위치적 연관성과 어휘적 유사성을 이용한 웹 이미지 캡션 추출

(Web Image Caption Extraction using Positional Relation and Lexical Similarity)

이 형 규 † 김 민 정 †† 홍 금 원 † 임 해 창 †††
(Hyoung-Gyu Lee) (Min-Jeong Kim) (Gumwon Hong) (Hae-Chang Rim)

요 약 이 논문은 웹 문서의 이미지 캡션 추출을 위한 방법으로서 이미지와 캡션의 위치적 연관성과 본문과 캡션의 어휘적 유사성을 동시에 고려한 방법을 제안한다. 이미지와 캡션의 위치적 연관성은 거리와 방향 관점에서 캡션이 이미지에 상대적으로 어떻게 위치하고 있는지를 나타내며, 본문과 캡션의 어휘적 유사성은 이미지를 설명하고 있는 캡션이 어휘적으로 본문과 어느 정도 유사한지를 나타낸다. 이미지와 캡션을 독립적으로 고려한 자질만을 사용한 캡션 추출 방법을 기저 방법으로 놓고 제안하는 방법들을 추가적인 자질로 사용하여 캡션을 추출하였을 때, 캡션 추출 정확률과 캡션 추출 재현율이 모두 향상되며, 캡션 추출 F-measure가 약 28% 향상되었다.

키워드 : 이미지 캡션, 이미지 캡션 추출, 위치적 연관성, 어휘적 유사성

Abstract In this paper, we propose a new web image caption extraction method considering the positional relation between a caption and an image and the lexical similarity between a caption and the main text containing the caption. The positional relation between a caption and an image represents how the caption is located with respect to the distance and the direction of the corresponding image. The lexical similarity between a caption and the main text indicates how likely the main text generates the caption of the image. Compared with previous image caption extraction approaches which only utilize the independent features of image and captions, the proposed approach can improve caption extraction recall rate, precision rate and 28% F-measure by including additional features of positional relation and lexical similarity.

Key words : Image caption, Image caption extraction, Positional relation, Lexical similarity

1. 서 론

이미지 캡션 추출이란 문서 내에 존재하는 텍스트 중 이미지를 직접적으로 설명하고 있는 캡션 텍스트를 추출하는 기술을 의미한다.

웹 문서로부터 이미지 캡션을 추출하는 기술은 웹 이미지 검색 시스템에서 이미지 색인에 사용될 수 있다. 이미지 캡션은 이미지와 직접적으로 연관된 내용을 담고 있으므로 이미지 캡션 추출의 성능은 이미지 검색의 성능을 좌우한다. 따라서 정확한 이미지 캡션을 추출하는 기술은 웹 검색 시스템을 구축하는 중요 요소 기술로 사용될 수 있다[1-6]. 또한, 이러한 캡션 추출 기술은 음악, 동영상 등의 다른 멀티미디어 검색에도 적용이 가능하기 때문에 웹 문서 내의 이미지 캡션을 추출하는 기술은 그 응용 분야가 넓다.

이미지 검색의 색인어를 쉽게 추출할 수 있는 방법

· 본 연구는 2단계 BK21사업과 2008년도 NHN Corp.의 지원을 받아 수행된 연구임

† 비 회 원 : 고려대학교 컴퓨터.전파통신공학과
hglee@nlp.korea.ac.kr
gwhong@nlp.korea.ac.kr
†† 학생회원 : 고려대학교 컴퓨터.전파통신공학과
mjkim@nlp.korea.ac.kr
††† 종신회원 : 고려대학교 컴퓨터통신공학부 교수
rim@nlp.korea.ac.kr

논문접수 : 2008년 7월 10일
심사완료 : 2009년 2월 6일

정보과학회논문지: 소프트웨어 및 응용 제36권 제4호(2009.4)

중 하나는 웹 문서의 구조적인 정보만을 고려하여 색인어를 추출하는 경우이다. 예를 들어, 웹 문서의 <img> 태그에 포함된 alt속성의 텍스트나 <title> 태그의 텍스트, 그리고 이미지 경로 및 파일명에 나타나는 단어 등을 대상으로 텍스트를 추출한다면 이미지를 설명하는 텍스트를 쉽고 빠르게 추출할 수 있다[2,4].

