpha,\beta}$ of fuzzy sets **A, B** in U with respect to α, β is given by $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \leq \frac{1-\rho_{\mathbf{A}}^{\alpha,\beta}\rho_{\mathbf{B}}^{\alpha,\beta}}{2-(\rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta})}$, where $0 < \beta \leq \alpha \leq 1$.*

Proof. From Property 1(a) and a basic set property, we have $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \leq 1 - \frac{\max\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}{|\overline{\mathscr{A}}_\beta| + |\overline{\mathscr{B}}_\beta|}$. If $|\underline{\mathscr{A}}_\alpha| \geq |\underline{\mathscr{B}}_\alpha|$ (or vice versa), then we obtain $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \leq 1 - \frac{1}{(|\overline{\mathscr{A}}_\beta|/|\underline{\mathscr{A}}_\alpha|)+(|\overline{\mathscr{B}}_\beta|/|\underline{\mathscr{A}}_\alpha|)}$. Thus, $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \leq \frac{1-\rho_{\mathbf{A}}^{\alpha,\beta}\rho_{\mathbf{B}}^{\alpha,\beta}}{2-(\rho_{\mathbf{A}}^{\alpha,\beta}+\rho_{\mathbf{B}}^{\alpha,\beta})}$ by Definition 7. □

Theorem 3. *An upper bound of the roughness measure $\rho_{\mathbf{A\cap B}}^{\alpha,\beta}$ of fuzzy sets **A, B** in U with respect to α, β is given by $\rho_{\mathbf{A\cap B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 + U^*$, where $0 < \beta \leq \alpha \leq 1$ and $U^* = \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$.*

Proof. From Property 1(b) and a basic set property, $\rho_{\mathbf{A\cap B}}^{\alpha,\beta} = 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|} + \frac{|\underline{\mathscr{A}}_\alpha \cup \underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$. With respect to Proposition 1(d), $\overline{\mathscr{A} \cap \mathscr{B}}_\beta \subseteq \overline{\mathscr{A}}_\beta$ (and $\subseteq \overline{\mathscr{B}}_\beta$) implies $|\overline{\mathscr{A} \cap \mathscr{B}}_\beta| \leq |\overline{\mathscr{A}}_\beta|$ (and $\leq |\overline{\mathscr{B}}_\beta|$); we therefore have $\frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} \leq \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$ and $\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|} \leq \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$. Then $\rho_{\mathbf{A\cap B}}^{\alpha,\beta} \leq 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|} + \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$ by Proposition 1(c). According to Definition 7, $\rho_{\mathbf{A\cap B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 + \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$. We finally have $\rho_{\mathbf{A\cap B}}^{\alpha,\beta} \leq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 + U^*$, where $U^* = \frac{|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|}{|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|}$. □

The bound in Theorem 2 depends on the roughness measures of fuzzy sets **A** and **B** while the bound in Theorem 3 depends on the roughness measures of fuzzy sets **A** and **B** in addition to $|\underline{\mathscr{A} \cup \mathscr{B}}_\alpha|$ and $|\overline{\mathscr{A} \cap \mathscr{B}}_\beta|$.

Theorem 4. *A lower bound on the roughness measure $\rho_{\mathbf{A\cup B}}^{\alpha,\beta}$ of fuzzy sets **A, B** in U with respect to α, β is given by $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 - L_*$, where $0 < \beta \leq \alpha \leq 1$ and $U_* = \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$.*

Proof. From Property 3(a) and a basic set property, we have that $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha| + |\underline{\mathscr{B}}_\alpha| + |\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$. For $|\overline{\mathscr{A}}_\beta| \geq |\overline{\mathscr{B}}_\beta|$, $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha| + |\underline{\mathscr{B}}_\alpha| + |\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|} = 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$. In accordance with Definition 7 and $\frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} \leq \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|}$, we have $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha|}{|\overline{\mathscr{A}}_\beta|} - \frac{|\underline{\mathscr{B}}_\alpha|}{|\overline{\mathscr{B}}_\beta|} - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|} = 1 - (1 - \rho_{\mathbf{A}}^{\alpha,\beta}) - (1 - \rho_{\mathbf{B}}^{\alpha,\beta}) - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$. Therefore, $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{A}}_\beta|}$. Similarly, for $|\overline{\mathscr{A}}_\beta| < |\overline{\mathscr{B}}_\beta|$, $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq 1 - \frac{|\underline{\mathscr{A}}_\alpha| + |\underline{\mathscr{B}}_\alpha| + |\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{B}}_\beta|}$. Therefore, $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 - \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{|\overline{\mathscr{B}}_\beta|}$. We finally have $\rho_{\mathbf{A\cup B}}^{\alpha,\beta} \geq \rho_{\mathbf{A}}^{\alpha,\beta} + \rho_{\mathbf{B}}^{\alpha,\beta} - 1 - L_*$, where $L_* = \frac{|\underline{Z}_{\mathscr{A}_\alpha}(\mathscr{B}_\alpha)|}{\max\{|\overline{\mathscr{A}}_\beta|, |\overline{\mathscr{B}}_\beta|\}}$. □

Theorem 5. *A lower bound of the roughness measure $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$ of fuzzy sets* **A, B** *in U with respect to α, β is given by* $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_*(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$, *where* $0 < \beta \leq \alpha \leq 1$ *and* $I_* = \frac{|\overline{\mathscr{A}}_\beta \cup \overline{\mathscr{B}}_\beta| + |\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{\min\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}$.

Proof. From Property 3(b), and basic set properties, we have $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{\min\{|\underline{\mathscr{A}}_\alpha|, |\underline{\mathscr{B}}_\alpha|\}}{|\overline{\mathscr{A}}_\beta|+|\overline{\mathscr{B}}_\beta|-|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|-|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}$. For $|\underline{\mathscr{A}}_\alpha| \leq |\underline{\mathscr{B}}_\alpha|$, we have that $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1}{\frac{|\overline{\mathscr{A}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}+\frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}-\frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}-\frac{|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}}$. From Definition 7 and $\frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|} \geq \frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{B}}_\alpha|}$, $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1}{\frac{|\overline{\mathscr{A}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}+\frac{|\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{B}}_\alpha|}-\frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|}{|\underline{\mathscr{A}}_\alpha|}-\frac{|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}} = 1 - \frac{1}{\frac{1}{1-\rho^{\alpha,\beta}_{\mathrm{A}}}+\frac{1}{1-\rho^{\alpha,\beta}_{\mathrm{B}}}-\frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}}$. We define $I_{|\underline{\mathscr{A}}_\alpha|} = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{A}}_\alpha|}$; thus $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_{|\underline{\mathscr{A}}_\alpha|}(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$. Similarly, for $|\underline{\mathscr{A}}_\alpha| > |\underline{\mathscr{B}}_\alpha|$, if we define $I_{|\underline{\mathscr{B}}_\alpha|} = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{|\underline{\mathscr{B}}_\alpha|}$, therefore we have $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_{|\underline{\mathscr{B}}_\alpha|}(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$. Thus, $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}} \geq 1 - \frac{1-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}+\rho^{\alpha,\beta}_{\mathbf{A}}\rho^{\alpha,\beta}_{\mathbf{B}}}{2-\rho^{\alpha,\beta}_{\mathbf{A}}-\rho^{\alpha,\beta}_{\mathbf{B}}-I_*(1-\rho^{\alpha,\beta}_{\mathbf{A}})(1-\rho^{\alpha,\beta}_{\mathbf{B}})}$, where $I_* = \frac{|\overline{\mathscr{A}}_\beta\cup\overline{\mathscr{B}}_\beta|+|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|}{\min\{|\underline{\mathscr{A}}_\alpha|,|\underline{\mathscr{B}}_\alpha|\}}$. □

The lower bounds for $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$ differ from the upper bounds on $\rho^{\alpha,\beta}_{\mathbf{A}\cap\mathbf{B}}$, in that they depend on the roughness measures of the fuzzy sets **A** and **B** and also $|\overline{\mathscr{A}}_\beta|$, $|\overline{\mathscr{B}}_\beta|$, $|\underline{\mathscr{A}}_\alpha|$, $|\underline{\mathscr{B}}_\alpha|$ and $|\overline{Z}_{\mathscr{A}_\beta}(\mathscr{B}_\beta)|$.

6. Concluding remarks

The roughness measure is an important indicator of uncertainty and accuracy associated with a given fuzzy set. More importantly, these roughness measures propagate through various set operations. Thus, before completing large volume operations involving two large fuzzy sets, we should have bounds on the roughness measure of the result, as proven in this work.

Acknowledgements

This research was supported by the grant MRG5180071 from the Thailand Research Fund and the Commission on Higher Education. Thanks are due to N. Cercone, G.M. Zaverucha and anonymous reviewers for their comments.

References

[1] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[2] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–356.
[3] G.J. Klir, U.H. St. Clair, B. Yuan, Fuzzy Set Theory Foundations and Applications, Prentice-Hall, London, 1997.
[4] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[5] X. Hu, N. Cercone, Data mining via discretization, generalization and rough set feature selection, Knowledge and Information Systems 1 (1) (1999) 33–60.
[6] A. An, N. Cercone, Rule quality measures improve the accuracy of rule induction: An experimental approach computational intelligence, Computational Intelligence 17 (3) (2001) 409–424.
[7] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (1) (2001) 87–104.
[8] P. Pattaraintakorn, N. Cercone, K. Naruedomkul, Rule learning: Ordinal prediction based on rough set and soft-computing, Applied Mathematics Letters 19 (12) (2006) 1300–1307.
[9] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–208.
[10] M. Banerjee, S.K. Pal, Roughness of a fuzzy set, Information Sciences 93 (1996) 235–246.
[11] V. Huynh, Y. Nakamori, A roughness measure for fuzzy sets, Information Sciences 173 (2005) 255–275.
[12] Y. Yang, R. John, Roughness bounds in rough set operations, Information Sciences 176 (2006) 3256–3267.
[13] H. Zhang, H. Liang, D. Liu, Two new operations in rough set theory with applications to fuzzy sets, Information Sciences 166 (2004) 147–165.
[14] D.V. Janardhan Rao, M. Banerjee, P. Mitra, Object extraction in gray-scale images by optimizing roughness measure of a fuzzy set, in: S.K. Pal, et al. (Eds.), Pattern Recognition and Machine Intelligence, in: LNCS, vol. 3776, Springer-Verlag, Heidelberg, 2005, pp. 744–749.

Capacity-Based Definite Rough Integral and Its Application

Puntip Pattaraintakorn, James F. Peters, and Sheela Ramanna

Abstract. This paper introduces an extension of the original capacity-based rough integral defined over a specific interval. The approach hearkens back to the pioneering work on capacities and the generalization of the Lebesgue integral by Gustav Choquet during the 1950s. Variations in the definition of the capacity function has led to various forms of the discrete Choquet integral. In particular, it is the rough capacity function (also called a rough membership function) introduced by Zdzisław Pawlak and Andrzej Skowron during the 1990s that led to the introduction of a capacity-based rough integral. By extension of the original work on the rough integral introduced in 2000, a discrete form of a capacity-based definite rough integral is introduced in this paper. This new form of the rough integral provides a means of measuring the relevance of functions representing features useful in the classification of sets of sample objects.

Keywords: capacity, Choquet integral, feature selection, rough set theory, rough membership function, definite rough integral.

