

# 잡지기사 관련 상품 연계 추천 서비스를 위한 하이퍼네트워크 기반의 상품이미지 자동 태깅 기법

하정우<sup>○</sup> 김병희 이바도 장병탁

서울대학교 전기컴퓨터공학부

[jwha@bi.snu.ac.kr](mailto:jwha@bi.snu.ac.kr), [bhkim@bi.snu.ac.kr](mailto:bhkim@bi.snu.ac.kr) [bdlee@bi.snu.ac.kr](mailto:bdlee@bi.snu.ac.kr), [btzhang@bi.snu.ac.kr](mailto:btzhang@bi.snu.ac.kr)

## Auto-tagging Method for Unlabeled Product Images with Hypernetworks for Article-related Product Recommender System

Jung-Woo Ha<sup>○</sup> Byoung-Hee Kim Bado Lee Byoung-Tak Zhang

School of Computer Science and Engineering, Seoul National University

### 서 론

최근 새로운 상품 추천 서비스로서 잡지 기사검색과 온라인 상품 쇼핑을 연계 추천하는 서비스가 새롭게 등장하고 있다. 이 서비스에서는 사용자가 온라인상으로 잡지 기사를 검색하면 기사 하단에 잡지 기사의 내용과 관련된 상품이 자동으로 연계되며 이를 기사를 검색한 사용자에게 추천하고 사용자가 상품 이미지를 클릭했을 때 쇼핑몰의 구매 사이트의 페이지로 연결된다. 일반적으로 쇼핑몰에서 상품을 추천할 때 가장 널리 활용되는 방법은 협력적 여과 (collaborative filtering)이다[1]. 그러나 협력적 여과는 상품의 컨텍스트 자체 보다는 상품과 사용자와의 이력 관계를 기반으로 추천 여부를 결정하기 때문에 이 서비스에서 적용되기에는 제한적이다. 그러므로 잡지 기사 컨텍스트와 상품 정보간의 연관성을 모델링할 새로운 방법이 필요하다. 특히 잡지 기사와 상품 정보 모두 멀티모달(multimodal) 특성을 갖는 데이터이므로 모달리티 간의 크로스모달(cross-modal) 연관성을 학습하는 기술[2-5]이 필요하다. 이를 위해 본 연구진은 멀티모달 연상 추천시스템(MARS) 프로젝트를 통하여 이미지와 텍스트 문서로 구성된 잡지기사의 컨텍스트와 이미지와 메타정보로 구성된 상품 정보간의 멀티모달 연상 연관관계를 모델링하는 방법을 연구 중이며[6] 본 논문에서는 해당 기술의 한 구성요소로서 상품의 이미지 정보를 이용하여 상품의 종류를 분류하고 이를 태그로서 상품에 부여하는 방법을 제안 한다.

### 본 론

본 연구에서 사용된 하이퍼네트워크 모델은 데이터를 구성하는 인자(feature)들 간의 고차 연관관계를 표현 및 학습 가능한 확률그래프 모델 기반의 기계학습 방법론이다[7]. 하이퍼네트워크의 특성은 회상기억 (Recall memory), 전역지역성(Glocality), 조합성(Compositionality), 자가조립(Self-assembly) 등으로 요약된다. 본 연구에서는 시각언어(visual words)의 벡터로 표현된 상품이미지 데이터 집합이 주어지면, 랜덤 샘플링 기반으로 생성된 하이퍼에지들로 초기 하이퍼네트워크  $H$ 를 구성한다. 그리고 훈련데이터와의 비교를 통해 하이퍼에지들의 가중치가 학습이 된다. 그리고 학습된 하이퍼네트워크  $H'$ 를 이용하여 훈련데이터의 분류 성능을 측정한다. 그 후 가중치가 낮은 일부 하이퍼에지는 하이퍼네트워크로부터 제거되고 새로운 하이퍼에지가 생성되어 하이퍼네트워크에 포함된다. 이러한 과정을 정해진 회수만큼 반복하여 학습이 진행된다. 이때 학습과정에서 하이퍼에지  $E_j$ 의 가중치  $w_j$ 를 계산하는 수식은 다음과 같이 정의된다.

$$w_j = \frac{\alpha}{\#_w} + \beta \times \#_c \quad (1)$$

식 (1)에서  $\alpha$ 와  $\beta$ 는 임의의 상수로서 덜 틀린 경우와 더 많이 맞추는 경우에 대하여 가중치에 영향을 주는 요소이다. 본 연구의 실험을 위해 사용된 데이터는 디자인시계, 아동용완구 등 디자인제품과 와인잔, 식기 등을 포함한 주방기기 등 총 10개의 카테고리 1,251개의 이미지로 구성되어 있다. 각각의 상품 이미지는 전처리로서 Surf [8]와 mean-shift 클러스터링을 기반으로 하여 추출된 4,022개의 시각언어 인자 집합 (feature set) 중에서 모든 샘플에서 동일한 값을 갖는 인자들을 제외한 81개의 인자들에 대하여 히스토그램 벡터로 표현되며, 표현된 히스토그램은 TF-IDF를 이용하여 정규화한다. 그리고 마지막으로 하이퍼네트워크 모델은 이산화된 데이터에 대해서만 효과적으로 학습이 가능하므로 정규화된 시각언어 벡터 데이터를 0에서 4까지 5단계로 이산화 표현하였다. 일반화 성능을 검증하기 위해 전체 데이터 샘플

