การหาความหมายภาพด้วยการวัดความคล้ายภาพด้วยเปรียบเทียบคู่ที่เหมาะสม

Semantic Images by Similarity Measure Images with Appropriate Pairwise comparisons

นัศพ์ชาณัณ ชินปัญช์ธนะ

Nutchanun Chinpanthana

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ 110/1-4 ถ.ประชาชื่น เขตหลักสี่ กรุงเทพฯ 10210

College of Innovative Technology and Engineering Dhurakij Pundit University 110/1-4 Prachachuen Rd. Laksi, Bangkok 10210,

E-mail nutchanun.cha@dpu.ac.th, Tel 02-954-7300

บทคัดย่อ

การค้นคืนภาพและการจำแนกภาพกำลังเป็นปัญหาที่น่าสนใจสำหรับการประมวลผลภาพ โดยส่วนใหญ่จะทำการใช้คุณลักษณะฟีเจอร์ที่ถูกสกัดออกมาด้วยวิธีการโครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อทำมาปรับปรุงการหาโมเลความหมายภาพด้วยการใช้คำหลักที่ถูกแท็กลงในภาพ ซึ่งส่วนใหญ่โมเดลที่ถูกสร้างมานั้นมักจะนำพาด้วยความสัมพันธ์ของคำหลักที่ถูกสกัดลงในภาพ ทำยังไม่สามารถแปลความหมายภาพได้อย่างแท้จริง ดังนั้นในงานวิจัยนี้พยายามแก้ปัญหาด้วยการวัดความคล้ายกันของภาพด้วยวิธีการเลือกคู่กราฟที่เหมาะสม โดยระบบทั้งหมดประกอบด้วย 4 ส่วนคือ (1) การเตรียมข้อมูล (2) การใส่ข้อมูลภาพ (3) การวัดความคล้ายกันของภาพด้วยวิธีการเลือกคู่กราฟที่เหมาะสม (4) การวัดและการประเมินผลการทำงาน จากการทดลองได้ทำการเปรียบกับเครื่องมือเปรียบเทียบจำแนกภาพทั่วไป เช่น ซัพพอร์ตเวกเตอร์แมกชีน, โครงข่ายประสาทเทียมแบบหลายชั้น โครงข่ายประสาทเทียมแบบสังวัตนาการ และ โครงข่ายประสาทเทียมแบบบริเวณสังวัตนาการ ผลที่ได้จากการทดลองแสดงให้เห็นว่าวิธีการวัดความคล้ายกันของภาพที่นำเสนอสามารถจำแนกภาพที่มีความหมายได้มากกว่าด้วยค่าความถูกต้องที่สุดกว่า

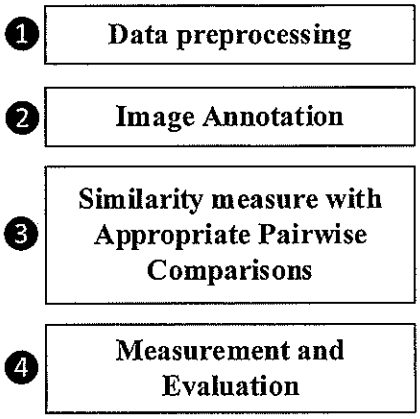
คำสำคัญ: การประมวลผลภาพ จับคู่กราฟ การค้นคืนภาพ ความหมายภาพ

Abstract

Recently, image retrieval and image classification are an active problem in digital image processing. Existing approaches typically focus on visual features extracted form an object-oriented convolutional neural network (CNN) and then improve semantic models by using keyword annotation. This model leads to an inadequate situation in the relation with keywords that are extracted from the images. It does not specific enough for representing the actual meaning. To overcome this problem, we suggest a new measure of similarity with appropriate pairwise comparisons (SAPC). The overall system process is divided into four steps: (1) data preprocessing; (2) image annotation; (3) similarity measure with appropriate pairwise comparisons and (4) measure and evaluation. The experiment is compared with traditional classification methods including Support Vector Machine Multiple Feedforward Neural Network, Convolutional Neural Networks, and Regions with Convolutional Neural Networks. The results show that our proposed approach offers significant performance improvements in the interpretation of semantic images, compared, with the higher accuracy.