이와 같은 방법은 이미지를 설명하는 텍스트를 쉽고 빠르게 추출할 수는 있지만, 위 정보들이 항상 이미지를 설명하고 있는 텍스트라고 볼 수는 없다. 예를 들어, <img> 태그의 alt 속성이나 <title> 태그 내에 텍스트가 이미지를 설명하지 않는 경우가 많이 있고, 이미지 파일명이 항상 그 이미지를 설명한다고 보기 어렵기 때문이다. 이러한 이유로, 본 논문에서는 웹 브라우저 상에서 사용자에게 보여지는 텍스트만을 대상으로 이미지 캡션을 추출하고자 한다.

이미지 캡션 추출에 대한 기존 연구로는 단어, HTML 태그, 이미지 속성 등의 각종 자질을 사용하여 학습 기반으로 접근한 연구[2]와 텍스트의 글꼴과 같은 레이아웃 정보를 이용하여 규칙 기반으로 PDF 문서에서의 캡션 추출을 시도한 연구[5]가 있다. [2]는 웹 이미지 캡션 추출에 유용한 여러 자질들을 제안하였고, 캡션 기반 이미지 검색 시스템을 구현하였다는 점에서 의미가 있다. 그러나 [2]에서 제안된 자질들은 이미지, 캡션 각각을 독립적으로 고려한 자질들이라 할 수 있으며, 이 자질들로는 이미지와 캡션의 위치 관계, 형태 관계를 고려하지 못해 캡션의 전형적인 스타일에서 어긋나는 텍스트를 캡션으로 추출하여 오류를 만드는 문제점이 있다. 또한 캡션과 이미지의 의미적 연관성 또한 전혀 고려하지 못하였다. [5]는 PDF문서에서 텍스트 레이아웃을 이용하여 이미지 캡션 추출의 높은 성능을 달성하였지만, 어휘 정보, 의미 정보를 전혀 고려하지 않아 의미적 연관성이 없는 캡션을 추출하는 한계를 보여주었다.

본 연구에서는 기존 연구의 한계점을 극복하고 웹 문서에서의 이미지 캡션 추출의 성능을 향상시키기 위해, 캡션과 이미지의 연관성을 이용한 자질들의 사용을 제안한다. 우선, 캡션과 이미지의 위치적 연관성을 이용한 자질들을 제안한다. 캡션과 이미지의 정확한 위치적 연관성을 알아내기 위해, 웹 이미지 분류 연구[7,8]에서 사용한 HTML 요소(element)들의 웹 브라우저 상 위치 값 획득 방법을 사용한다. 다음으로, 캡션과 본문의 어휘적 유사성을 이용한 자질의 사용을 제안한다. 이 자질은 이미지를 포함한 문서에는 이미지와 관련된 내용을 설명하고 있는 본문이 존재할 것이라는 가정을 포함한다. 이미지 캡션 추출에 관한 기존 연구들은 이미지와 캡션의 의미적 연관성에 대한 고려가 필요하다는 인식은 공통적으로 가지고 있으나, 이미지를 포함한 문서의 본문의 분석을 통한 캡션과 이미지 간의 연관성 파악에 대한 시도는 없었다.

본 논문은 다음과 같은 순서로 구성된다. 2장에서는 이미지 캡션 추출에 관련된 기존 연구들을 설명한다. 3장에서는 제안하는 자질들을 설명하고, 4장에서는 실험 및 평가, 5장에서는 본 연구의 결론을 설명한다.

2. 관련 연구

2.1 이미지 캡션 추출 관련 연구

이미지 검색에 초점을 맞추어 웹 문서 내에서 이미지를 색인하기 위한 텍스트를 찾는 연구는 다수 존재하지만[9-14], 이미지 캡션을 찾는 것에 초점을 맞춘 연구는 많지 않다. 이미지 캡션 추출에 관련된 대표적인 연구로 Rowe의 연구[1-4]와 Maderlechner의 연구[5]를 꼽을 수 있다.

Rowe는 [2]에서 HTML 문서 내의 이미지 캡션을 추출하는 연구를 수행하였다. 이 연구는 캡션 기반 이미지 검색 시스템인 MARIE-4를 제안하면서 캡션 추출을 위한 각종 자질들의 유용성을 실험하였다.

