รหัสโครงการ:	TRG5680090
ชื่อโครงการ:	การใช้เทคนิคการเรียนรู้ของเครื่องจักรเพื่อทำนายการมองของมนุษย์บนรูปภาพ
หัวหน้าโครงการ:	ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
อีเมล:	kitsuchart@it.kmitl.ac.th
ระยะเวลาโครงการ:	3 มิถุนายน 2556 ถึง 2 พฤศจิกายน 2559

บทคัดย่อ

การเก็บข้อมูลการเคลื่อนไหวของตานั้นมีค่าใช้จ่ายสูงและลำบาก อีกทั้งมีโอกาสบ่อยครั้งที่จะเกิดข้อมูลสูญหาย สมมติว่าเรากำลังจะเก็บข้อมูลการเคลื่อนไหวของตาจากผู้ใช้ต่าง ๆ ที่มองที่รูปภาพชุดหนึ่ง มีโอกาสเป็นไปได้ที่เราจะไม่สามารถเก็บข้อมูลได้ทุก ๆ ผู้ใช้ในแต่ละรูปภาพในชุดนั้น (หนึ่งหรือมากกว่าหนึ่งมุมมอง (View) อาจจะไม่ได้ถูกเก็บเพื่อเป็นตัวแทนของภาพ) เราสมมติว่าความสัมพันธ์ระหว่างมุมมองต่าง ๆ สามารถเรียนรู้ได้การเก็บมุมมองทั้งหมดหรือรายการ (Item) ดังนั้นงานนี้จะเป็นการสร้างข้อมูลขึ้นมาแทนในมุมมองที่ขาดหายไปของแต่ละรายการที่ไม่สมบูรณ์ โดยสร้างจากความสัมพันธ์ที่ได้มาจากรายการที่สมบูรณ์โดยใช้คุณสมบัติพีชคณิตของเทนเซอร์ (Tensor Algebra) เราได้แสดงว่า ปัญหานี้สามารถพิจารณาให้เป็นการเรียนรู้แบบถดถอย ยิ่งไปกว่านั้น เราใช้กรอบคิดของการหาค่าที่เหมาะสมโดยใช้พื้นฐานของการเรียนรู้ที่มีมาร์จินสูงสุด (Maximum Margin Learning) มาแก้ไขปัญหา ซึ่งวิธีนี้สามารถควบคุมได้ง่าย ปัญหานี้จะคล้ายกับการเรียนรู้ที่จะทำนายการมองของมนุษย์บนรูปภาพ โดยเราได้นำเสนออัลกอริทึมใหม่ที่เรียกว่า Tensor-based Multi-View Learning (TMVL) อีกทั้งได้นำเสนอเทคนิคในการเพิ่มประสิทธิภาพของการทำนายโดยใช้ลักษณะเด่นที่ได้มาจากการแยกองค์ประกอบโดยวิธีโครเนคเกอร์ (Kronecker Decomposition) ของรูปภาพรวมกับข้อมูลการเคลื่อนไหวของตา โดยการใช้ลักษณะเด่นที่นำเสนอนี้สามารถเพิ่มประสิทธิภาพในการทำนายได้อย่างมาก วิธีที่นำเสนอนี้ได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพมากขึ้นกว่าเทคนิคอื่น ๆ

คำหลัก: การเรียนรู้หลายมุมมอง; ข้อมูลสูญหาย; พีชคณิตของเทนเซอร์; การเรียนรู้ที่มีมาร์จินสูงสุด; การเคลื่อนไหวของตา; การแยกองค์ประกอบโดยวิธีโครเนคเกอร์.

Project Code: TRG5680090

Project Title: Predicting where Humans Look at in Images by Machine Learning Technique

Investigator: Asst. Prof. Dr. Kitsuchart Pasupa

King Mongkut's Institute of Technology Ladkrabang

Email address: kitsuchart@it.kmitl.ac.th

Project period: 3 June 2013 – 2 November 2016

Abstract

Eye movement data collection is very expensive and laborious. Moreover, there are usually missing values. Assuming that we are collecting eye movement data from a set of images viewed by different users, there is a possibility that we will not able to collect the data of every user from every image–one or more views may not be represented in the image. We assume that the relationships among the views can be learnt from the whole collection of views (or items). The task is then to reproduce the missing part of the incomplete items from the relationships derived from the complete items and the known part of these items. Using certain properties of tensor algebra, we showed that this problem can be formulated consistently as a regression type learning task. Furthermore, there is a maximum margin based optimisation framework in which this problem can be solved in a tractable way. This problem is similar to learning to predict where a person is looking in an image. Therefore, we proposed an algorithm called "Tensor-based Multi-View Learning" (TMVL) in this report. Furthermore, we also proposed a technique for improving prediction by introducing a new feature set obtained from Kronecker decomposition of the image fused with user's eye movement data. Using this new feature can improve prediction performance markedly. The proposed approach was proven to be more effective than two well-known saliency detection techniques.

Keyword: multi-view learning; missing data; tensor algebra; maximum margin learning; eye movements; Kronecker decomposition.