ที่เข้ามาช่วยการค้นคืนความหมายภาพ แต่ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเป็นส่วนใหญ่

มีกลุ่มนักวิจัยอธิบายโครงสร้างความหมายในรูปแบบลำดับในเชิงการมองเห็น (semantic hierarchy for vision) [8] เพื่อนำเสนอความสัมพันธ์วัตถุภายในภาพเพื่อสร้างความสัมพันธ์ และหาความสัมพันธ์ของภาพจากคำหลักภายใน ความสอดคล้องกันด้วยการเรียนรู้ทำให้ครอบคลุมรูปแบบการค้นคืนภาพ บางกลุ่มใช้กราฟเพื่อสร้างความสัมพันธ์ของกราฟ ระหว่างข้อมูลสำหรับการจำแนกที่ดีขึ้นและยังไม่สามารถได้ความหมายที่ถูกต้อง ปัจจุบันงานวิจัยพยายามที่จะสร้างความสัมพันธ์ภายในภาพ [9] เช่น (girl, on, horse) หรือ (man, eat, apple) เพื่อตอบสนองการแปลความหมายภาพที่มีความซับซ้อนขึ้น บางวิจัยของได้นำเสนอวิธีการที่ทำการจับคู่วัตถุและหัวเรื่องและความสัมพันธ์ระหว่างกัน [10] ตัวอย่างเช่น ภาพที่ประกอบด้วย บุคคล (person), มอเตอร์ไซค์ (motorcycle) และ หมวกกันน็อค (helmet) สามารถสร้างความสัมพันธ์ได้เป็น (person - on – motorcycle), (person -wear – helmet) หรือ (motorcycle - has – wheel) แต่ในความเป็นจริงแล้วการวัดความคล้ายกันของภาพบนการใส่ความหมายวัตถุด้วยคำหลักเกิดความซ้ำซ้อนกันเนื่องจากการให้ความหมายของคำหลักมีความหมายเหมือนกัน (Synonym) เช่น “stone” กับ “rock” มีความหมายคล้ายกันคือหิน “people” กับ “human” มีความหมายคล้ายกันคือ คน , “kid”กับ“child” มีความหมายคล้ายกันคือเด็ก หรือ “hat”กับ “cap” มีความหมายคือหมวก เป็นต้น ดังนั้นจะเห็นว่าการพยายามใช้คำหลักแทนชื่อวัตถุนั้นยังเกิดข้อผิดพลาดซึ่งบางภาพอาจมีความหมายที่ตรงกันแต่การใช้คำหลักที่แตกต่างกันก็เป็นได้ดังนั้นการสร้างกลุ่มคำที่มีความเหมือนกันหรือกลุ่มคำที่มีความสัมพันธ์กับตามลำดับออนโทโลยี (Ontology) [11][12] เป็นโครงสร้างหนึ่งที่มีหลายกลุ่มวิจัยใช้ในการแทนรูปแบบการเชื่อมโยงของกลุ่มคำในรูปแบบกราฟลำดับขึ้นเพื่อให้ง่ายต้องการสร้างความเชื่อมโยงตามกลุ่มคำที่สัมพันธ์กัน เช่น “horse” กับ “camel” ตามลำดับขั้นที่ถูกเชื่อมโยงกัน แต่อย่างไรก็ตามการหาความสัมพันธ์ของลำดับขึ้นที่เหมือนกันตามกลุ่มคำหลักบนภาพ ยากที่จะเกิดความสัมพันธ์ ดังนั้นในงานวิจัยนี้