1 Introduction

During the early 1990s, Zdzisław Pawlak introduced a discrete form of the definite Riemann integral of a continuous, real function defined over intervals representing

Puntip Pattaraintakorn
Department of Mathematics and Computer Science, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 10520
e-mail: kppuntip@kmitl.ac.th

James F. Peters
Department of Electrical and Computer Engineering,
University of Manitoba Winnipeg, Manitoba R3T 5V6 Canada
e-mail: jfpeters@ee.umanitoba.ca

Sheela Ramanna
Department of Applied Computer Science, University of Winnipeg,
Winnipeg, Manitoba R3B 2E9 Canada
e-mail: s.ramanna@uwinnipeg.ca

K.A. Cyran (Eds.): Man-Machine Interactions, AISC 59, pp. 299–308.
springerlink.com

equivalence classes [6]. Since that time, work on various forms of rough integrals has continued fairly steadily (see, e.g., [5, 8, 9, 10]). The original integral was called a *rough integral* because it was defined over intervals represented equivalence classes in a partition of sequences of reals defined by an equivalence relation.

Based on work on the Choquet integral $(C)\int f\ d\mu$ considered in the context of rough sets [12], a new form of discrete rough integral $\int f\ d\mu_x^B$ was introduced [8, 9] and elaborated in [10]. The Choquet integral is a generalization of the Lebesgue integral defined relative to a capacity μ [1, 2]. A *capacity* is a function μ that assigns a non-negative real number to every subset of a finite set X and satisfies $f(\emptyset) = 0$ [4]. When the discrete form of the Choquet integral $(C)\int f\ d\mu$ is defined relative to a finite universe, the Lebesgue integral reduces to a (convex) linear combination, where each individual integrand function value is weighted with a capacity function value [2].

This article introduces a new form of the Choquet integral called a *capacity-based definite rough integral* because its capacity is a function defined relative to equivalence classes. The extension of the capacity-based rough integral has a number applications. In particular, we show in this article how this integral can be used in feature selection with the DRI tool implemented in MATLAB®.

This article is organized as follows. A capacity-based rough integral is defined in Sect. 3. A discrete form of the definite rough integral (DRI) is defined in Sect. 4. An illustration of the application of the DRI is presented in Sect. 5.

2 Rough Capacity Function

Rough capacity functions were introduced during the mid-90s [11]. A rough capacity function returns the degree of overlap between a fixed set containing objects of interest and a set of sample objects.

Definition 1 (Rough Capacity Function). Let $S = (\mathscr{O}, \mathscr{F})$ denote an information system. Assume $X \subseteq \wp(\mathscr{O})$, $B \subseteq \mathscr{F}$, $x \in X$ and $[x]_B \subseteq X/\sim_B$. The capacity μ_x^B : $\wp(\mathscr{O}) \longrightarrow [0,1]$ is defined:

$$\mu_x^B(X) = \begin{cases} \frac{|X \cap [x]_B|}{|[x]_B|}, & \text{if } X \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The capacity μ_x^B is an example of a set function, i.e., a function whose domain is a family of sets [3]. This set function measures the degree of overlap between the set X and the class $[x]_B$, i.e., the extent that X is covered by $[x]_B$. Recall that $[x]_B$ is a set of objects having matching descriptions. This is important in evaluating the set X, since $\mu_x^B(X)$ is a measure of the extent that the objects in X are part of the classification represented by $[x]_B$. In the context of capacity-based integrals, the function μ_x^B is a source of weights, i.e., degree of importance of each set X in a weighted sum for the discrete form of the integral.

It can also be observed that μ_x^B is a real-valued set function that is additive. That is, for X, X' in $\wp(\mathscr{O})$, it can be shown that

$$\mu_x^B(X \cup X') = \mu_x^B(X) + \mu_x^B(X').$$

3 Capacity-Based Discrete Integrals

This section gives a brief introduction to the Choquet integral, which eventually led to the introduction of a capacity-based rough integral.

3.1 Discrete Choquet Integral

Recall that the Choquet integral $(C)\int f\, \mathrm{d}\mu$ is a generalization of the Lebesgue integral defined with respect to a non-classical measure μ called a capacity. Also recall that a capacity is a real-valued set function

$$\mu : \wp(X) \longrightarrow \Re^+,$$

such that $\mu(\emptyset) = 0$ and $X' \subset X'' \subset \wp(X)$ implies $\mu(X') \leq \mu(X'')$ (monotonicity). When the Choquet integral is defined relative to a finite sets, then the Choquet integral reduces to a weighted sum that has a variety of applications, especially in multi-criteria decision-making (see, e.g., [1, 2]).

Definition 2 (Discrete Choquet Integral [2]). Let μ be a capacity defined on a finite set X. The discrete Choquet integral of a function $f : X \longrightarrow \Re^+$ with respect to capacity μ is defined by

$$(C)\int f\, \mathrm{d}\mu = \sum_{i=1}^{n} \left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right) \right] \cdot \mu(X_{(i)}),$$

where $\cdot_{(i)}$ denotes a permuted index so that $0 \leq f\left(x_{(1)}\right) \leq f\left(x_{(2)}\right) \leq \cdots \leq f\left(x_{(i)}\right) \leq \cdots \leq f\left(x_{(n)}\right) \leq 1$. Also, $X_{(i)} = \left\{x_{(i)}, \ldots, x_{(n)}\right\}$, and $f(x_{(0)}) = 0$.

3.2 Discrete Rough Integral

The introduction of the rough capacity function μ_x^B paved the way for a discrete rough integral $(P)\int f\, \mathrm{d}\mu_x^B$ named after Zdzisław Pawlak. This rough integral is a variation of the discrete Choquet integral [1, 2, 7, 9, 10].

Definition 3 (Discrete Rough Integral). Let μ_x^B be a rough capacity function defined on a finite set X. The discrete rough integral of a function $f : X \rightarrow \Re^+$ with respect to capacity μ_x^B is defined by

$$(P)\int f d\mu_x^B = \sum_{i=1}^{n} \left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right) \right] \cdot \mu_x^B(X_{(i)}),$$

where $\cdot_{(i)}$ denotes a permuted index so that $0 \leq f\left(x_{(1)}\right) \leq f\left(x_{(2)}\right) \leq \cdots \leq f\left(x_{(i)}\right) \leq \cdots \leq f\left(x_{(n)}\right) \leq 1$. Also, $X_{(i)} = \left\{x_{(i)}, \ldots, x_{(n)}\right\}$, and $f(x_{(0)}) = 0$.

If f is non-negative, then $(P)\int f d\mu_x^B$ represents the lower approximation of the area under the graph of f.

Proposition 1. *If* $Min\mu \leq \mu_x^B(X_{(i)}) \leq Max\mu,\ 1 \leq i \leq n,$ *then* $0 \leq (P)\int f\ \mathrm{d}\mu_x^B \leq Max\mu$.

Proof

$$\begin{aligned}
(P)\int f\ \mathrm{d}\mu_x^B &= \sum_{i=1}^{n} \left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right) \right] \cdot \mu_x^B(X_{(i)}) \\
&\leq \sum_{i=1}^{n} \left[f\left(x_{(i)}\right) - f\left(x_{(i-1)}\right) \right] \cdot Max\mu \\
&= Max\mu \cdot \left[\left(f\left(x_{(1)}\right) - f\left(x_{(0)}\right) \right) + \left(f\left(x_{(2)}\right) - f\left(x_{(1)}\right) \right) + \right. \\
&\quad \left. \cdots + \left(f\left(x_{(n)}\right) - f\left(x_{(n-1)}\right) \right) \right] \\
&= Max\mu \cdot \left[-f\left(x_{(0)}\right) + f\left(x_{(n)}\right) \right] \\
&\leq Max\mu \qquad \text{(because } f(x_{(0)}) = 0 \text{ and } f\left(x_{(n)}\right) \leq 1\text{)}.
\end{aligned}$$

It can be proven in a straightforward way that $(P)\int f\ \mathrm{d}\mu_x^B \geq 0$. □

Consider a specialized capacity $\mu_x^{\{\phi\}}$ defined in terms of a single function $\phi \in B$, where B is a set of functions representing features of objects in a finite set X.

Proposition 2. *[9] Let* $0 < s \leq r$ *and* $\phi \in B$. *If* $f(x) \in [s, r]$ *for all* $x \in X$, *then* $(P)\int f\ \mathrm{d}\mu_x^{\{\phi\}} \in (0, r]$.

4 Definite Discrete Rough Integral

In this section, we introduce a discrete form of definite rough integral (DRI) of a function f denoted by $\int_a^b f\ \mathrm{d}\mu_x^B$, where $\cdot_{(i)}$ is a permuted index and a, b such that $x_{(1)} \leq a \leq b \leq x_{(n)}$ are the lower and upper integral limits, respectively. The limits on the rough integral specify the interval over which a function f is integrated and it is assumed that f is continuous over $[a, b]$. This integral is defined in terms of an upper integral $\overline{\int_a^b} f\ \mathrm{d}\mu_x^B$ and a lower integral $\underline{\int_a^b} f\ \mathrm{d}\mu_x^B$.

$$\int_a^b f\ \mathrm{d}\mu_x^B = \overline{\int_a^b} f\ \mathrm{d}\mu_x^B - \underline{\int_a^b} f\ \mathrm{d}\mu_x^B.$$

The discrete forms of the lower and upper integral are defined w.r.t. $[x_1]_B$ in (2) and (3), respectively.

$$\underline{\int_a^b} f \, \mathrm{d}\mu_{x_1}^B = \sum_{i=a}^{b} \left[f(x_{(i-1)}) \right] \Delta_{(i)} \mu_x^B, \tag{2}$$

$$\overline{\int_a^b} f \, \mathrm{d}\mu_{x_1}^B = \sum_{i=a}^{b} \left[f(x_{(i)}) \right] \Delta_{(i)} \mu_x^B, \tag{3}$$

where $\Delta_{(i)} \mu_x^B$ is defined in (4).

$$\Delta_{(i)} \mu_x^B = \left| \mu_x^B(X_{(i)}) - \mu_x^B(X_{(i-1)}) \right|. \tag{4}$$

Definition 4 (Definite Discrete Rough Integral). Let μ_x^B denote a rough capacity function with domain $X_{(i)}$ that is the set

$$X_{(i)} = \left\{ x_{(i)}, x_{(i+1)}, \ldots, x_{(n)} \right\},$$

where $\cdot_{(i)}$ is a permuted index, $X_{(0)} = \emptyset$, and a, b such that $x_{(1)} \leq a \leq b \leq x_{(n)}$ are the lower and upper integral limits, respectively. It is assumed that $f\left(x_{(0)}\right) = 0$. The difference $\Delta_{(i)} \mu_x^B$ is defined in (4) relative to the set $X_{(i)}$. The discrete definite rough integral of $f : X \to \Re^+$ is defined by

$$\int_a^b f \, \mathrm{d}\mu_x^B = \overline{\int_a^b} f \, \mathrm{d}\mu_x^B - \underline{\int_a^b} f \, \mathrm{d}\mu_x^B,$$

where the lower and upper integrals are defined in (2) and (3), respectively.

4.1 Interpretation

Observe that the capacity function μ_x^B is defined in terms of a set of functions B representing the features of sample objects of interest. The set B provides a basis for the relation $\sim_B$ that defines a partition of the sample objects X (source of integral limits). Then a function f is integrated with respect to μ_x^B. In the discrete form of the DRI, μ_x^B provides a weight on each summand $X_{\{i\}}$, a set of sample objects. The capacity μ_x^B computes the degree of overlap between a set X and a class $[x]_B$ representing objects that have been classified relative to the features represented by B (see Sect. 2). In effect, B is a source of criteria for grouping together objects matching the criteria represented by B. Hence, the definite rough integral indicates the importance and relevance of a function integrated with respect to μ_x^B. Hence, if a function ϕ representing an object feature is integrated with respect to μ_x^B, the DRI provides an effective means of selecting features that can be used to discriminate objects. In effect, the definite rough integral is useful for feature selection within the prescribed limits of the integral. The novelty here is that the limits on the DRI can be varied to measure varying importance of an individual feature represented by a function ϕ.

