

폭소노미에서 이미지 자동 태깅을 위한 사회적 관계 추출에 관한 연구

엄원용*, 이시형*, 노용만*¹

*한국과학기술원 영상 및 비디오 시스템 연구실

e-mail : ewony@kaist.ac.kr

Study for social relationship extraction for automatically image tagging in Folksonomy

Wonyong Eom*, Sihyoung Lee*, Yong Man Ro*

*Image and Video Systems Lab., Korea Advanced Institute of Science and Technology

요 약

멀티미디어 기기의 확산과 인터넷의 발달로 Flickr, Facebook 과 같은 사회적 네트워크를 기반으로 이미지 공유가 활발해졌다. 사회적 네트워크 사이트에서 이미지의 효율적인 검색과 관리를 위해서 태그를 이용하는 방법이 많이 사용되고 있다. 하지만 많은 양의 이미지에 수동으로 태그를 등록하는 것은 사용자에게 많은 시간과 노력을 요구한다. 태그 추천 기술은 자동으로 사용자에게 태그를 추천함으로써, 수동 태깅의 한계를 극복할 수 있는 방법이다. 본 논문에서는 사회적 네트워크를 기반으로 하는 폭소노미에서 사용자 사이의 사회적 관계를 사용자 들의 얼굴 정보를 이용하여 측정하고, 이를 활용하여 이미지 태그를 추천하는 기술을 제안한다. 제안하는 방법은 이미지의 시각 정보와 태그 분포뿐만 아니라 사용자 사이의 사회적 관계 정보를 추가로 활용한다. 실험을 통해서 제안하는 방법이 기존의 이미지 태그 추천 방법에 비해서 7% 향상된 태그 추천의 정확성을 보장하는 것을 증명하였다.

1. 서론

최근 멀티미디어 기술의 발전과 인터넷의 보급으로 인해서 웹을 통해서 이미지를 공유하는 환경이 확산되고 있다. Flickr [1]와 Facebook [2]은 온라인을 통한 이미지 공유를 지원하는 대표적인 사회적 네트워크 (social network) 사이트로써, Flickr 의 경우 매년 수천장의 이미지가 새롭게 등록되고 있다 [3]. 이처럼 급속하게 증가하고 있는 이미지의 효율적인 검색 및 관리 방법이 중요한 문제로 인식되고 있다.

사회적 네트워크 사이트는 태깅 (tagging)을 통해서 이미지의 검색 및 관리를 수행한다. 태깅은 사용자가 이미지와 관련된 태그 (tag)를 검색을 목적으로 이미지와 함께 사회적 네트워크에 등록하는 것을 의미한다. Flickr 와 Facebook 등의 사이트도 태그를 통한 이미지 검색이 보편적이다. 그리고 사회적 네트워크 사이트를 통한 협력 태깅 (collaborative tagging)을 통해서 폭소노미 (folksonomy)를 구성한다.

태깅을 통해서 이미지에 태그를 등록하는 것은 효율적인 이미지 검색과 관리에 필수적이다. 그러나, 현재 사회적 네트워크 사이트는 사용자가 수동으로 태그를 등록하는 방법에 의존하고 있다. 많은 양의 이미지에 태그를 등록하는 것은 사용자에게 많은 노력

과 시간을 요구한다. 그래서 Flickr 에는 전체 이미지 중에서 태그가 등록되어 있지 않거나 매우 적은 수의 태그가 등록되어 있는 이미지의 비율이 80%가 넘는다 [4].

이미지 태그 추천 (image tag recommendation) 기술은 수동 태깅의 한계를 극복할 수 있는 방법으로써, 이미지의 특징 분석을 통해서 해당하는 이미지와 관련된 있는 태그를 자동으로 사용자에게 추천한다. 사용자의 행위는 이미지 태그 추천 기술에 의해서 추천된 태그 중에서 이미지와 가장 적합한 태그를 선택하는 과정으로 축소된다. 이미지 태그 추천 기술은 특정 태그의 이미지 특징 모델링을 통한 모델 기반 (model based) 방법과 모델링을 기반으로 하지 않는 모델 자유 (model free) 방법으로 분류할 수 있다. 모델 기반 태그 추천 방법은 트레이닝 이미지로부터 추출한 하위 특징 (low-level feature)을 통해서 미리 정해놓은 숫자의 높은 수준 컨셉 (high-level concept)을 모델링하고, 이것을 통해서 태그를 추천하는 방법이다. 모델 기반 방법은 미리 지정해 놓은 제한된 개수의 컨셉에 대해서만 태그를 추천할 수 있다는 것과 트레이닝을 해야 한다는 단점이 있다 [5]-[7].