중 70%를 훈련데이터로 활용하고 나머지 30%를 테스트 데이터로 사용하였으며, 실험 결과의 재현성을 위하여 매 실험마다 훈련 데이터와 테스트 데이터의 구성이 변하도록 실험하였다. 또한 비교분석을 위하여 Weka 3.6.1[9]에 구현된 알고리즘을 사용하였다. 그리고 각 상품에 부여되는 태그의 수가  $m$ 인 경우 ‘ $m$ 차 태깅’이라고 정의한다. 마지막으로 이 실험에서 사용된 하이퍼네트워크의 파라미터는 표 1과 같다. 표 2는 하이퍼네트워크 기반의 상품이미지 태깅 정확도를 일반적으로 널리 활용되는 기계학습 기법을 이용한 태깅 결과와 비교한 것이다. 표 2에서 알 수 있듯이 하이퍼네트워크는 SVM과 Bayesian Network와 비교했을 때 경쟁력 있는 태깅 성능을 보이며 그 외의 알고리즘들에 비해서 우수한 태깅 정확도를 보여주고 있다. 그림 1은 3차 태깅이 부여된 상품의 이미지의 예를 보여주고 있으며 이를 통해 상품이미지가 나타내는 컨텍스트와 부여된 태그들 사이에는 의미의 유사성이 있음을 알 수 있다.

표 1. 하이퍼네트워크 파라미터

파라미터	값	
샘플링 수	20	
하이퍼에지 교체 비율	10%	
반복실험 횟수	10회	
가중치	$\alpha$	10
	$\beta$	0.0001

표 2. 다른 알고리즘과의 태깅 정확도 비교. 비교된 기계학습 모델의 정확도는 10회 실행 후 평균 값

기계학습 (세부 세팅)	1차 태깅 정확도(%)	표준편차
<b>30-하이퍼네트워크</b>	<b>36.15</b>	<b>2.44</b>
SVM (SMO + 2차 다항식 커널)	35.24	1.95
Decision Tree (J48)	28.33	2.10
Random Forest	29.51	1.68
kNN (k=1)	30.6	2.71
Bayesian Network (K2, #Parents =3)	35.78	2.28
Naive Bayes	33.48	1.49



그림 1. 3차 태깅된 상품이미지 예. 밑줄 표시가 된 태그가 실제 상품의 종류임.

감사의 글

이 논문은 지식경제부 및 한국산업기술평가관리원의 IT산업원천기술개발사업(KI002138, 차세대 맞춤형 서비스를 위한 기계학습 기반 멀티모달 복합 정보 추출 및 추천기술 개발, MARS), 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 연구(314-2008-1-D00377, Xtran 및 No.2010-0017734) 및 교육과학기술부의 BK21-IT사업에 의해 지원되었음.

참고문헌

[1] Vucetic, S. and Obradovic, Z., A regression-based approach for scaling-up personalized recommender systems in e-commerce. In *ACM WebKDD 2000 Web Mining for E-Commerce Workshop*, 2000.

[2] Li, D. Dimitrova, N. Li, M., and Sethi, K., Multimedia content processing through cross-modal association, In *Proc. of the 11th annual ACM International Conference on Multimedia (MM'03)*, 604-611. 2003.

[3] Fuster, J. M. Bodner, M. and Kroger, J. K., Cross-modal and cross-temporal association in neurons of frontal cortex, *Nature*, 405, 347-351. 2000.

[4] Goh, K.-S. Chang, E. Y. and Lai, W.-C., Multimodal concept-dependent active learning for image retrieval, In *Proc. of the 12th annual ACM international conference on Multimedia (MM'04)*, 564-571. 2004.

[5] Quek, F. McNeil, D. Bryll, R. Duncan, S. Ma, X.-F. Kirbas, C. McCullough, K. E. and Ansari, R., Multimodal human discourse: gesture and speech, *ACM Trans. on Computer-Human Interaction*, 9(3), 171-193. 2002.

[6] Ha, J.-W. Kim, B.-H. Kim, H.-W. Yoon, W. C. Eom, J.-H. and Zhang, B.-T., Text-to-image cross-modal retrieval of magazine articles based on higher-order pattern recall by hypernetworks. *The 10th International Symposium on Advanced Intelligent Systems (ISIS 2009)*, 274-277. 2009.

[7] Zhang, B. -T., Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *IEEE Computational Intelligence Magazine*, 3(3), 49-63, 2008.

[8] Bay, H. Tuytelaars, T. and Gool, T. V., Surf: Speed up robust features, In *Proc. of the 9th European Conference on Computer Vision (ECCV '06)*, 404-417. 2006.

[9] University of Waikato New Zealand, Waikato Environment for Knowledge Analysis (Weka) Version 3.6.1, Available: <http://www.cs.waikato.ac.nz/ml/weka/index.html>