5 Feature Selection

In this section, we briefly illustrate an application of the discrete definite rough integral (DRI). For simplicity, we assume that each vector of function values used to describe a sample object is evaluated, e.g., acceptable ($d = 1$) vs. unacceptable ($d = 0$). Put $B = \{d\}$, where $d \in \{0,1\}$ in defining the capacity μ_x^B. Then any set of objects X can be partitioned using the set B.

5.1 Illustration

Next, consider, for example, a set of objects described with two functions, namely, ϕ_1, ϕ_2 representing features of objects in a sample X. For the purposes of illustration, we treat ϕ_1, ϕ_2 abstractly, i.e., without considering specific functions. Here is a partial, sample description table with the evaluation (decision) column d included. Note that $x_{(i)}$ indicates a permuted object with values in ascending order.

Table 1 Sample data

X	ϕ_1	ϕ_2	d	X	ϕ_1	ϕ_2	d
$x_{(1)} = x_4$	0.79224	29.988	0	$x_{(11)} = x_{11}$	0.92282	13.787	1
$x_{(2)} = x_3$	0.79467	30.114	0	$x_{(12)} = x_{12}$	0.93357	11.387	1
$x_{(3)} = x_2$	0.80596	27.402	0	$x_{(13)} = x_{13}$	0.94553	9.7302	1
$x_{(4)} = x_5$	0.81286	27.633	0	$x_{(14)} = x_{14}$	0.90996	13.979	1
$x_{(5)} = x_1$	0.85808	21.754	0	$x_{(15)} = x_{15}$	0.95608	7.1776	1
$x_{(6)} = x_6$	0.86020	22.866	0	$x_{(16)} = x_{16}$	0.94722	11.387	1
$x_{(7)} = x_7$	0.84569	24.43	0	$x_{(17)} = x_{17}$	0.94467	9.0222	1
$x_{(8)} = x_8$	0.87886	20.16	0	$x_{(18)} = x_{18}$	0.96424	6.3259	1
$x_{(9)} = x_9$	0.88235	19.945	0	$x_{(19)} = x_{19}$	0.92804	12.398	1
$x_{(10)} = x_{10}$	0.88094	19.227	0	$x_{(20)} = x_{20}$	0.93925	9.9671	1
				$x_{(21)} = x_{21}$	0.99435	2.3081	1

We now compute the DRI relative to $\left[x_{(1)}\right]_{\{d=0\}}$ starting with integration of ϕ_1 where $B = \{d = 0\}$. The lower integral $\underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \, d\mu_{x_1}^{\{d=0\}}$ is

$$\underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 d\mu_{x_1}^{\{d=0\}} = \phi_1(x_{(0)}) \cdot \Delta_{(1)}\mu_{x_1}^{\{d=0\}} + \phi_1(x_{(1)}) \cdot \Delta_{(2)}\mu_{x_1}^{\{d=0\}} + \cdots = 0.7,$$

and the upper integral $\overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}}$ is

$$\overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} = \phi_1(x_{(1)}) \cdot \Delta_{(1)}\mu_{x_1}^{\{d\}} + \phi_1(x_{(2)}) \cdot \Delta_{(2)}\mu_{x_1}^{\{d=0\}} + \cdots = 1.45.$$

Using these results, the definite integral $\int_{x_{(1)}}^{x_{(10)}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}}$ is

$$\int_{x_{(1)}}^{x_{(10)}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} = 0.75.$$

Next, integrate ϕ_2 with respect to $\mu_{x_1}^{\{d=0\}}$ using class $\left[x_{(1)}\right]_{\{d=0\}}$, and obtain

$$\int_{x_{(1)}}^{x_{(10)}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(10)}}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(10)}}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=0\}} = 2.3.$$

This means that for class $\left[x_{(1)}\right]_{\{d=0\}}$, the feature represented by ϕ_2 is more important than the feature represented by ϕ_1. The computations have been performed using the DRI tool (see Fig. 1) implemented in MATLAB ®. From the plots in Fig. 2, notice that there is less dispersion of the values for the plot for $\int_{x_{(1)}}^{x_{(10)}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=0\}}$, i.e., the values for the successive lower and upper values are tightly groups around the $\int_{x_{(1)}}^{x_{(10)}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=0\}}$ values compared with the $\int_{x_{(1)}}^{x_{(10)}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=0\}}$ values.

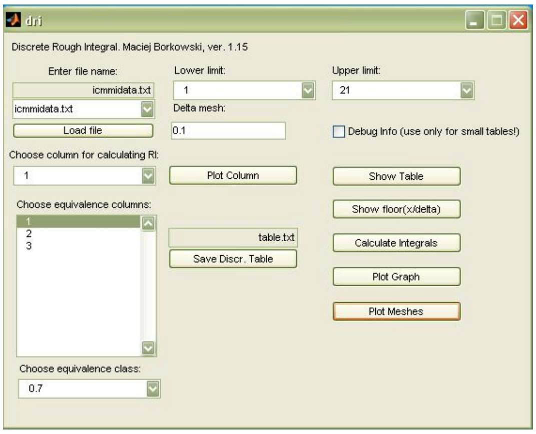


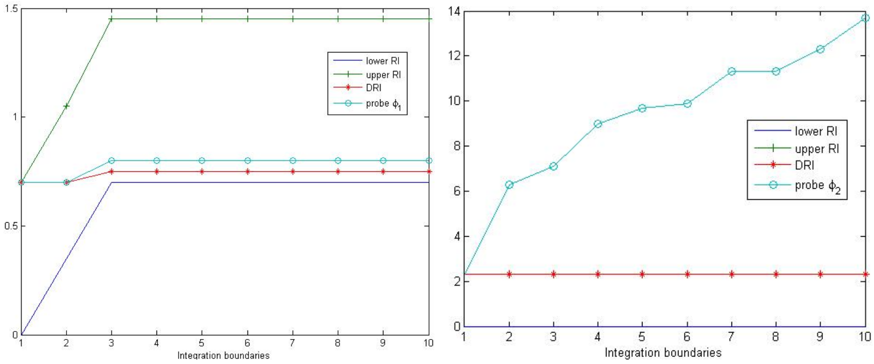
Fig. 1 DRI Tool interface

Similarly, consider the case for class $\left[x_{(1)}\right]_{\{d=1\}}$ and use the DRI to discover which function has greater importance, i.e., carries more weight. Then the DRI computed relative to ϕ_1 is

$$\int_{x_{(11)}}^{x_{(21)}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} = \overline{\int_{x_{(11)}}^{x_{(21)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} - \underline{\int_{x_{(11)}}^{x_{(21)}}} \phi_1 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} = 0.$$

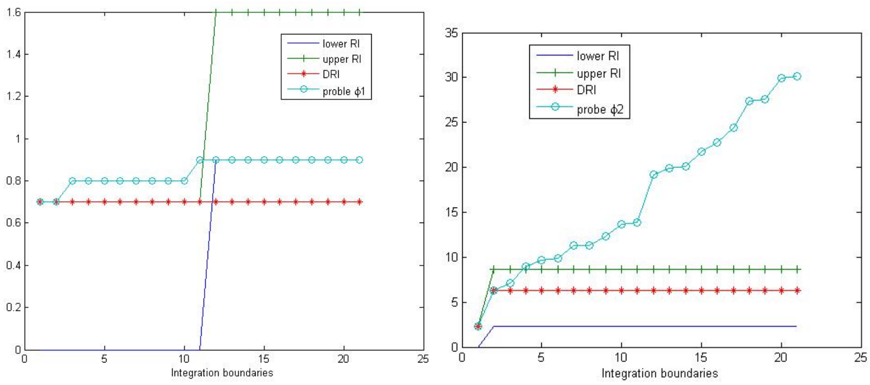
Next, integrate ϕ_2 for the same class, and obtain

$$\int_{x_{(11)}}^{x_{(21)}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} = \overline{\int_{x_{(11)}}^{x_{(21)}}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} - \underline{\int_{x_{(11)}}^{x_{(21)}}} \phi_2 \, \mathrm{d}\mu_{x_1}^{\{d=1\}} = 0.1.$$



(a) (b)

Fig. 2 DRI for ϕ_1, ϕ_2 for class $\left[x_{(1)}\right]_{\{d=0\}}$ **a** DRI $\phi_1, d = 0$. **b** DRI $\phi_2, d = 0$



(a) (b)

Fig. 3 DRI for ϕ_1, ϕ_2 for class $\left[x_{(2)}\right]_{\{d=1\}}$ **a** DRI $\phi_1, d = 1$. **b** DRI $\phi_2, d = 1$

We now calculate the DRI over the entire data set, for both classes $\left[x_{(2)}\right]_{\{d=0\}}$ and $\left[x_{(2)}\right]_{\{d=1\}}$ for the two features ϕ_1 and ϕ_2 respectively:

$$\int_{x_{(1)}}^{x_{(21)}} \phi_1 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(21)}}} \phi_1 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(21)}}} \phi_1 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} = 0.7.$$

$$\int_{x_{(1)}}^{x_{(21)}} \phi_2 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} = \overline{\int_{x_{(1)}}^{x_{(21)}}} \phi_2 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} - \underline{\int_{x_{(1)}}^{x_{(21)}}} \phi_2 \, \mathrm{d}\mu_{x_2}^{\{d=0\}} = 6.3.$$

From this, we can conclude that the feature represented by ϕ_2 is more important than the feature represented by ϕ_1 with respect to both classes and for different equivalence classes. The plots in Figs. 2 and 3 reveal an interesting feature of the sample objects that are a source of limits on the DRI, i.e., at $x_{(12)}$ there is a sharp change in the difference between ϕ_2 and DRI values (this is especially evident in the plot). This suggests a place to begin experimenting with different limits on the integral. This break in the DRI values also indicates a change in the evaluation of the functions representing object features.

6 Conclusion

This paper introduces the definite rough integral defined relative to a rough capacity function. It is the capacity μ_x^B that distinguishes the rough integral from the original capacity-based integral introduced by Choquet during the 1950s. In addition to the geometric interpretation of the definite rough integral already mentioned, this integral is highly significant because it offers a measure of the importance and relevance of features of sample objects. This measurement of the relevance of a feature is a byproduct of the rough capacity function that provides the basis for the rough integral. Now, with the introduction of rough integral limits, it is possible to measure the relevance of features relative to selected ranges of objects. In effect, the definite rough integral provides a new basis for feature selection in the classification of objects. Future work includes a study of the properties of the definite rough integral and its applications.

Acknowledgements. The authors gratefully acknowledge the helpful insights concerning topics in this paper by Andrzej Skowron and Maciej Borkowski for the DRI tool. This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grants 185986 and 194376, Canadian Arthritis Network grant SRI-BIO-05, Manitoba Hydro grant T277, Thailand Research Fund and the Commission on Higher Education grant MRG5180071.