모델 자유 태그 추천 방법은 태그를 추천 받으려는 이미지의 저 수준 특징이나 사용자가 과거에 등록했

¹ Correspondence: ymro@ee.kaist.ac.kr

던 태그들의 확률 혹은 폭소노미 내의 태그 분포 등의 다양한 방법을 이용하여 태그를 추천하는 방법이다. 이 방법은 트레이닝을 해야 할 필요가 없고, 추천할 수 있는 태그의 수가 제한되어 있지 않다. [4]은 이미지에 이미 등록된 소수의 초기 태그를 기반으로 추가로 추천될 태그를 제시한다. 초기 태그와 다른 태그 사이의 공기 (co-occurrence) 정보를 측정하고, 이것을 통해서 추천될 태그를 결정한다. 하지만, 이 방법은 초기 태그가 없으면 태그 추천을 할 수 없을 뿐만 아니라, 이미지의 시각적 특징을 고려하지 못한다. [8]는 폭소노미에 존재하는 태그의 분포와 이미지의 시각 특징을 통해서 태그를 추천한다. 먼저 입력 이미지와 시각적으로 유사한 이미지를 하위 특징 비교를 통해서 획득한다. 그리고 이렇게 획득한 유사 이미지에 등록된 태그의 빈도를 바탕으로 태그를 추천한다. 이 방법은 시각 정보를 활용한 태그 추천은 가능하지만, 태그들간의 관계를 고려하지 못한다. [9]는 폭소노미를 통한 태그 추천 과정을 MAP (maximum a posteriori)을 통해서 모델링 하였다. 입력 이미지와 폭소노미의 이미지 사이의 시각 특징 분포와 태그의 분포를 확률적으로 계산하여 태그를 추천한다. 기존의 태그 추천 기술은 이미지의 시각적 특징과 같은 이미지 자체의 특징, 또는 태그 분포와 같은 정보만 활용한다.

최근 많은 사회적 네트워크 사이트에서는 사용자간의 관계가 발달되고 있다. 특히 사회적 관계 (social relationship)가 높은 사용자간에는 유사한 관심 영역이나 취미를 갖고 있거나 실제 사회에서 직장 동료나 친구와 같은 사이일 확률이 높다. 다시 말해 사회적 관계가 높은 사용자일수록 유사한 의미의 이미지 데이터를 가지고 있을 확률이 높다 [10]-[13]. 그러나 기존의 태그 추천은 사회적 네트워크에 참여하는 사용자 사이의 사회적 관계를 활용하지 않은 접근이 대부분이었다. 그렇기 때문에 실제로 사람들이 유사성이 없다고 느끼는 이미지들도 비슷한 것으로 분류되는 경우가 많았다. 이에 본 논문에서는 폭소노미에서 사용자 간의 사회적 관계 (social relationship)를 이용하여 태그 추천의 정확도를 향상 시키는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 장에서는 얼굴 정보를 통한 사회적 관계를 모델화 하는 방법을 설명하고, 3 장에서는 사회적 관계 모델을 이용한 태그 추천 방법을 설명한다. 그리고 4 장에서는 실험을 통해서 사회적 관계를 활용한 태그 추천 방법이 기존 태그 추천 방법보다 태그 추천의 정확성을 향상 시켰음을 보인다. 마지막으로 5 장에서 본 논문의 결론을 내린다.