รูปที่ 2 ขั้นตอนการประมวลผลข้อมูลภาพ

จึงสร้างรูปแบบความสัมพันธ์ของความหมายภาพจากการวัดความคล้ายด้วยการเปรียบเทียบจากการจับคู่วัตถุที่เหมาะสม (similarity measure with appropriate Pairwise comparisons) เพื่อทำการจำแนกข้อมูลภาพให้ได้ตามกลุ่มความหมายอย่างแท้จริง

2. ขั้นตอนการประมวลผล

การเตรียมข้อมูลภาพเพื่อเข้าสู่กระบวนการประมวลผลจะแบ่งขั้นตอนวิธีการดำเนินงานวิจัย ออกเป็น 4 ส่วนหลักสามารถอธิบายดังรูปที่ 2 ดังนี้ (1) ขั้นตอนการเตรียมข้อมูล (Data preprocessing) (2) การใส่ข้อมูลภาพ (Image annotation) (3) การวัดความคล้ายกันของภาพด้วยวิธีการเลือกคู่กราฟที่เหมาะสม (Similarity measure with Appropriate Pairwise Comparisons) (4) การวัดและการประเมินผลการทำงาน (Measurement and Evaluation)

2.1 ขั้นตอนการเตรียมข้อมูล เริ่มจากคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลังภาพที่ถูกคัดเลือกเข้ามานั้น เป็นภาพที่มีความหมายภาพชัดเจน ภาพที่ถูกคัดเลือกออกไปไม่นำมาใช้ทดลองจะเป็นภาพที่มีลักษณะผิดปกติ (outlier) คุณลักษณะวัตถุไม่ชัดคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถบ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ทำให้ต้องมีการคัด

$$S_{ij} = \begin{cases} e^{\frac{\|p_i - p_j\|_2^2}{\sigma^2}} \; if \; \tilde{g}_i \in N_k(\tilde{I}_j), \tilde{g}_j \in N_k(\tilde{g}_j), \\ 0 \qquad\qquad otherwise \end{cases} \quad (2)$$

เมื่อกำหนดให้ σ คือจำนวนพารามิเตอร์ และ $\tilde{g}_i$ เป็นจำนวนระดับการเรียนรู้ของฟีเจอร์จากการวนซ้ำก่อนหน้าในการเรียนรู้ทั้งหมด และ $N_k(\tilde{g}_i)$ แทนด้วยชุดข้อมูลของ K-nearest neighbors ของ $\tilde{g}_i$

กำหนดให้ ลาปลัสกราฟ (Graph Laplacian) $\mathbf{L} = \mathbf{D} - \mathbf{S}$, เมื่อ $\mathbf{D}$ แทนเมทริกซ์ทแยงมุม (Diagonal Matrix) ที่กำหนดให้มีอิลิเมนต์เป็น $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{ij}$, จากสมการที่ (1) สามารถเขียนใหม่เป็น

$$P_\delta = tr(\mathbf{G}^T \mathbf{L} \mathbf{G}), \quad (3)$$

$$\frac{\partial P_\delta}{\partial \mathbf{G}} = 2\mathbf{L}\mathbf{G}. \quad (4)$$

การเขียนสมการใหม่เพื่อลดทอนสมการบางตัวที่ไม่จำเป็นและบางโครงสร้างถูกสืบทอดมาและอยู่ในการเรียนรู้ของทั้งระบบแล้ว

5. การวัดและประเมินผลการทำงาน

การวัดและประเมินผลการทำงาน (Measurement and Evaluation) จะเป็นการทดสอบการจำแนกความหมายภาพที่มีวัตถุในภาพเป็นหลักจากฐานข้อมูลจาก PASCAL action และ [13]

5.1 ข้อกำหนดและฐานข้อมูล

สำหรับข้อมูลภาพที่ใช้ในการทดลองทั้งหมด 1,500 ภาพ คัดเลือกฐานข้อมูลภาพแอ็คชันจาก PASCAL action [13] ข้อมูลภาพประกอบด้วย 12 วัตถุดังนี้ bike, bird, boat, car, cat, chair, table, dog, horse, person, plant และ sofa ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะไม่ถูกนำมาพิจารณาและจำแนกข้อมูลภาพให้อยู่ในหมวดหมู่ตามความหมายภาพ เช่น “birthday party”, “leisure”, “couple love”, “sport racing”, “office working”, “playground”, “graduation ceremony”, “city” เป็นต้น