References

1. Grabisch, M.: Modelling data by the Choquet integral. In: Torra, V. (ed.) Information Fusion in Data Mining, pp. 135–148. Physica Verlag, Heidelberg (2001)
2. Grabisch, M., Murofushi, T., Sugeno, M., Kacprzyk, J.: Fuzzy Measures and Integrals. Theory and Applications. Physica Verlag, Heidelberg (2000)
3. Halmos, P.R.: Measure Theory. Van Nostrand and Co., London (1950)
4. Lehrer, E.: A new integral for capacities. Tech. rep., School of Mathematical Sciences, Tel Aviv University (2005)
5. Nguyen, H.S., Jankowski, A., Peters, J.F., Skowron, A., Szczuka, M., Stepaniuk, J.: Discovery of process models from data and domain knowledge: A rough-granular approach. In: Yao, J.T. (ed.) Novel Developements in Granular Computing: Applications of Advanced Human Reasoning and Soft Computation. Information Science Reference, Hersey (2009) (in press)
6. Pawlak, Z.: On rough derivatives, rough integrals and rough differential equations. Tech. Rep. 41/95, Institute of Computer Science, Warsaw Institute of Technology (1995)

7. Pawlak, Z., Peters, J.F., Skowron, A.: Approximating functions using rough sets. In: Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society, vol. 2, pp. 785–790 (2004)
8. Pawlak, Z., Peters, J.F., Skowron, A., Suraj, Z., Ramanna, S., Borkowski, M.: Rough measures, rough integrals and sensor fusion. In: Inuiguchi, M., Tsumoto, S., Hirano, S. (eds.) Rough Set Theory and Granular Computing, pp. 263–272. Springer, Heidelberg (2000)
9. Pawlak, Z., Peters, J.F., Skowron, A., Suraj, Z., Ramanna, S., Borkowski, M.: Rough measures: theory and applications. Tech. Rep. 1/2, Bulletin of the International Rough Set Society (2001)
10. Pawlak, Z., Peters, J.F., Skowron, A., Suraj, Z., Ramanna, S., Borkowski, M.: Rough measures and integrals: a brief introduction. In: Terano, T., Nishida, T., Namatame, A., Tsumoto, S., Ohsawa, Y., Washio, T. (eds.) JSAI-WS 2001. LNCS (LNAI), vol. 2253, pp. 375–379. Springer, Heidelberg (2001)
11. Pawlak, Z., Skowron, A.: Rough membership functions. In: Yager, R., Fedrizzi, M., Kacprzyk, J. (eds.) Advances in Dempster-Shafer Theory of Evidence, pp. 251–271. John Wiley & Sons, New York (1994)
12. Peters, J.F., Han, L., Ramanna, S.: The Choquet integral in a rough software cost decision system. In: Grabisch, M., Murofushi, T., Sugeno, M., Kacprzyk, J. (eds.) Fuzzy Measures and Integrals. Theory and Applications, pp. 392–414. Physica Verlag, Heidelberg (2000)

Entropy Measures of Flow Graphs with Applications to Decision Trees

Puntip Pattaraintakorn

Department of Mathematics and Computer Science, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520 Thailand
kppuntip@kmitl.ac.th

Abstract. Entropy is a fundamental principle in many disciplines such as information theory, thermodynamics, and more recently, artificial intelligence. In this article, a measure of entropy on Pawlak's mathematical flow graph is introduced. The predictability and quality of a flow graph can be derived directly from the entropy. An application to decision tree generation from a flow graph is examined. In particular, entropy measures on flow graphs lead to a new methodology of reasoning from data and shows rigorous relationships between flow graphs, entropy and decision trees.

Keywords: Flow graphs, Entropy, Decision trees.

1 Introduction

Flow graphs, invented by Pawlak as an extention of *rough set theory* [7], model the information flow of a given data set [8,9,10,11]. When starting from a large data set (as in databases around the world), reasoning is referred to as *inductive reasoning.* Reasoning using flow graphs is included in inductive reasoning. This is in contrast to *deductive reasoning*, where axioms expressing some universal truths are used as a departure point of reasoning [8]. We can discover dependencies, correlations and decision rules within a data set without reference to its probabilistic nature by using flow graphs [8]. It is an efficient method for uncertainty management, partly because the branches of a flow graph are interpreted as decision rules. Flow graphs play an important role in reasoning from uncertain data and have been successfully applied in many areas e.g., fuzzy sets [3], search engines [4], rule analysis [6], conflict analysis [10] and data mining [11].

We look at two developments here. One concerns the quality of an individual flow graph. A promising measure considered in this paper is entropy. A decision tree can be constructed as a unique flow graph by removing the root while its nodes are labeled by the same attribute [11]. We further investigate decision tree generation from flow graphs, which is the inverse problem. Thus, creation of decision trees can be accomplished without referring to decision tables but using the information flow about the problem we are interested in.

This paper is organized as follows. Section 2 introduces preliminary definitions of flow graphs. Section 3 describes basic notions of entropy. Next, we state entropy measures of flow graphs (Section 4). Section 5 contains an illustrative data analysis example, followed by an application to decision trees (Section 6).

P. Wen et al. (Eds.): RSKT 2009, LNCS 5589, pp. 618–625, 2009.

2 Flow Graphs

In this section we breifly review and discuss basic definitions and some mathematical properties of flow graphs from the studies of Pawlak [9,11].

Flow graphs have traditionally been used for managing uncertainty [1,3,4,6,8,9,10,11]. In order to demonstrate interesting relationships between flow graphs and other disciplines, we consider the normalized version of flow graphs.

A *normalized flow graph* is a *directed, acyclic, finite graph* $G = (N, B, \sigma)$, where N is a set of *nodes*, $B \subseteq N \times N$ is a set of *directed branches* φ: $B \rightarrow R^+$ is a flow function, $\varphi(G)$ is a *throughflow* of flow graph G, σ: $B \rightarrow [0,1]$ is a *normalized flow* of (x, y) and $\sigma(x)$ is a *normalized throughflow* of x.

With every decision rule, there are three associated coefficients: *strength*, *certainty* and *coverage*. The *strength* of (x, y) is given by

$$\sigma(x,y) = \frac{\varphi(x,y)}{\varphi(G)}. \tag{1}$$

For every node x of a flow graph G, the associated *normalized inflow* and *outflow* are defined respectively as $\varphi_+(x) = \sum_{y \in I(x)} \sigma(y,x)$, $\varphi_-(x) = \sum_{y \in O(x)} \sigma(x,y)$. For every branch (x, y) of a flow graph G, the *certainty* and the *coverage* of (x, y) are defined respectively as:

$$cer(x,y) = \frac{\sigma(x,y)}{\sigma(x)}, \tag{2}$$

$$cov(x,y) = \frac{\sigma(x,y)}{\sigma(y)} \tag{3}$$

where $\sigma(x), \sigma(y) \neq 0$. As a consequence of the previous definitions, the following properties hold: $\sum_{y \in O(x)} cer(x,y) = \sum_{y \in I(x)} cov(x,y) = 1$, $cer(x,y) = \frac{cov(x,y)\sigma(y)}{\sigma(x)}$ and $cov(x,y) = \frac{cer(x,y)\sigma(x)}{\sigma(y)}$ [9,11]. The two last equations are Bayes' rules [8] which simplifies computation. Furthermore, *flow conservations* of flow graphs are discussed in [9].

Example 1. Consider the well known weather data set for data mining, given in Table 1 [5], where *Outlook*, *Temperature*, *Humidity*, *Wind* are condition attributes and *PlayTennis* is the decision attribute. Fig. 1 depicts the flow graph corresponding to this data set. Each group of nodes (vertically) in the flow graph is referred to as a *layer*, for example, Fig. 1 has five layers. Every layer corresponds to a particular attribute, and all nodes of a layer correspond to possible values of the attribute. We can interpret some patterns e.g., the database shows 36% sunny outlook, 29% overcast outlook and 36% rain outlook. We also know that 40% of sunny days are high, 40% are mild and 20% are low, etc. Briefly, the flow graph visualizes the information of the data given in Table 1.

If we focus on *Wind* and the decision *PlayTennis*, then we can construct a flow graph to analyze its information flow and decision rules as shown in Fig. 2(a)[1]. Nodes in the first layer are the possible values of *Wind* labelled by Weak and Strong with their normalized throughflow, $\sigma(x)$, (calculated from $\frac{8}{14}$

[1] The computations may contain roundoff errors.

Table 1. Weather data set [5]

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

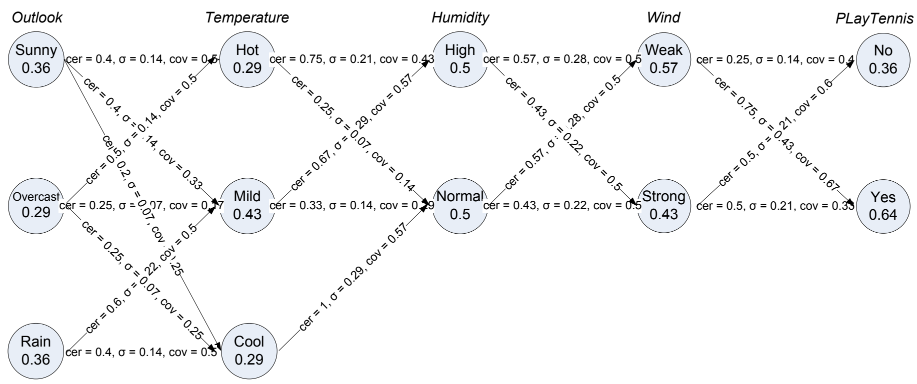
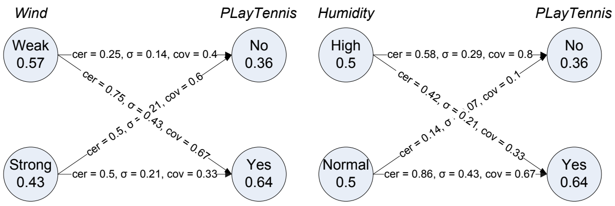


Fig. 1. Flow graph weather data

and $\frac{6}{14}$) indicate there are 57% and 43% of days which have weak and strong wind, respectively. The nodes in the second layer are the possible values of *PlayTennis*, labelled by No and Yes. These nodes indicate that 64% and 36% of the days they do and do not play tennis, respectively.

All branches are interpreted as decision rules with certainty, strength and coverage coefficients computed by (1)–(3). In the branches starting from Weak, $\sigma = 0.14$ and $\sigma = 0.43$, there are 14% and 43% of days that wind is weak but they do not play tennis and play tennis, respectively. Accordingly, $cer = 0.25$ and $cer = 0.75$ indicate that there are 25% and 75% of the weak wind days that they do not play tennis and play tennis, respectively. Finally, for the branches ending at No, $cov = 0.4$ and $cov = 0.6$ indicate that there are 40% and 60% of the do not play tennis days which are weak and strong wind, respectively. Similarly, Fig. 2(b) illustrates the flow graph of *Humidity* and *PlayTennis*. Traditionally, decision rules from flow graphs with large values of certainty are included in the classifier system e.g., IF *Wind* = Weak THEN *PlayTennis* = Yes or IF *Humidity* = High THEN *PlayTennis* = No. The respective values of coverage are useful to give explanations (reasons) for these decision rules.



(a) *Wind* and *PlayTennis* (b) *Humidity* and *PlayTennis*

Fig. 2. Flow graphs of weather data

3 Entropy

Entropy-based measurements of uncertainty and predictability of flow graphs are considered in this paper. The entropy of a random variable measures the uncertainty associated with that variable (sometimes called the *Shannon entropy* [12]).

Definition 1. *If X is a discrete random variable and $p(x)$ is the value of probability distribution, then the entropy of X is $H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$.*

Definition 2. *If X and Y are discrete random variables and $p(x, y)$ is the value of their joint probability distribution at (x, y), then the joint entropy of X and Y is $H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y)$.*

Joint entropy is the amount of information in two (or more) random variables whereas conditional entropy is the amount of information in one random variable given we already know the other.