2. 얼굴 정보를 통한 사회적 관계 추출

이미지 q 는 태그 추천 알고리즘을 통해서 태그를 추천 받을 이미지를 의미하고, u_q 는 이미지 q 를 획득한 사용자로 정의한다. 그리고 사용자 u_q 와 u_n 사이의 사회적 관계의 크기는 $w_{q,n}$ 로 정의한다. 본 논문에서 사용자 사이의 사회적 관계는 폭소노미에 등록된 이미지에 존재하는 사용자 얼굴의 공기 정보 (co-

occurrence information)를 통해서 모델링한다. 두 사용자 사이의 사회적 관계의 크기 $w_{q,n}$ 는 다음과 같이 수식화한다.

$$w_{q,n} = \exp(P(u_n) + P(u_q, u_n)), \quad (1)$$

여기서 $P(u_n)$ 은 폭소노미에 등록된 이미지에서 사용자 u_n 의 얼굴이 나타난 확률을 의미하고, $P(u_q, u_n)$ 은 폭소노미의 이미지 중에서 사용자 u_q 와 u_n 의 얼굴이 동시에 나타난 이미지의 확률을 의미한다. 수식 (1)은 앞에서 설명한 두 확률의 합에 지수함수를 취해 합이 클수록 합이 작은 것과의 차이를 크게 해준다.

식 (1)의 첫 번째 항은 다음과 같이 정의 할 수 있다.

$$P(u_n) = \frac{\sum_{i \in I_{u_q}} \Delta_1(u_n, i)}{|I_{u_q}|}, \quad (2)$$

여기서, I_{u_q} 는 사용자 u_q 가 소유하고 있는 이미지 집합을 의미하고, $|\cdot|$ 는 집합의 원소 수를 나타낸다. $\Delta_1(u_n, i)$ 는 이미지 i 에 사용자 u_n 이 나타나면 1, 아니면 0인 값을 갖는 함수이다.

식 (1)의 두 번째 항은 식 (3)과 같이 정의된다.

$$P(u_q, u_n) = \frac{\sum_{i \in I} \Delta_2(u_q, u_n, i)}{|I|}, \quad (3)$$

식 (3)에서 I 는 주어진 폭소노미에 등록된 이미지 집합을 의미하고, $\Delta_2(u_q, u_n, i)$ 는 이미지 i 에 u_q 와 u_n 의 사용자 모두가 나타나면 1을 그렇지 않은 경우 0인 값을 갖는 함수이다.

3. 사회적 관계를 통한 태그 추천 방법

본 절에서는 앞에서 설명한 사회적 관계를 활용한 이미지 태그 추천 방법을 제안한다. 이를 위해서 먼저 시각 폭소노미를 정의하고 정의된 폭소노미를 통한 태그 추천 방법을 설명한다.

3.1 제한된 시각 폭소노미 정의

시각 폭소노미 (visual folksonomy) \mathcal{F} 는 사용자 집합 U , 이미지 집합 I , 태그 집합 T 그리고 이들 사이의 관계를 정의하는 어사인먼트 (assignment) 집합 A 로 구성된다. 어사인먼트 a 는 사용자 u 와 이미지 i 그리고 태그 t 들의 관계를 나타내 준다.

본 논문에서는 폭소노미 \mathcal{F} 로부터 제한된 시각 폭소노미 \mathcal{G} 와 \mathcal{Q} 를 정의한다. 이미지 q 가 태그를 추천 받으려 하는 이미지라고 한다면, 먼저 \mathcal{G} 는 사용자 u_q 와 사회적 관계 $w_{q,n}$ 가 임계 값 (threshold value) 이상의 값을 갖는 사용자들 기반으로 구성된 폭소노미를 의미한다. 폭소노미 \mathcal{Q} 는 \mathcal{G} 내에서 q 와 시각적으로 유사한 이미지를 중심으로 구성된 폭소노미로 정의한다.