이 연구는 이미지 캡션을 추출하는 데에 유용한 자질로서 picture, show 등과 같이 캡션에 자주 출현하는 단어가 나타났는지 여부, 캡션에 자주 사용되는 HTML 태그가 나타났는지 여부, 이미지 파일명에 button과 같이 캡션 추정의 부정적 단서가 되는 용어와 media와 같이 긍정적 단서가 되는 용어의 출현 여부, 이미지 파일명에 숫자 존재 여부, 이미지 파일명과 캡션에 동시에 나타나는 단어 존재 여부, 이미지 포맷(gif/jpg), 캡션 길이, 이미지 크기 등 8가지 자질을 제안하였다. 반면, 이미지와 캡션 간의 거리는 이미지 캡션을 추출하는 데에 유용하지 않음을 실험 결과를 통해 밝혔다. 이미지와 캡션 간의 거리를 측정하는 데에 HTML 코드 상으로 몇 라인 떨어져 있는지를 측정하였는데, 이렇게 측정된 거리는 웹 브라우저 상에서 보여지는 실제 거리와는 차이를 보였다.

이 연구에서 제안된 캡션 추출을 위한 자질들은 캡션 특유의 어휘적 특성, 캡션을 갖는 이미지의 파일명의 특성, 이미지 속성 등을 이용하였다. 그러나, 제안된 자질들은 이미지와 캡션을 독립적으로 고려한 자질들로서 캡션과 이미지 간의 위치적 연관성 및 의미적 연관성은 고려하지 못한 한계가 있다. 예를 들어, 캡션에 자주 출현하는 단어가 나타났는지 여부는 캡션만을 고려한 자질이며, 이미지 크기는 이미지만을 고려한 자질이다. 이 연구에서는 이미지와 캡션의 위치적 연관성을 고려하지 못해, 캡션의 전형적인 스타일에서 어긋나는 텍스트를 캡션으로 추출하여 캡션 추출의 정확률을 떨어뜨리는 요인이 되었으며, 이미지와 캡션의 의미적 연관성을 고

려하지 못해, 이미지를 설명하고 있지 않은 텍스트를 캡션으로 추출하는 한계가 있었다.

[4]에서는 웹 문서의 이미지, 음악, 동영상 등 멀티미디어의 캡션을 추출하기 위한 자질들을 조사하였다. 이 조사에서는 구문적(syntactic) 자질과 의미적(semantic) 자질로 구분하여 많은 자질들을 설명하였는데, 대다수의 자질들이 실험 결과로 검증되지 못한 한계가 있다.

Maderlechner는 [5]에서 텍스트 레이아웃을 이용하여 PDF 문서의 이미지 캡션을 추출하였다. 텍스트 레이아웃이란 캡션의 글꼴(font type), 글자 크기(font size), 그리고 이탤릭, 진한글자 등의 글자 스타일(font style) 등의 외형적 특징을 말한다. PDF 문서는 대체로 일관된 형태로 캡션이 작성되기 때문에 레이아웃 분석만으로도 높은 정확률과 재현율을 달성할 수 있었다. 이 연구에서는 제안하는 방법의 범용성을 실험하기 위해 HTML 문서를 PDF 문서로 변환하여 제안하는 방법을 적용하였다. 그러나, HTML 문서의 특성 상 캡션이 아닌 텍스트가 캡션의 전형적인 위치와 형태로 존재하는 경우가 있기 때문에 이미지와 의미적으로 연관이 없는 캡션이 추출되는 한계를 드러내었다.

이미지 캡션의 추출이 아닌 생성에 관한 연구 중에는 이미지와 캡션 간의 의미적 연관성을 고려한 연구가 있었다. [6]은 이미지 처리 기법을 통해 이미지의 개념 부류를 결정하고, 미리 정의해 놓은 개념 부류 키워드를 캡션의 일부로 사용하였다. 그러나, 이 연구는 이미지의 개념을 캡션 생성에 활용했다는 측면에서 캡션과 이미지 간의 의미적 연관성을 고려했다고 볼 수 있지만, 웹 문서 내의 텍스트를 대상으로 이미지 캡션을 추출한 것이 아니라는 점에서 본 연구와는 방향의 차이를 보인다.

2.2 어휘적 유사성 측정 관련 연구

캡션과 본문의 어휘적 유사성을 이용한 자질과 관련된 연구로는 문서 간 유사도 측정에 관한 연구들이 있다.