Definition 3. *If X and Y are discrete random variables and $p(x, y)$ and $p(y|x)$ are the values of their joint and conditional probability distributions, then the conditional entropy of Y given X is $H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x)$.*

4 Entropy Measures of Flow Graphs

Similar to the standard definition of entropy, we give definitions of the entropy of a flow graph in this section (measured in *bits*).

Traditionally, given a collection (training data) S, the entropy is defined relative to its classification to characterizes the purity of this collection of examples. This is essentially the entropy of the probability distribution defined by the data set for the decision attribute Y. We define the entropy of a flow graph by replacing this probability with the normalized throughflow of value y, which is natural, since $\sigma(y) = p(Y = y)$.

Definition 4. *The entropy, $H(G)$, of a flow graph G is defined as*

$$H(G) = -\sum_{y \in Y} \sigma(y) \log_2 \sigma(y)$$

where $\sigma(y)$ is the normalized throughflow of the decision attribute value $Y = y$.

The joint entropy and conditional entropy between attributes X and Y are also defined similar to existing definitions.

Definition 5. *If X and Y are attributes in the flow graph G then the joint entropy of X and Y is*

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} \sigma(x,y) \log_2 \sigma(x,y)$$

where $\sigma(x,y)$ is the strength coefficient of the attribute values $X = x$ and $Y = y$.

The task of inference in knowledge discovery is related to computing $p(Y = y|X = x)$ where x and y are values of condition and decision attributes. This how the entropy measures of a flow graph describe the predictive performance of an attribute. Recall that $cer(x,y) = p(Y = y|X = x)$. We define the conditional entropy of attributes in a flow graph below.

Definition 6. *If X and Y are attributes in the flow graph G then the conditional entropy of Y given X is defined as*

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} \sigma(x,y) \log_2 cer(x,y).$$

where $\sigma(x,y)$ and $cer(x,y)$ are the strength and certainty coefficients of the attribute values $X = x$ and $Y = y$, respectively.

Next, we give a formula to compute the information gain, $Gain(G,X)$ of a particular condition attribute X. It is the original entropy of the flow graph (Definition 4) minus the conditional entropy of Y given X.

Definition 7. *If X and Y are attributes in the flow graph G then the information gain, $Gain(G,X)$ of an attribute X, relative to a flow graph G, is defined as*

$$Gain(G,X) = H(G) - H(Y|X).$$

Information gain can serve as a tool for predictive performance discovery. It measures the effectiveness of an attribute to classify the given (training) flow graph. In other words, it indicates the best prediction attribute (having highest $Gain(G,X)$) of decision attributes in flow graph.

5 Illustrative Example

The attributes of the data[2] given in Fig. 2, can be regarded as discrete random variables where *Wind* takes values $w = \{\text{Weak, Strong}\}$, *Humidity*—$h = \{\text{High, Normal}\}$ and *PlayTennis*—$t = \{\text{Yes, No}\}$.

First let us focus on *Wind* and *PlayTennis* in the flow graph in Fig. 2(a), their normalized throughflows are:

- $\sigma(x)$ of *Wind* are $\sigma(Weak) = 0.57$ and $\sigma(Strong) = 0.43$, and
- $\sigma(y)$ of *PlayTennis* are $\sigma(No) = 0.36$ and $\sigma(Yes) = 0.64$.

[2] $Wind$ and $PlayTennis$ are not independent since $p(Wind = w, PlayTennis = t) \neq p(Wind = w) \times p(PlayTennis = t)$, c.f. [1] for dependency issue.

The strength coefficients $\sigma(x, y)$ of *Wind* and *PlayTennis* are:

- $\sigma(Weak, No) = 0.14$,
- $\sigma(Weak, Yes) = 0.43$,
- $\sigma(Strong, No) = 0.21$,
- $\sigma(Strong, Yes) = 0.21$.

First, let us calculate the entropy of the flow graph G in Fig. 2(a) by using Definition 4.

$$\begin{aligned} H(G) &= -\sum_{y \in PlayTennis} \sigma(y) \log_2 \sigma(y) \\ &= -(0.36 \log_2 0.36 + 0.64 \log_2 0.64) \\ &= 0.94. \end{aligned}$$

We calculate the certainty coefficients $cer(x, y)$ of $Wind$ and $PlayTennis$ as

- $cer(Weak, No) = \frac{\sigma(Weak,No)}{\sigma(Weak)} = 0.25$,
- $cer(Weak, Yes) = 0.75$,
- $cer(Strong, No) = 0.5$,
- $cer(Strong, Yes) = 0.5$.

According to Definition 6, conditional entropy of *PlayTennis* given *Wind* is

$$\begin{aligned} H(PlayTennis|Wind) &= -\sum_{x \in Wind} \sum_{y \in PlayTennis} \sigma(x, y) \log_2 cer(x, y) \\ &= -(0.14 \log_2 0.25 + 0.21 \log_2 0.5 \\ &\quad +0.43 \log_2 0.75 + 0.21 \log_2 0.5) \\ &= 0.89. \end{aligned}$$

Thus, the information gain of attribute *Wind*, relative to the flow graph G given in Fig. 2(a) is

$$\begin{aligned} Gain(G, Wind) &= H(G) - H(PlayTennis|Wind) \\ &= 0.94 - 0.89 \\ &= 0.05. \end{aligned}$$

Similarly, the information gain of *Humidity*, relative to a flow graph G in Fig. 2(b) is $Gain(G, Humidity) = 0.15$

According to the discussion provided in [5], our definition of information gain of a particular attribute relative to a flow graph G is similar to the information gain they used to build decision trees. $Gain(G, Wind)$ is less than $Gain(G, Humidity)$ indicates condition attribute $Wind$ has less predictive power than $Humidity$. In other words, knowing the $Humidity$ value helps to predict outcome better than knowing the $Wind$ value. This is thus referring to the quality of the flow graph in Fig. 2(a) as a classifier is not as predictive as the flow graph in Fig. 2(b).

6 An Application to Decision Trees

Entropy in machine learning has been successful in computing information gain in decision trees [5]. In further discussion, we adopt standard terminology concerning decision trees like root, branches, etc. Starting from a decision tree, it can be constructed as a unique flow graph by removing the root while its nodes are labeled by the same attribute [11]. Theoretically, Butz et al. showed that a flow graph is a special case of chain Bayesian network [1]. On the contrary, in some situations, such as voting analysis and supply–demand problems (discussed in [8]), the available format is not a decision table or a database but is a flow graph. Hence, a classifier system constructed directly from a flow graph instead of a decision table is required. It is shown that classification of objects in flow graph representation boils down to finding the maximal output flow [8]. For these reasons, flow graph information gain is applied to solve this problem.

In this section, a new construction for decision trees from a flow graph is established. It is not sufficient to just precede in the inverse order as a (unique) flow graph construction from a decision tree, since a flow graph can be rearranged its layer order. Thus, it can construct several decision trees with distinct predictabilities.

Example 2. We are given the initial flow graph in Fig. 1 and we aim to construct a decision tree classifier. We can construct the decision tree by adding its root. All corresponding nodes and branches are inherited from the flow graph. Although, a decision tree constituted directly from this graph will have the level of nodes as appeared in the flow graph which has less predictive performance as discussed in [5]. Alternatively, from Section 5, the information gain (predictive) order of condtional attributes is *Outlook*, *Humidity*, *Wind* and *Temperature*. Then we can generate a decision tree classifier (Fig. 3). Its levels are determined according to their flow graph information gains. As one can see, this flow graph can be used as a classifier for *PlayTennis*. Hence, a more predictive decision tree can be constructed by using the proposed flow graph entropy and information gain.

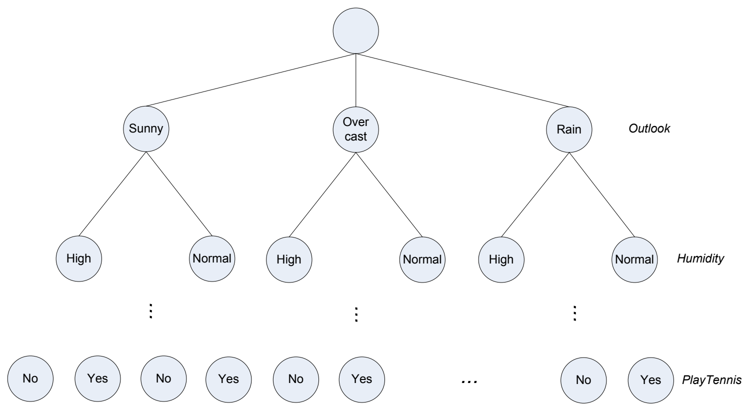


Fig. 3. Predictive decision tree from weather data set

7 Concluding Remarks

We propose flow graphs' analysis based on entropy computation. We have shown a new mathematical relationship between flow graphs and entropy, which can be used for data analysis. In particular, an information gain derived from entropy is suitable for creating and analyzing decision trees from flow graphs when starting from a specified format. The entropy of flow graphs may have applications not necessarily associated with decision trees, but these require further study. Future works are an exploration of such problems and a more formal scheme for decision tree generation.

Acknowledgement

This research was supported by the grant MRG5180071 from the Thailand Research Fund and the Commission on Higher Education. Thanks are due to G.M. Zaverucha.

References

1. Butz, C.J., Yan, W., Yang, B.: An Efficient Algorithm for Inference in Rough Set Flow Graphs. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets V. LNCS (LNAI), vol. 4100, pp. 102–122. Springer, Heidelberg (2006)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Inc., New York (1991)
3. Chitcharone, D., Pattaraintakorn, P.: Towards Theories of Fuzzy Set and Rough Set to Flow Graphs. In: The 2008 IEEE World Congress on Computational Intelligence, pp. 1675–1682. IEEE Press, Hong Kong (2008)
4. Kostek, B., Czyzewski, A.: Processing of Musical Metadata Employing Pawlak's Flow Graphs. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS (LNAI), vol. 3100, pp. 279–298. Springer, Heidelberg (2004)
5. Mitchell, T.M.: Machine Learning. McGraw-Hill Higher Education, New York (1997)
6. Pattaraintakorn, P., Cercone, N., Naruedomkul, K.: Rule Learning: Ordinal Prediction Based on Rough Set and Soft-Computing. Appl. Math. Lett. 19(12), 1300–1307 (2006)
7. Pawlak, Z.: Rough Sets. In: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
8. Pawlak, Z.: In Pursuit of Patterns in Data Reasoning from Data–the Rough Set Way. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 1–9. Springer, Heidelberg (2002)
9. Pawlak, Z.: Decision Algorithms and Flow Graphs: a Rough Set Approach. J. of Telecom. and Inform. Tech. 3, 98–101 (2003)
10. Pawlak, Z.: Some Remarks on Conflict Analysis. European J. of Oper. Res. 166, 649–654 (2005)
11. Pawlak, Z.: Decision Trees and Flow Graphs. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H.S., Słowiński, R. (eds.) RSCTC 2006. LNCS (LNAI), vol. 4259, pp. 1–11. Springer, Heidelberg (2006)
12. Shannon, C.E.: A Mathematical Theory of Communication. AT&T Tech. J. 27, 379–423, 623–656 (1948)

An Extension of Rough Set Approximation to Flow Graph Based Data Analysis

Doungrat Chitcharoen and Puntip Pattaraintakorn

School of Mathematics, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 10520
s0067111@kmitl.ac.th, kppuntip@kmitl.ac.th

Abstract. This paper concerns some aspects of mathematical flow graph based data analysis. In particular, taking a flow graph view on rough sets' categories and measures leads to a new methodology of inductively reasoning form data. This perspective shows interesting relationships and properties among rough set, flow graphs and inverse flow graphs. A possible car dealer application is outlined and discussed. Evidently, our new categories and measures assist and alleviate some limitations in flow graphs to discover new patterns and explanations.