3.2 폭소노미를 통한 태그 추천 방법

이미지 q 에 추천되는 태그를 \hat{t} 로 정의하면, 폭소노미 내의 태그 t 가 \hat{t} 가 될 확률은 식 (4)와 같다.

$$\hat{t} = \arg \max_t \{P(t|q, \mathcal{Q}, \mathcal{G})\}. \quad (4)$$

식 (4)의 확률은 베이즈 룰 (Bayes' rule)을 이용하여 아래와 같이 표현할 수 있다.

$$P(t|q, \mathcal{Q}, \mathcal{G}) = \frac{P(q|t, \mathcal{Q}, \mathcal{G})P(t|\mathcal{Q}, \mathcal{G})}{P(q|\mathcal{Q}, \mathcal{G})}, \quad (5)$$

여기서, 분모 항은 t 에 관한 함수에서 보면 상수로 취급할 수 있다. 따라서 식 (5)는 식 (6)으로 나타낼 수 있다.

$$P(t|q, \mathcal{Q}, \mathcal{G}) \cong P(q|t, \mathcal{Q}, \mathcal{G})P(t|\mathcal{Q}, \mathcal{G}), \quad (6)$$

여기서, 첫 번째 항은 우도 확률 (likelihood probability) 두 번째 항은 사전 확률(prior probability)를 의미한다. 첫 번째 항인 우도 확률은 \mathcal{Q} 와 \mathcal{G} 를 만족하며 특정한 t 가 등록된 이미지가 q 일 확률을 의미 한다. 여기서 \mathcal{Q} 와 \mathcal{G} 를 만족하면서 특정한 t 가 등록되어 있다는 조건을 만족하는 것은 \mathcal{R} 이다. 이 때 \mathcal{R} 내의 이미지들의 특징과 이미지들의 평균 특징 벡터 사이의 거리가 가우시안 분포 (Gaussian distribution)를 나타낸다 [9]. 이를 근거로 하여 첫 번째 항인 우도 확률은 식 (7)과 같이 풀어낼 수 있다.

$$P(q|t, \mathcal{Q}, \mathcal{G}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(d(q, \bar{i}_t) - \mu_t)^2}{2\sigma_t^2}\right), \quad (7)$$

여기서, \bar{i}_t 는 \mathcal{Q} 에서 태그 t 를 가지고 있는 이미지의 특징 벡터 평균이며, $d(q, \bar{i}_t)$ 는 q 의 특징 벡터와 \bar{i}_t 간의 거리를 나타낸다. μ_t, σ_t^2 는 각각 q 의 특징 벡터와 \mathcal{Q} 에서 태그 t 를 가지고 있는 이미지의 사이의 특징벡터 거리의 평균과 분산을 의미한다. 식 (6)에서 두 번째 항인 사전 확률은 폭소노미 \mathcal{Q} 에서의 특정 태그 t 의 확률을 의미하며 식 (8)과 같이 나타낼 수 있다.

$$P(t|\mathcal{Q}, \mathcal{G}) = \frac{|T_q \cap T_t|}{|T_q|}, \quad (8)$$

여기서, T_q 는 폭소노미 \mathcal{Q} 를 구성하는 태그 집합을 나타내고, T_t 는 폭소노미 \mathcal{F} 에서 태그 t 의 집합을 의미한다. $|\cdot|$ 는 집합의 원소 수를 의미한다.

4. 실험

본 절에서는 제안된 방법의 우수성을 증명하기 위하여 폭소노미를 구성하고 실험 결과를 어떤 방법으로 수치화 시킬지를 정하며 그 성능을 사회적 관계를 사용하지 않고 이미지 태그를 추천하는 방법 [9]과 비교한다.

4.1 실험 환경

본 논문에서 수행하는 실험을 위해서 한국의 대표적인 사회적 네트워크 웹사이트 중 하나인 싸이월드 [14]를 통해서 폭소노미를 구축하였다. 본 논문에서는 얼굴을 이용한 사회적 관계를 구축 하기 위해 얼굴이 존재하는 이미지를 이용하여 폭소노미를 구축하였다. 구축한 폭소노미는 30 명의 사용자, 4,010 장의 이미지

그리고 17,103 개의 태그로 구성되어 있다. 30 명의 사용자중 2 명의 사용자를 선택해 u_{q_A} 와 u_{q_B} 로 지정하고 이들이 업로드한 이미지 348 장을 테스트 이미지로 사용하여 제안된 태그 추천 방법의 성능을 측정한다. 이 때 각 u_q 들에 대해 \mathcal{G} 를 구성하는 사용자들의 수와 이미지 데이터 수, 그리고 추천 받고자 하는 테스트 이미지 q 의 수는 <표 1>과 같다. 이 때 기준 값은 각 u_q 와 다른 29 명의 사용자 간의 평균 $w_{q,n}$ 값을 이용 하였다.