2.2.1 벡터 공간 모델 관점에서 문서 간 유사도 측정

벡터 공간 모델 관점의 유사도 측정은 하나의 문서를 벡터로 표현하고, 벡터의 각 요소를 어휘 출현 회수로 표현한다. 이러한 관점에서 matching 계수, 다이스 계수, 자카드 계수, Overlap 계수, 코사인 유사도 등의 다양한 유사도 척도가 연구되어 왔다[15]. 이 척도들은 공통적으로 두 문서 간에 공통된 어휘가 많을수록 높은 유사도를 나타낸다.

2.2.2 언어 모델 관점에서 질의 생성 확률 추정

[16]에서 소개하고 있는 언어 모델 관점의 질의 생성 확률 추정 방법은 비교적 짧은 질의와 긴 문서 간에 공통적으로 나타나는 어휘를 통해 확률값을 계산한다는 점에서 문서 간 유사도 측정 방법의 일종으로 볼 수 있다. 이 방법에서는 긴 문서를 하나의 언어 모델로 간주하고 질의가 언어 모델로부터 생성될 확률을 계산한다. 이렇게 계산된 질의 생성 확률이 높을수록 질의와 문서의 어휘적인 유사성이 높다. 이 방법은 질의와 문서의 관계처럼 길이 차이가 큰 문서 간에 유용한 방법이라 할 수 있다.

2.2.3 심층적인 자연어처리 기법을 활용한 연구

앞에서 설명한 어휘적 유사성 측정 방법들은 공통적으로 Bag of words 가정을 하고 있어서 어휘간 연관성, 어휘의 의미, 문장의 구조를 고려하지 못하는 한계가 있다. 이러한 문제점을 완화시키기 위해 워드넷과 같은 온톨로지 활용, 구문분석 등의 심층적인 자연어처리 기술을 사용할 수 있다.

그러나 심층적인 분석을 본 연구에 적용하기에는 다음과 같은 문제가 있음을 관찰하였다.

첫째, 웹 이미지 캡션 추출은 웹의 특성상 높은 성능의 언어처리 결과를 기대하기 힘들다. 특히 다수의 저품질 문서, 띄어쓰기 오류, 철자 오류, 신조어 등의 문제가 있어서 형태소분석조차 실패하는 경우가 적지 않았으며 구문분석, 의미분석의 성능은 매우 낮을 것으로 판단된다.

둘째, 이미지 캡션은 완전한 문장이 아닌 경우가 많다. 따라서 문장과 문장 사이의 유사도 측정의 경우에는 구문구조가 유용한 자질로 사용될 수 있으나, 본문과 캡션의 유사도를 측정하는 경우에는 구문구조가 큰 도움이 되지 않을 가능성이 크다.

셋째, 구문분석이나 의미분석이 캡션 추출의 속도를 저하시키는 문제가 있다. 이미지 검색 시스템이 자동 추출된 이미지 캡션을 색인에 사용하기 위해서는 방대한 양의 웹에서 캡션을 추출해야 하기 때문에 속도 역시 중요한 문제라 할 수 있다.

따라서 본 논문은 복잡한 자연어처리 기법은 적용하지 않으며 캡션이 문서 내에 위치하는 정보와 더불어 본문과의 유사성이 캡션 추출에 어떤 도움을 줄 수 있는지 알아보기 위한 시도에 의의를 둔다. 본 논문에서는 코사인 유사도와 언어 모델 관점의 질의 생성 확률을 이용하여 어휘적 유사성을 측정하며, 자세한 내용은 3.2절에서 설명한다.

3. 제안하는 자질들

이 장에서는 이미지와 캡션의 위치적 연관성과 캡션과 본문의 어휘적 유사성을 이용한 캡션 추출을 위한 자질들에 대해 설명한다. 3.1절에서는 이미지와 캡션의 위치적 연관성을 이용한 자질들, 3.2절에서는 캡션과 본문의 어휘적 유사성을 이용한 자질을 설명한다.