Key words: Flow graphs, rough sets and decision rules

1 Introduction

A mathematical flow graph, invented by Pawlak in 2002, is an extension of *rough set theory* [9]. A flow graph represents the information flow from the given data set [10–14]. The branches of a flow graph can be constructed as decision rules, with every decision rule, there are three associated coefficients: *strength*, *certainty* and *coverage* [14]. These coefficients satisfy Bayes' theorem. Inference in flow graphs has polynomial time and flow conservation comes with probabilistic conditional independencies in the problem domain [1]. Flow graphs have led to many interesting applications and extensions such as preference analysis [11], decision tree [13], survival analysis [7], association rule [3], data mining [14], search engines [2], fuzzy set [4, 6], entropy measures [8] and granular computing [5]. More studies involving rough sets are discussed and provided in [15].

Flow distribution in flow graphs can be exploited for approximation and reasoning. Based on flow graph contexts, we define fundamental definitions for rough sets: four categories of vagueness, accuracy of approximation, roughness of approximation and dependency degree. In addition, we state formulas to conveniently compute these measures for inverse flow graphs. To illustrate, a possible car dealer preference analysis is provided to support our propositions. New categories and measures assist and alleviate some limitations in flow graphs to discover new patterns and explanations.

This paper is organized as follows. In Section 2, we present the basic concepts of rough sets. In Section 3, we recall preliminary definitions of flow graphs. In Section 4, we present a new bridge between rough sets and flow graphs with an example throughout, followed by a conclusion in the last section.

2 Rough Set Theory

The following rough sets preliminary is taken from [9]. Rough sets are based on an *information system*. Formally, it is a pair $S = (U, A)$, where U is a nonempty finite set of objects called the *universe* and A is a nonempty finite set of attributes such that $a \colon U \to V_a$ for every $a \in A$. The set V_a is called the *domain* of a.

If we partition an information system into two disjoint classes of attributes, called *condition* and *decision attributes*, then the information system will be called a *decision system*, denoted by $S = (U, C, D)$, where $C \cap D = \emptyset$. Any subset B of A determines a binary relation $I(B)$ on U called an *indiscernibility relation*. It is defined as $(x, y) \in I(B)$ if and only if $a(x) = a(y)$ for every $a \in A$, where $a(x)$ denotes the attribute value of element x. Equivalence classes of the relation $I(B)$ are referred to as *B-elementary sets* or *B-elementary granules* denote by $B(X)$, i.e., $B(X)$ describes X in the terms of attribute values from B [11]. Below, we recall key feature definitions of approximations in rough sets.

Definition 1 *[15] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$. The B-lower approximations, B-upper approximations and B-boundary region of X are defined as $\underline{B}(X) = \bigcup_{x\in U} \{B(X) | \, B(X) \subseteq X\}$, $\overline{B}(X) = \bigcup_{x\in U} \{B(X) | \, B(X) \cap X \neq \emptyset\}$ and $BN_B(X) = \overline{B}(X) - \underline{B}(X)$, respectively.*

If the boundary region of X is the empty set (i.e., $BN_B(x) = \emptyset$), then X is *crisp*. On the contrary, if $BN_B(X) \neq \emptyset$, then X is *rough*. In what follows we recall four basic classes of rough sets, i.e., four categories of vagueness.

Definition 2 *[15] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$, the four categories of vagueness are defined as*

- $\underline{B}(X) \neq \emptyset$ *and* $\overline{B}(X) \neq U$ *iff X is roughly B-definable,*
- $\underline{B}(X) = \emptyset$ *and* $\overline{B}(X) \neq U$ *iff X is internally B-indefinable,*
- $\underline{B}(X) \neq \emptyset$ *and* $\overline{B}(X) = U$ *iff X is externally B-definable,*
- $\underline{B}(X) = \emptyset$ *and* $\overline{B}(X) = U$ *iff X is totally B-indefinable.*

Approximation of a rough set can be characterized numerically by some measurements as follows.

Definition 3 *[15] Let $S = (U, A)$ be an information system. For $X \subseteq U$, $B \subseteq A$, the accuracy of approximation, $\alpha_B(X)$, and roughness of approximation, $\gamma_B(X)$, are defined respectively as $\alpha_B(X) = \frac{card(\underline{B}(X))}{card(\overline{B}(X))}$ and $\gamma_B(X) = 1 - \alpha_B(X) = 1 - \frac{card(\underline{B}(X))}{card(\overline{B}(X))}$, where $card(X)$ denotes the cardinality of X.*

Let us observe that, $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, then X is *crisp* with respect to B and otherwise, if $\alpha_B(X) < 1$, then X is *rough* with respect to B.

Definition 4 *Let $S = (U, A)$ be an information system and $F = \{X_1, X_2, \ldots, X_n\}$ be a partition of the universe U. For $B \subseteq A$, F depends on B to a degree $k_B(F) = \frac{\sum_{i=1}^{n} card(\underline{B}(X_i))}{card(U)}$.*

Definitions $2 - 4$ will be stated in the context of flow graphs in Section 4.

3 Flow Graphs

In this section, we recall some concepts of flow graphs which were introduced by Pawlak in [10–14].

A *flow graph* is a *directed, acyclic, finite graph* $G = (\mathcal{N}, \mathcal{B}, \varphi)$, where $\mathcal{N}$ is a set of nodes, $\mathcal{B} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of *directed branches*, φ: $\mathcal{B} \to R^+$ is a *flow function* and R^+ is the set of non-negative real numbers. If $(x, y) \in \mathcal{B}$ then x is an *input* of node y denoted by $I(y)$ and y is an *output* of node x denoted by $O(x)$. The *input* and *output* of a flow graph G are defined by $I(G) = \{x \in \mathcal{N} \mid I(x) = \emptyset\}$ and $O(G) = \{x \in \mathcal{N} \mid O(x) = \emptyset\}$. These inputs and outputs of G are called *external nodes* of G whereas other nodes are called *internal nodes* of G. If $(x, y) \in \mathcal{B}$ then we call (x, y) a *throughflow* from x to y. We will assume in what follows that $\varphi(x, y) \neq 0$ for every $(x, y) \in \mathcal{B}$. With every node x of a flow graph G, we have its associated *inflow* and *outflow* respectively as: $\varphi_+(x) = \sum_{y \in I(x)} \varphi(y, x)$ and $\varphi_-(x) = \sum_{y \in O(x)} \varphi(x, y)$. Similarly, an *inflow* and an *outflow* for the flow graph G are defined as: $\varphi_+(G) = \sum_{x \in I(G)} \varphi_-(x)$ and $\varphi_-(G) = \sum_{x \in O(G)} \varphi_+(x)$. We assume that for any internal node x, $\varphi_-(x) = \varphi_+(x) = \varphi(x)$, where $\varphi(x)$ is a *throughflow* of node x. Similarly then, $\varphi_-(G) = \varphi_+(G) = \varphi(G)$ is a *throughflow* of graph G. As discussed by Pawlak [11], the above equations can be considered as *flow conservation equations* (or *pairwise consistent* [1]).

Normalized Flow Graphs, Paths and Connections

In order to demonstrate interesting relationships between flow graphs and other disciplines (e.g., statistics), we come to the normalized version of flow graphs.

A *normalized flow graph* is a *directed, acyclic, finite graph* $G = (\mathcal{N}, \mathcal{B}, \sigma)$, where $\mathcal{N}$ is a set of *nodes*, $\mathcal{B} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of *directed branches* and σ: $\mathcal{B} \to [0, 1]$ is a *normalized flow function* of (x, y). The *strength* of (x, y) is $\sigma(x, y) = \frac{\varphi(x,y)}{\varphi(G)}$. With every node x of a flow graph G, the associated *normalized inflow* and *outflow* are defined as: $\sigma_+(x) = \frac{\varphi_+(x)}{\varphi(G)} = \sum_{y \in I(x)} \sigma(y, x)$, $\sigma_-(x) = \frac{\varphi_-(x)}{\varphi(G)} = \sum_{y \in O(x)} \sigma(x, y)$. For any internal node x, it holds that $\sigma_+(x) = \sigma_-(x) = \sigma(x)$, where $\sigma(x)$ is a *normalized throughflow* of x. Similarly, *normalized inflow* and *outflow* for the flow graph G are defined as: $\sigma_+(G) = \frac{\varphi_+(G)}{\varphi(G)} = \sum_{x \in I(G)} \sigma_-(x)$, $\sigma_-(G) = \frac{\varphi_-(G)}{\varphi(G)} = \sum_{x \in O(x)} \sigma_+(x)$. It also holds that $\sigma_+(G) = \sigma_-(G) = \sigma(G) = 1$. With every branch (x, y) of a flow graph G, the *certainty* and the *coverage* of (x, y) are defined respectively as: $cer(x, y) = \frac{\sigma(x,y)}{\sigma(x)}$, $cov(x, y) = \frac{\sigma(x,y)}{\sigma(y)}$, where $\sigma(x), \sigma(y) \neq 0$. Properties of these coefficients were studied by Pawlak in [10–14].

Next, if we focus on sequence of nodes in a flow graph, we can find them by using the concept of a directed simple path. A (directed) *path* from x to y $(x \neq y)$ in G, denoted by $[x \dots y]$, is a sequence of nodes $x_1, \dots, x_n$ such that $x_1 = x$ and $x_n = y$ and $(x_i, x_{i+1}) \in B$ for every i, $1 \leq i \leq n - 1$. The *certainty, coverage* and *strength of the path* $[x_1 \dots x_n]$ are defined respectively as: $cer[x_1 \dots x_n] = \prod_{i=1}^{n-1} cer(x_i, x_{i+1})$, $cov[x_1 \dots x_n] = \prod_{i=1}^{n-1} cov(x_i, x_{i+1})$, $\sigma[x \dots y] = \sigma(x) cer[x...y] = \sigma(y) cov[x \dots y]$.

The set of all paths from x to y ($x \neq y$) in G, denoted by $\langle x, y\rangle$, is a *connection* of G determined by nodes x and y. For every connection $\langle x, y\rangle$, the associated *certainty, coverage* and *strength of the connection* $\langle x, y\rangle$ are defined as: $cer\,\langle x,y\rangle = \sum_{[x\ldots y]\in\langle x,y\rangle} cer[x\ldots y]$, $cov\,\langle x,y\rangle = \sum_{[x\ldots y]\in\langle x,y\rangle} cov[x\ldots y]$, $\sigma\,\langle x,y\rangle = \sum_{[x\ldots y]\in\langle x,y\rangle} \sigma[x\ldots y] = \sigma(x)cer\,\langle x,y\rangle = \sigma(y)cov\,\langle x,y\rangle$. If $[x\ldots y]$ is a path such that x and y are the input and output of G, then $[x\ldots y]$ will be referred to as a *complete path.* The set of complete paths from x to y will be called a *complete connection* from x to y in G.