<표 1> u_q 에 대한 \mathcal{F}, \mathcal{Q} 의 이미지 데이터 숫자와 테스트 이미지 q 의 개수

User	The number of test images	The number of users in \mathcal{Q}	The number of images in \mathcal{Q}
u_{q_A}	139	8	1,615
u_{q_B}	209	6	734

<표 1>에서 확인할 수 있는 것처럼, u_{q_A} 의 경우 정해놓은 기준 이상의 사회적 관계 크기를 갖는 사용자가 8 명이었고 u_{q_B} 의 경우 6 명이였다. 각 u_q 에 대한 \mathcal{F} 의 T 는 5 명의 사람이 직접 정확한 태그를 등록해 주었다. 이 때 보다 신뢰성 있는 실험을 위하여 375 개의 태그 목록 [15]을 이용하여 목록에 존재하는 태그만을 등록하였다.

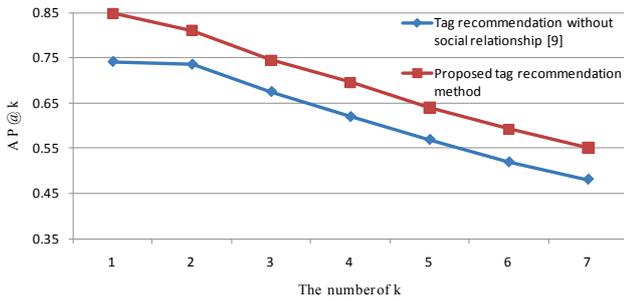
4.2 이미지 태그 추천 성능 평가 방법

추천된 태그의 정확도를 수치적으로 확인하기 위하여 본 논문에서는 추천된 태그들 중 상위 k 개의 태그만을 가지고 정확도를 측정하고 이들의 평균값을 나타내는 AP@k (Average Precision at k)를 이용하여 태그 추천 성능을 비교한다. 이 때 k 는 테스트 이미지들에 전문가가 직접 등록한 기준 (ground truth) 태그의 평균 개수인 7 개를 최대로 하여 성능을 평가한다.

4.3 이미지 태그 추천 결과

u_{q_A} 와 u_{q_B} 의 결과를 모두 합쳐 평균을 내서 최종 실험 결과를 얻었다. k 를 1 부터 7 까지 바꿔가며 얻은 AP@k 는 그림 1 와 같다. 그림 1 의 x 축은 k 의 숫자를 나타내고 y 축은 AP@k 값을 나타낸다. AP@1 이 84.8%로 [9]와 약 10.7%정도의 차이로 가장 컸고 AP@7 에서 55.6%로 [9]와 비교해 약 7%정도의 차이로 가장 작은 차이를 보였다. 평균적으로는 [9]의 방법보다 7.9%정도의 성능 향상을 보였는데 이는 [9]의 경우 시각적 저 수준 특징은 비슷하지만 실제 태그 입장에서는 무관한 이미지들이 많이 나타나기 때문이다. 하지만 제안하는 방법에서는 상대적으로 태그 입장에서 유사성이 높은 이미지를 갖고 있을 확률이 높은 사용자들로부터 시각적 저 수준 특징이 비슷한 이미지를 찾기 때문에 [9]의 방법보다는 무관한 이미지가 적게 나타나게 되고 비교적 높은 성능을 보여준다. 그림 1 에서 k 가 증가할수록 성능이 낮아지는 이유는 높은 k 일수록 $P(t|q, \mathcal{Q}, \mathcal{G})$ 값이 낮은 태그가 선택되기

때문이다.



(그림 1) k의 변화에 따른 AP@k의 변화

그림 2는 실험 결과의 예제를 보여준다. 3개의 이미지에 대해서 각 이미지에 추천된 태그들을 높은 확률 순으로 나열하였다. 그림 2에서 밑줄 친 태그는 추천된 태그들 중 실제로 이미지와 관련된 태그를 나타낸다.

