

Abstract

Project Code: MRG6080252

Project Title: Study of Algorithmic Evolution in Scientific Literature

Investigator: Assistant Professor Dr. Suppawong Tuarob

Faculty of Information and Communication Technology, Mahidol University

Mentor: Professor Dr. Peter Haddawy

Faculty of Information and Communication Technology, Mahidol University

E-mail Address: suppawong.tua@mahidol.edu

Project Period: 2 years

Computer sciences and related disciplines evolve around developing, evaluating, and applying algorithms. Typically, an algorithm is not developed from scratch, but uses and builds upon existing ones, which often are proposed and published in scholarly articles. The ability to capture this evolution relationship among these algorithms in scientific literature would not only allow us to understand how an algorithm is composed, but also shed light on large-scale analysis of algorithmic evolution through different temporal spans and thematic scales. We propose to capture such evolution relationship between two algorithms by investigating the knowledge represented in citation contexts, where authors explain how cited algorithms are used in their works. A set of heterogeneous ensemble machine-learning methods is proposed, where the combination of two base classifiers trained with heterogeneous feature types is used to automatically identify the algorithm usage relationship. The proposed heterogeneous ensemble methods achieve the best average F1 of 0.749 and 0.905 for fine-grained and binary algorithm citation function classification, respectively. The success of this study will allow us to generate a large-scale algorithm citation network from a collection of scholarly documents representing multiple time spans, venues, and fields of study. Such a network will be used as an instrument not only to answer critical questions in algorithm search, such as identifying the most influential and generalizable algorithms, but also to study the evolution of algorithmic development and trends over time.

Keywords: Algorithm Citation, Ensemble Machine Learning, Scholarly Big Data, Algorithmic Evolution

บทคัดย่อ

รหัสโครงการ: MRG6080252
ชื่อโครงการ: การศึกษาวิวัฒนาการของอัลกอริทึมในวรรณกรรมทางวิทยาศาสตร์
ชื่อนักวิจัย: ผู้ช่วยศาสตราจารย์ ดร. ศุภวงศ์ ทวีรอบ
คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยมหิดล
ชื่อนักวิจัยที่ปรึกษา: Professor Dr. Peter Haddawy
คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยมหิดล
E-mail Address: suppawong.tua@mahidol.edu
ระยะเวลาโครงการ: 2 ปี

อัลกอริทึม (Algorithm) คือกระบวนการ การทำงานที่ใช้การตัดสินใจ โดยนำหลักเหตุผล และคณิตศาสตร์มาช่วยเลือกวิธีการหรือขั้นตอนการดำเนินงานต่อไป จนกระทั่งถึงขั้นตอนสุดท้าย เป็นวิธีการที่ใช้แยกย่อยและเรียงลำดับขั้นตอนของกระบวนการในการทำงานต่างๆ เพื่อเพิ่มประสิทธิภาพในการค้นหาและแก้ไขปัญหา อัลกอริทึมมักถูกค้นคว้าพัฒนาในสาขาวิชา วิทยาศาสตร์คอมพิวเตอร์และศาสตร์ด้านการคำนวณต่างๆ ตัวอย่างเช่น Dijkstra Algorithm ถูกคิดค้นโดยนักวิทยาศาสตร์คอมพิวเตอร์ Edsger Dijkstra ในปีค.ศ. 1956 เพื่อจุดประสงค์ในการ ค้นหาเส้นทางที่ใกล้ที่สุดระหว่างจุดหนึ่งไปยังจุดต่างๆ ในกราฟ ในขณะที่วิทยาศาสตร์มีความก้าวหน้าอย่างก้าวกระโดด โจทย์ปัญหาในศาสตร์ต่างๆ มักจะเกี่ยวข้องกับข้อมูลขนาดใหญ่และซับซ้อน ซึ่งจะเห็นได้ว่าปัญหาเหล่านี้มักจะถูกแปรให้อยู่ในรูปที่สามารถนำเอา อัลกอริทึมมาใช้งานได้ ตัวอย่างเช่น อัลกอริทึมที่ใช้ในการกระจายความเสี่ยงในการลงทุนได้ถูกนำมาใช้ในการเสนอผลลัพธ์ของระบบสืบค้น (Search Engine) ในงานวิจัยนี้ ผู้วิจัยได้นำเสนอวิธีการในการหาความสัมพันธ์ในเชิงวิวัฒนาการของแต่ละอัลกอริทึมในวรรณกรรมทาง วิทยาศาสตร์ ด้วยการใช้วิธีการ Ensemble Machine Learning จากผลการทดลอง วิธีการที่ นำเสนอได้ค่า F1 0.749 สำหรับการหาความสัมพันธ์เชิงละเอียด และ F1 0.905 สำหรับการหา ความสัมพันธ์แบบหยาบ องค์ความรู้ที่ได้จากงานวิจัยนี้สามารถนำไปต่อยอดเพื่อการพัฒนา งานวิจัยหรืองานประยุกต์ใช้อื่นๆ ในด้านวิทยาศาสตร์คอมพิวเตอร์และสารสนเทศศาสตร์ โดยเฉพาะการประยุกต์ใช้ในด้านสืบค้นและการทำเหมืองข้อมูลขนาดใหญ่จากห้องสมุด ดิจิทัล การศึกษาเครือข่ายอ้างอิงของอัลกอริทึม (Algorithm Citation Network) สามารถนำไป ตอบคำถามวิจัยหลายอย่างได้เช่น อัลกอริทึมไหนมีอิทธิพลมากที่สุด หรืออัลกอริทึมไหนมีความ เป็นทั่วไป (Generalizability) ซึ่งสามารถนำไปประยุกต์ใช้กับปัญหาในศาสตร์อื่นๆ ได้

คำหลัก: การอ้างอิงอัลกอริทึม, การเรียนรู้เครื่องทั้งมวล, ข้อมูลวิชาการขนาดใหญ่, วิวัฒนาการ ของอัลกอริทึม