If we substitute every complete connection $\langle x, y\rangle$ in G, where x and y are an input and an output of a graph G with a single branch (x, y) such that $\sigma(x, y) = \sigma\,\langle x, y\rangle$, $cer(x, y) = cer\,\langle x, y\rangle$ and $cov(x, y) = cov\,\langle x, y\rangle$ then we have a new flow graph G' with the property: $\sigma(G) = \sigma(G')$. G' is called a *combined flow graph.*

Starting from a flow graph, if we invert the direction of all branches in G, then the resulting graph G^{-1} will be called the *inverted graph of* G (or the *inverse flow graph of* G) [14]. Essentially, three coefficients of an inverse flow graph can be computed from its flow graph as follows: $\sigma_{G^{-1}}(y, x) = \sigma_G(x, y)$, $cer_{G^{-1}}(y, x) = cov_G(x, y)$ and $cov_{G^{-1}}(y, x) = cer_G(x, y)$.

4 Rough Set Approximations and Flow Graphs

In this section, we provide a bridge between flow graphs and rough approximation. From standard definitions of approximations made by rough sets, we give these definitions in the context of flow graphs below.

Suppose we are given a normalized flow graph $G = (A, \mathcal{B}, \sigma)$, where $A = \{A_{l_1}, A_{l_2}\ , \ldots, A_{l_n}\}$ is a set of attributes[1], $\mathcal{B}$ is a set of directed branches and σ is a normalized flow function. A set of nodes in a flow graph G corresponding to A_{l_i} is referred to as a *layer i*. For $A = C \cup D$, we have that every layer corresponding to C will be called a *condition layer* whereas every layer corresponding to D will be called a *decision layer.* If an attribute A_{l_i} contains n_{l_i} values, we say that it contains n_{l_i} nodes.

We now consider how to approximate an attribute value $Y \in A_{l_{i+1}}$ from attribute values of A_{l_i} where $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, to indicate lower approximation, upper approximation and boundary region of Y. In Definition 5, we recall Pawlak's definitions of lower approximation, upper approximation and boundary region for flow graphs.

Definition 5 *[11] Let $G = (A, \mathcal{B}, \sigma)$ be a normalized flow graph, $A_{l_i} = \{X_1, X_2, \ldots, X_{n_{l_i}}\Big\}$, $1 \leq i \leq k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of the flow graph G, the union of all inputs X_j of Y is the upper approximation of Y (denoted $\overline{A_{l_i}}(Y)$), the union of all inputs X_j of Y, such that $cer(X_j, Y) = 1$, is the lower approximation of Y (denoted $\underline{A_{l_i}}(Y)$). Moreover, the union of all inputs X_j of Y, such that $cer(X_j, Y) < 1$, is the boundary region of Y (denoted $A_{l_i}N_{A_{l_i}}(Y)$).*

[1] In what follows, we regard $\mathcal{N}$ as A for simplicity.

In Definition 6, we state four categories of rough sets mentioned in Definition 2 in terms of flow graph.

Definition 6 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \leq i \leq k-1$, be an attribute in layer i and Y be a node in $A_{l_{i+1}}$. For any branch (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, of G, we define four categories of vagueness as*

- $\exists X_j\, [cer(X_j, Y) = 1]$ *and* $\exists X_j\, [X_j \notin I(Y)]$ *iff Y is roughly A_{l_i}-definable,*
- $\forall X_j\, [cer(X_j, Y) \neq 1]$ *and* $\exists X_j\, [X_j \notin I(Y)]$ *iff Y is internally A_{l_i}-indefinable,*
- $\exists X_j\, [cer(X_j, Y) = 1]$ *and* $\forall X_j\, [X_j \in I(Y)]$ *iff Y is externally A_{l_i}-definable,*
- $\forall X_j\, [cer(X_j, Y) \neq 1]$ *and* $\forall X_j\, [X_j \in I(Y)]$ *iff Y is totally A_{l_i}-indefinable.*

From the definition we obtain the following interpretation:

- if Y is roughly A_{l_i}*-definable*, this means that we are able to decide for some elements of U whether they belong to Y or $-Y$[2], using A_{l_i},
- if Y is internally A_{l_i}*-indefinable*, this means that we are able to decide whether some elements of U belong to $-Y$, but we are unable to decide for any element of U, whether it belongs to Y or not, using A_{l_i},
- if Y is externally A_{l_i}*-indefinable*, this means that we are able to decide for some elements of U whether they belong to Y, but we are unable to decide, for any element of U whether it belongs to $-Y$ or not, using A_{l_i},
- if Y is totally A_{l_i}*-indefinable*, we are unable to decide for any element of U whether it belongs to Y or $-Y$, using A_{l_i}.

Property 1 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $2 \leq i \leq k$, be an attribute in layer i and W be a node in $A_{l_{i-1}}$. For any branch (X_j, W), $j \in \{1, \ldots, n_{l_i}\}$ in the inverse flow graph of G, the union of all output X_j of W in flow graph G is the upper approximation of W, the union of all outputs X_j of W in a flow graph G, such that $cov(W, X_j) = 1$, is the lower approximation of W. Moreover, the union of all outputs X_i of W, such that $cov(W, X_j) < 1$, is the boundary region of Y.*

Proof. It can be proved in a straightforward way according to definition and property of inverse flow graph and Definition 5. □

Example Suppose we are given the flow graph for the preference analysis problem depicted in Fig. 1, that describes four disjoint models of cars $X = \{X_1, X_2, X_3, X_4\}$. They are sold to four disjoint groups of customers $Z = \{Z_1, Z_2, Z_3, Z_4\}$ through three dealers $Y = \{Y_1, Y_2, Y_3\}$.

By Definition 5, when we consider customer Z_1: the lower approximation of Z_1 is an empty set, the upper approximation of Z_1 is $Y_1 \cup Y_2$ and the boundary region Z_1 is $Y_1 \cup Y_2$. Hence, by Definition 6, we conclude that Z_1 is internally Y-indefinable. In Fig. 1 (only limited information is available), by using the set of dealers (Y) to approximate the customer group Z_1 together with the flow distribution visualized in layers two and three, our results can be summarized as the following.

[2] Where $-Y = U - Y$.

- Since no branch connects Y_3 and Z_1, there is no customer Z_1 buys a car from dealer Y_3. As a result if dealer Y_3 plans to run new promotional campaigns, they do not need to pay attention to customer group Z_1 in these campaigns.
- If a customer buys a car through dealer Y_1 or Y_2, then we cannot conclude whether this is a customer in group Z_1 or not. Thus, if dealers Y_1 and Y_2 plan to run promotional campaigns, then they should, at least, target at customer group Z_1 in their campaigns.

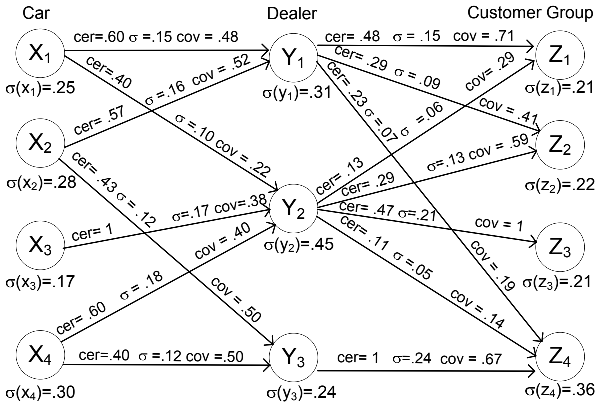


Fig. 1. A normalized flow graph.

Similarly, we can approximate all attribute values (node) in the inverse flow graph of G by using Property 1.

However, the flow graph perspective on rough sets' categories in Definition 6 do not provide approximations quantitively. Hence, in Definitions 7 and 8, we define two measures for flow graphs, the accuracy of approximation and the roughness of approximation.

Definition 7 *Let* $G=(A,\mathcal{B},\sigma)$ *be a flow graph,* $A_{l_i}=\left\{X_1,X_2,\ldots,X_{n_{l_i}}\right\}$, $1 \leq i \leq k-1$, *be an attribute in layer* i *and* Y *be a node in* $A_{l_{i+1}}$. *For any branch* (X_j,Y), $j \in \{1,\ldots,n_{l_i}\}$, *of* G, *the accuracy of approximation,* $\alpha_{A_{l_i}}(Y)$, *is defined as:* $\alpha_{A_{l_i}}(Y) = \frac{card(\underline{\underline{A_{l_i}}}(Y))}{card(\overline{\overline{A_{l_i}}}(Y))}$.

We can use the accuracy of approximation to specify the quality of an approximation. Obviously, $0 \leq \alpha_B(X) \leq 1$. If $\alpha_{A_{l_i}}(Y)=1$, then Y is crisp with respect to A_{l_i}, and otherwise, if $\alpha_{A_{l_i}}(Y)<1$, then Y is rough with respect to A_{l_i}.

Definition 8 *Let* $G=(A,\mathcal{B},\sigma)$ *be a flow graph,* $A_{l_i}=\left\{X_1,X_2,\ldots,X_{n_{l_i}}\right\}$, $1 \leq i \leq k-1$, *be an attribute in layer* i *and* Y *be a node in* $A_{l_{i+1}}$. *For any branch*

(X_i, Y), $i \in \{1, \ldots, n_{l_i}\}$, *of* G, *the roughness of approximation,* $\gamma_{A_{l_i}}(Y)$, *is defined as:* $\gamma_{A_{l_i}}(Y) = 1 - \alpha_{A_{l_i}}(Y) = \frac{card(\overline{A_{l_i}}(Y)) - card(\underline{A_{l_i}}(Y))}{card(\overline{A_{l_i}}(Y))}$.

Property 2 *Let* $G = (A, \mathcal{B}, \sigma)$ *be a flow graph,* $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $1 \leq i \leq k-1$, *be an attribute in layer* i *and* Y *be a node in* $A_{l_{i+1}}$. *For any branch* (X_j, Y), $j \in \{1, \ldots, n_{l_i}\}$, *of* G, *we have*

(1) $\alpha_{A_{l_i}}(Y) = \frac{\sum_{cer(X_j,Y)=1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_j)}$ *and (2)* $\gamma_{A_{l_i}}(Y) = \frac{\sum_{cer(X_j,Y)<1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_i)}$.

Proof. (1) From Definition 5, we have $card(\underline{A_{l_i}}(Y)) = \sum_{cer(X_j,Y)=1} card(X_j)$ and $card(\overline{A_{l_i}}(Y)) = \sum_{X_j \in I(Y)} card(X_i)$. Since $card(X_j) = \varphi(X_j) = \sigma(X_j)\varphi(G) = \sigma(X_j)\varphi(U)$ and by Definition 7, then $\alpha_B(Y) = \frac{\sum_{cer(X_j,Y)=1} \sigma(X_j)}{\sum_{X_j \in I(Y)} \sigma(X_j)}$.

(2) It can be proved similarly to (1). □

Let us briefly comment on Property 2(1) that the greater the boundary of Y, the lower is the accuracy. If $\alpha_{A_{l_i}}(Y) = 1$, the boundary region of Y is empty.

Property 3 *Let* $G = (A, \mathcal{B}, \sigma)$ *be a flow graph,* $A_{l_i} = \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$, $2 \leq i \leq k$, *be an attribute in layer* i *and* W *be a node in* $A_{l_{i-1}}$. *For any branch* (X_j, W), $j \in \left\{1, \ldots, n_{l_j}\right\}$ *in the inverse flow graph of* G, *we have*

(1) $\alpha_{A_{l_j}}(W) = \frac{\sum_{cov(W,X_j)=1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$ *and (2)* $\gamma_{A_{l_j}}(W) = \frac{\sum_{cov(W,X_j)<1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$.