5.2 วิธีการวัดประสิทธิภาพ

การประเมินด้วยการวัดประสิทธิภาพของการจำแนกภาพด้วย ซัพพอร์ตเวกเตอร์แมกชีน (Support Vector Machine : SVM) [15][16][17], โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network :MLPN) โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Networks : CNN) [18][19][20] โครงข่ายประสาทเทียมแบบบริเวณสังวัตนาการ (Regions with Convolutional Neural Networks: RCNN) และ นำเสนอวิธีการวัดความคล้ายกันของภาพด้วยวิธีการเลือกคู่กราฟที่เหมาะสม (Similarity measure with Appropriate Pairwise Comparisons: SAPC) การพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และ ค่าเฉลี่ยของความแม่นยำ

1. ค่าความแม่นยำ (False positive rate / Precision: Pr.) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหามาได้

$$Pr = \varepsilon / \partial \quad (5)$$

เมื่อกำหนดให้ ε แทนจำนวนข้อมูลที่เกี่ยวและถูกดึงออกมาได้อย่างถูกต้อง และ ∂ แทนจำนวนข้อมูลที่ถูกดึงออกมาทั้งหมด

2. ค่าความระลึก (True positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพที่ถูกต้องทั้งหมด

$$Re = \varepsilon / \sigma \quad (6)$$

เมื่อ σ แทนจำนวนข้อมูลที่เกี่ยวข้องทั้งหมด

3. ค่าเฉลี่ยของความแม่นยำ (mean average precision: MAP) เป็นค่าเฉลี่ยความแม่นยำของคำหลักค้นหลายคำที่เกี่ยวข้องกัน (relevance) และนำไปจัดอันดับ โดยมีความสัมพันธ์กับคิวรี (query) ค่าเฉลี่ยของความแม่นยำเมื่อกำหนดค่าสูงสุดของ k ที่มีความเกี่ยวข้องกับข้อมูลภายในที่ถูกค้นคืน สามารถเขียนเป็นกลุ่มของเซตของข้อมูลที่

ตารางที่ 3 การจำแนกความหมายภาพด้วยความสัมพันธ์ของวัตถุด้วย Data Set I

Categories/Method	SVM		MLPN		CNN		RCNN		SAPC	
	Prec.	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*
office working	54.0	50.5	65.3	65.3	71.6	69.4	75.7	75.0	83.0	85.6
city	51.5	46.8	65.7	65.0	75.8	70.6	70.1	75.8	79.2	82.4
birthday party	49.5	46.7	67.0	67.0	71.4	69.4	70.6	72.0	77.6	78.4
couple love	48.2	52.0	65.1	67.6	65.8	70.9	71.2	73.3	78.4	76.0
grad. ceremony	46.7	42.2	62.1	54.0	72.5	71.8	71.4	70.7	79.1	65.5
playground	46.1	47.5	67.4	62.6	63.2	67.7	73.2	71.0	69.5	75.3
family time	49.0	51.5	66.3	67.0	65.1	65.1	75.3	70.2	76.9	79.5
sport game	42.7	50.0	58.1	67.3	73.3	72.5	77.2	76.5	76.6	78.8
Accuracy	**48.37**		**64.49**		**69.67**		**73.05**		**77.48**	