Images	Recommended tags	
	Tag recommendation without social relationship [9]	Proposed tag recommendation method
	<u>man</u> , <u>people</u> , <u>sky</u> , tree, room, shadows, bench, food, <u>woman</u> , <u>sea</u>	<u>man</u> , <u>people</u> , <u>sky</u> , <u>beach</u> , <u>waves</u> , <u>light</u> , <u>reflection</u> , <u>tree</u> , <u>woman</u> , <u>sea</u>
	<u>man</u> , <u>people</u> , <u>window</u> , tree, water, room, sky, <u>food</u> , <u>wall</u> , <u>door</u>	<u>man</u> , <u>people</u> , <u>window</u> , room, <u>wall</u> , <u>door</u> , <u>food</u> , <u>clothes</u> , <u>wood</u> , tables
	<u>stone</u> , <u>bush</u> , sidewalk, bench, ground, food, snow, reflection, road, cafe	<u>mountain</u> , <u>stone</u> , stairs, <u>fence</u> , <u>tree</u> , food, ground, decoration, <u>wood</u> , <u>people</u>

(그림 2) 이미지 태그 추천 방법의 결과 예제

5. 결론

본 논문에서는 사회적 네트워크를 기반으로 하는 폭소노미에서 태그 추천을 받기 원하는 사용자를 중심으로 얼굴 정보를 통해 사회적 관계의 크기를 정량화한다. 이를 기반으로 제한된 폭소노미를 구성하고 태그 추천을 함으로써 기존의 방법보다 높은 성능을 가질 수 있는 태그 추천 방법을 제안하였다. 싸이월드라는 국내의 대표적인 사회적 네트워크 웹사이트의 데이터를 기반으로 한 실험을 통해서 실제로도 추천된 태그들이 기존 방법보다 높은 정확도를 보이는 것을 확인하였다.

참고문헌

[1] Flickr, Flickr Website. <http://www.flickr.com/>.
 [2] Facebook, Facebook Website. <http://www.facebook.com/>.
 [3] Flickr blog, "4,000,000,000," Oct. 2009. Available on <http://blog.flickr.net/en/2009/10/12/4000000000/>.

[4] B. Sigurbjornsson, R. van Zwol, "Flickr Tag Recommendation based on Collective Knowledge," Proceedings of 17th Int. World Wide Web Conference, pp. 417-426, 2008.
 [5] C. Cusano, G. Ciocca, R. Schettini, "Image annotation using SVM," Proceedings of Internet Imaging, Vol.4, pp. 330-338, SPIE, 2004.
 [6] E. Chang, G. Kingshy, G. Sychay, G. Wu, "CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines," IEEE Trans. CSVT, Vol. 13, No. 1, pp. 26-38, 2003.
 [7] G. Carneiro, N. Vasconcelos, "A database centric view of semantic image annotation," Proceedings of SIGIR, pp. 559-566, 2005.
 [8] S. G. Sevil, O. Kuckutunc, P. Duygulu, F. Can, "Automatic tag expansion using visual similarity for photo sharing websites," Multimedia Tools and Applications, Oct. 2009.
 [9] Sihyoung Lee, Wesley De Neve, Konstantinos N. Plataniotis, Yong Man Ro, "MAP-based image tag recommendation using a visual folksonomy," Pattern Recognition Letters, Jan. 2010.
 [10] X. Li, L. Guo, Y. E. Zhao, "Tag-based Social Interest Discovery," Proceeding of 17th International World Wide Web Conference, 2008.
 [11] D. M. Boyd, N. B. Ellison, "Social Network Sites: Definition, History and Scholarship," Journal of Computer-Mediated Communication, 2007.
 [12] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," Proceeding of Internet Measurement Conference, 2007.
 [13] J. Breslin, S. Decker, "The Future of Social Networks on the Internet," IEEE Internet Computing 11, pp. 86-90, 2007.
 [14] Cyworld, Cyworld Website. <http://www.cyworld.com/>.
 [15] V. Lavrenko, R. Manmatha, J. Jeon, "A model for learning the semantics of pictures," Conference on Advances in Neural Information Processing Systems, 2003.