Proof. (1) From Property 1, we have $card(\underline{A_{l_j}}(W)) = \sum_{cov(X_j,W)=1} card(X_j)$ and $card(\overline{A_{l_j}}(Y)) = \sum_{X_j \in O(W)} card(X_j)$. Since $card(X_j) = \varphi(X_j) = \sigma(X_j)\varphi(G) = \sigma(X_j)\varphi(U)$ and by Definition 7, then $\alpha_{A_{l_j}}(W) = \frac{\sum_{cer(X_j,W)=1} \sigma(X_j)}{\sum_{X_j \in O(W)} \sigma(X_j)}$.

(2) It can be proved similarly to (1). □

Example (Cont.) Consider the branches between dealer and customer group in Fig. 1. We can read from our flow graph that 24% of all customers buy cars through dealer Y_3 ($\sigma(Y_3) = 0.24$) and all of them are in customer group Z_3 ($cer(Y_3, Z_4) = 1$). There is only one branch (Y_3, Z_4) with $cer(Y_3, Z_4) = 1$[3]. Thus, by Property 2(1), we have $\alpha_Y(Z_1) = \alpha_Y(Z_2) = \alpha_Y(Z_3) = 0$ and $\alpha_Y(Z_4) = \frac{\sum_{cer(Y_i,Z_4)=1} \sigma(Y_i)}{\sum_{Y_i \in I(Z_4)} \sigma(Y_i)} = \frac{\sigma(Y_3)}{\sigma(Y_1)+\sigma(Y_2)+\sigma(Y_3)} = 0.24$.

These results imply that we should not make decisions involving customer groups Z_1, Z_2 and Z_3 solely by using dealers due to high imprecision. Nevertheless, we can partly check that it will be customer group Z_4 with low accuracy by using dealers. Similarly, if we consider the roughness of approximation between dealer and customer group, then by Property 2(2), we have $\gamma_Y(Z_1) = \gamma_Y(Z_2)$

[3] By employing the approach presented in our previous study [3], we can extract some interesting association rules. If the model of car X_2 (or X_4) is bought through dealer Y_3 then the customer group is Z_4 with support 0.12 and confidence 1.

$=\gamma_Y(Z_3) = 1$ and $\gamma_Y(Z_4) = 0.76$. We can draw a conclusion in a similar manner as we did for the roughness measure.

Please note that we can calculate the accuracy and the roughness of approximation between attributes in the inverse flow graph by using Property 3. Another important topic in data analysis is dependency between attributes. We introduce dependency degree between any two attributes in Definition 9.

Definition 9 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i}= \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$ and $A_{l_{i+1}}= \left\{Y_1, Y_2, \ldots, Y_{n_{l_{i+1}}}\right\}$, $1 \leq i \leq k$, be any two adjacent layers. $A_{l_{i+1}}$ depends on A_{l_i} to a degree $k_{A_{l_i}}(A_{l_{i+1}}) = \frac{\sum_{l=1}^{n_{l_{i+1}}} card(\underline{A_{l_i}}(Y_l))}{card(U)}$.*

If $k_{A_{l_i}}(A_{l_{i+1}}) = 1$, we say that $A_{l_{i+1}}$ depends totally on A_{l_i}, and if $k_{A_{l_i}}(A_{l_{i+1}}) < 1$, we say that $A_{l_{i+1}}$ depends partially in a degree $k_{A_{l_i}}(A_{l_{i+1}})$ on A_{l_i}. It is worth pointing out that our dependency measure is different to the one given by Pawlak [14]. The former gives dependency degree between two adjacent attributes (layers) while the latter gives dependency degree between two nodes connected by directed branch.

Property 4 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_i}= \left\{X_1, X_2, \ldots, X_{n_{l_i}}\right\}$ and $A_{l_{i+1}}=\left\{X_1, X_2, \ldots, X_{n_{l_{i+1}}}\right\}$, $1 \leq i \leq k$, be any two adjacent layers. $A_{l_{i+1}}$ depends on A_{l_i} to a degree $k_{A_{l_i}}(A_{l_{i+1}}) = \sum_{cer(X_i,X_j)=1} \sigma(X_i)$.*

Proof. From Definition 5, $\sum_{j=1}^{n_{l_{i+1}}} card(\underline{A_{l_i}}(X_j)) = \sum_{j=1}^{n_{l_{i+1}}} \sum_{cer(X_i,Y_j)=1} card(X_i)$. Since $X_n \cap X_m = \emptyset$, $1 \leq n \neq m \leq n_{l_i}$, then $\underline{A_{l_i}}(X_n) \cap \underline{A_{l_i}}(X_m) = \emptyset$. Thus $\sum_{j=1}^{n} card(\underline{A_{l_i}}(X_j)) = \sum_{cer(X_i,Y_j)=1} card(X_i)$. Since $\varphi(X_i) = \sigma(X_i)\varphi(G) = \sigma(X_i)\varphi(U)$ and by Definition 9, we can write $\gamma_B(D) = \sum_{cer(X_i,X_j)=1} \sigma(X_i)$. □

Property 5 *Let $G = (A, \mathcal{B}, \sigma)$ be a flow graph, $A_{l_j} = \left\{X_1, X_2, \ldots, X_{n_{l_j}}\right\}$ and $A_{l_{j-1}}= \left\{X_1, X_2, \ldots, X_{n_{l_{j-1}}}\right\}$, $1 \leq j \leq k+1$, be any two adjacent layers in the inverse flow graph of G. $A_{l_{j-1}}$ depends on A_{l_j} to a degree $k_{A_{l_j}}(A_{l_{j-1}}) = \sum_{cov(X_i,X_j)=1} \sigma(X_i)$.*

Proof. It can be proved similarly as Property 4 □

Example (Cont.) Consider model of car and dealer in the flow graph G in Fig. 1. By Property 4, dealer depends on model of car to a degree $\gamma_X(Y) = \sum_{cer(X_i,Y_j)=1} \sigma(X_i) = \sigma(X_3) = 0.17$. On the other hand, if we consider customer and dealer in the inverse flow graph of G, then by Property 5, we obtain that dealer depends on customer group to a degree $\gamma_Z(Y) = \sum_{cov(Y_i,Z_j)=1} \sigma(Z_i) = \sigma(Z_3) = 0.21$. These results give a conclusion that dealers depend on customer groups more than models of cars.

In what follows, we aim to approximate a specific attribute value by some attribute values such that they are not in adjacent layers. We can use the concept of a *connection* to do this. More specifically, if we aim to approximate an attribute value in an output layer by attribute values in an input layer, then we will use the concept of *complete connection*.

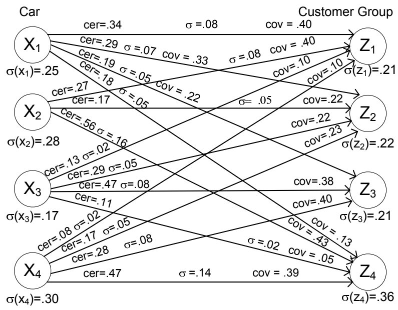


Fig. 2. A combined flow graph.

Example (Cont.) For model of car and customer group in Fig. 1, we give a combined flow graph in Fig. 2. By Definition 5, for Z_4, the lower approximation of Z_4 is an empty set, the upper approximation and the boundary region of Z_4 are $X_1 \cup X_2 \cup X_3 \cup X_4$. Hence, by Definition 6, Z_4 is totally X-indefinable.

By Property 2, we have the accuracy and the roughness approximation of customer Z_4 by model of car as: $\alpha_X(Z_4) = 0$ and $\gamma_X(Z_4) = 1$. Additionally, we can use Property 4 to compute the dependency between model of car and customer group, and the result is 0. From these results due to the imprecision and dependency, we should not make decisions involving customer group Z_4 by using only model of car. As before, we can approximate and measure them for the inverse flow graph in the same way. Comparing the obtained accuracy and roughness measures, we can draw a conclusion that from this population *dealer* is a better indicator for analyzing customer group Z_4 than *model of car*.

5 Conclusion

In this paper, we introduce definitions and properties of rough set approximations, accuracy and roughness of approximation which are defined in terms of a flow graph. They can be useful when the initial data is in the form of flow graph and contains some limitations. We illustrate a car dealer preference analysis to support our propositions.

Acknowledgments

This paper was supported by the grant MRG5180071 from the Thailand Research Fund and the Commission on Higher Education and King Mongkut's Institute of Technology Ladkrabang. Thanks also to G.M. Zaverucha.

References

1. Butz, C.J., Yen, W., Yang, B.: The Computational Complexity of Inference Using Rough Set Flow Graphs. In: Slezak, D. et al. (eds), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. LNAI, vol. 3641, pp. 335–344. Springer–Verlag, Berlin Heidelberg (2005)
2. Czyzewski, A., Kostek, B.: Musical Metadata Retrieval with Flow Graphs. In: Tsumoto, S. et al. (eds.), Rough Sets and Current Trends in Computing. LNAI, vol. 3066, pp. 691–698. Springer–Verlag, Berlin Heidelberg (2004)
3. Chitcharone, D., Pattaraintakorn, P.: Knowledge Discovery by Rough Sets Mathematical Flow Graphs and its Extension. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, pp. 340-345. (2008)
4. Chitcharone, D., Pattaraintakorn, P.: Towards Theories of Fuzzy Set and Rough Set to Flow Graphs. In: the 2008 IEEE World Congress on Computational Intelligence, pp. 1675–1682. IEEE Press, Hong Kong (2008)
5. Liu, H., Sun, J., Zhang, H.: Interpretation of Extended Pawlak Flow Graphs Using Granular Computing, In: Peters, J.F., Skowron, A. (eds.), T. Rough Sets VIII. LNCS, vol. 5084, pp. 93–115. Springer–Verlag, Heidelberg (2008)
6. Matusiewicz, Z., Pancerz, K.: Rough Set Flow Graphs and Max $-*$ Fuzzy Relation Equations in State Prediction Problems. In: Chan, C.C. et al. (eds), Rough Sets and Current Trends in Computing. LNAI, vol. 5306, pp. 359–368. Springer-Verlag, Berlin (2008)
7. Pattaraintakorn, P., Cercone, N., Naruedomkul, K.: Rule Learning: Ordinal Prediction Based on Rough Set and Soft-Computing. Appl. Math. Lett. 19(12), 1300-1307 (2006)
8. Pattaraintakorn, P.: Entropy Measures of Flow Graphs with Applications to Decision Trees. In: Wen, P. et al. (eds.), Rough Sets and Knowledge Technology. LNCS, vol. 5589, pp. 618–625. Springer–Verlag, Berlin (2009)
9. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
10. Pawlak, Z.: Rough Sets, Decision Algorithms and Bayes' Theorem. European J. of Oper. Res. 136, 181–189 (2002)
11. Pawlak, Z.: Rough Set and Flow Graphs. In: Slezak, D. et al. (eds), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. LNAI, vol. 3641, pp. 1–11. Springer–Verlag, Berlin Heidelberg (2005)
12. Pawlak, Z.: Inference Rules and Decision Rules. In: Rutkowski, L. et al. (eds.) International Conference on Artificial Intelligence and Soft Computing. LNAI, vol. 3070, pp. 102–108. Springer–Verlag, Berlin Heidelberg (2004)
13. Pawlak, Z.: Decision Trees and Flow Graphs. In Greco, S. et al. (eds.) Rough Sets and Current Trends in Computing. LNAI, vol. 4259, pp. 1-11. Springer–Verlag, Heidelberg (2006)
14. Pawlak, Z.: Flow Graphs and Data Mining. In: Peters, J.F., Skowron, A. (eds.) Transcations on Rough Sets III. LNCS, vol. 3400, pp. 1–36. Springer–Verlag, Berlin Heidelberg (2005)
15. Pawlak, Z., Skowron, A.: Rudiments of Rough Sets. Inform. Sciences. 177, 3–20 (2007)