ตารางที่ 4 การจำแนกความหมายภาพด้วยความสัมพันธ์ของวัตถุด้วย Data Set II

Categories/Method	SVM		MLPN		CNN		RCNN		SAPC	
	Prec.	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*	*Prec.*	*Recall*
office working	54.2	52.7	68.4	65.0	73.5	72.0	74.7	72.4	80.8	83.2
city	53.8	55.4	68.5	63.0	70.4	69.0	70.7	70.0	81.0	81.0
birthday party	53.6	48.6	68.7	68.0	70.6	72.7	72.0	71.3	75.3	77.7
couple love	50.0	51.4	64.8	69.3	69.4	75.0	71.2	73.1	79.6	78.0
grad. ceremony	45.7	41.6	65.2	58.6	75.3	70.0	73.0	72.3	78.7	66.1
playground	42.7	40.6	63.1	65.7	74.7	70.3	71.2	74.0	69.1	76.0
family time	49.5	50.0	65.4	68.7	69.9	72.0	75.3	73.0	78.7	73.3
sport game	43.3	52.0	64.2	69.3	72.8	75.0	77.5	79.0	75.9	84.2
Accuracy	**49.09**		**65.96**		**72.00**		**73.14**		**77.26**	

เปรียบเทียบ การจำแนกด้วยวิธีการ ได้ค่าความถูกต้องเพียง 48.37%, 64.49%, 69.67% และ 73.05% ด้วยวิธีการ SVM, MLPN, CNN และ RCNN และด้วยวิธีนำเสนอการวัดค่าความแตกต่างด้วยกราฟ SAPC สามารถได้ค่าเฉลี่ยความถูกต้องสูงถึง 77.48% ในชุดข้อมูล Data Set I และเช่นเดียวกัน การจำแนกด้วย SAPC สามารถจำแนกได้ค่าความถูกต้องเฉลี่ยสูงถึง 77.26% กับชุดข้อมูล Data Set II แสดงให้เห็นว่าการใช้ความสัมพันธ์ของวัตถุที่เกิดขึ้นภายในภาพ และการวัดค่าความแตกต่างมีผลต่อประสิทธิภาพในการจำแนกข้อมูลตามความหมายภาพ แต่ถ้ามีการนำมาใช้เพียงจำแนกวัตถุจะได้ค่าความถูกต้องที่น้อยกว่าอย่างเด่นชัด

จากการทดลองที่ใช้ชุดข้อมูลเดียวกันแต่มีการแยกตามกลุ่มวัตถุ และ กลุ่มความหมายของภาพด้วยการวัดประสิทธิภาพเดียวกัน SVM, MLPN, CNN และ RCNN และด้วยวิธีนำเสนอการวัดค่าความแตกต่างด้วยกราฟ SAPC จะเห็นว่าเมื่อมีการแทนข้อมูลภาพด้วยกราฟและใช้การวัดประสิทธิภาพด้วย SAPC จะเห็นว่าได้ค่าความถูกต้องที่สูงกว่าถึง 77.48%, 77.26% ในชุดข้อมูล Data set I และ Data set IIตามลำดับ

7. สรุป

งานวิจัยนี้ได้นำเสนอรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพและการสร้างรูปแบบความสัมพันธ์ของความหมายภาพจากการ

[14] Hanwang Zhang, Zawlin Kyaw, Shih-Fu Chang, Tat-Seng ChuaVisual. (2017). Translation Embedding Network for Visual Relation Detection. Conference: Conference: 2017 IEEE Conference on Computer Vision and Pattern: 5532-5540.

[15] B. Sch¨olkopf and A. Smola. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge. MA.

[16] O. Chapelle, P. Haffner, and V. Vapnik. (1999). Support vector machines for histogram-based image classification. NN.

[17] S. Melacci and M. Belkin. (2011). Laplacian support vector machines trained in the primal. Journal of Machine Learning Research. 12: 1149–1184.

[18] Joachims, T. (1998). Text categorization with support vector machines—learning with many relevant features. In Proceedings of the 10th European Conference on Machine Learning. Chemnitz. Germany. (Berlin: Springer): 137–142.

[19] Y. Tsuruoka, J. Tsujii, S. Ananiadou. (2009). Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty. In Proceedings of the AFNLP/ACL.

[20] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems: 91-99.