3.1 이미지와 캡션의 위치적 연관성을 이용한 자질들

[2]에서 제안한 8가지의 자질에는 캡션의 이미지에 상대적인 위치나 형태를 이용한 자질은 없었다. 그러나 웹

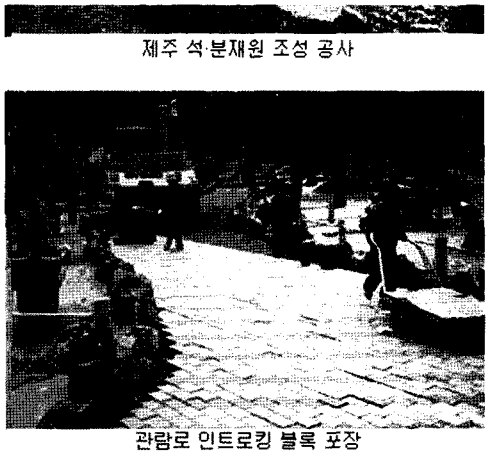

그림 1 이미지와 캡션의 위치적 연관성이 유용한 예

문서의 저자가 이미지의 캡션을 작성할 때, 대체로 이미지에 가까우면서 다른 텍스트와는 구별되는 위치에 캡션이라 인식되는 적절한 형태로 작성한다. 이로 인해 대다수의 캡션은 그 이미지에 상대적인 위치나 형태가 전형적인 패턴을 갖는다. 예를 들어, 그림 1에서 이미지와 캡션의 위치적 연관성을 이용하지 않고 캡션 추출을 한다면, 이미지의 위에 위치한 텍스트와 아래에 위치한 텍스트 중 무엇이 캡션인지 알지 못하는 경우가 발생한다. 이러한 모호성은 이미지와 캡션 간의 거리 및 캡션의 상대적인 위치와 같은 정보를 고려했을 때 해결할 수 있다.

따라서 이미지 캡션을 추출하는 데에 이미지와 캡션의 위치적 연관성을 이용한 자질의 활용은 필수적이라 할 수 있다. 웹 브라우저에 보여지는 요소들의 정확한 위치 값을 사용한다면,[1] 이미지와 캡션 간 위치적 연관성을 이용할 수 있다. 이에 본 연구에서는 다음과 같은 자질들을 제안한다.

캡션과 이미지 간 거리 캡션은 이미지와 가까운 곳에 위치한다고 가정한다. 웹 브라우저에 보여지는 요소들의 정확한 위치 값을 사용하여 이미지와 캡션후보 간의 최단 거리를 픽셀 단위로 얻어 낸다. 그리고나서 거리값을 40픽셀 간격으로 5단계로 구간화한다. 구간의 간격과 개수는 실험에 의한 최적값으로 구하였다. 이렇게 구해진 0에서 4사이의 정수를 캡션 추출을 위한 하나의 자질로 제안한다.

$$f_{dist}(cap,img)=\begin{cases}\left\lfloor \dfrac{d(cap,img)}{40} \right\rfloor & if\ d(cap,img)<200\\ 4 & otherwise\end{cases}$$

f_{dist}는 캡션과 이미지 간 거리 자질값이며, cap은 특정 캡션후보, img는 특정 이미지이고 $d(cap,img)$는 cap과 img 간의 픽셀 거리값이다.

캡션의 이미지에 상대적 위치 캡션은 이미지를 기준으로 어느 한 방향에 많이 나타나는 패턴을 가지고 있을 것이라고 가정한다. 본 연구에서는 각 이미지에 대한 캡션후보들이 이미지로부터 어느 방향(상, 하, 좌, 우)에 위치하는지를 캡션 추출을 위한 하나의 자질로 제안한다.

$$f_{pos}(cap,img)=\begin{cases}\text{상} & if\ cap\text{이 } img\text{위에 위치}\\ \text{하} & if\ cap\text{이 } img\text{아래에 위치}\\ \text{좌} & if\ cap\text{이 } img\text{왼쪽에 위치}\\ \text{우} & if\ cap\text{이 } img\text{오른쪽에 위치}\end{cases}$$

캡션의 위치 독립성 캡션은 다른 텍스트나 다른 이미지와는 독립적으로 위치한다고 가정한다. 즉, 캡션은 설명하고자 하는 대상 이미지와는 가까운 곳에 위치하면서, 다른 텍스트나 다른 이미지와는 구분되는 곳에 위치한다. 본 연구에서는 캡션과 대상 이미지 간의 거리보다 더 가까운 거리에 다른 이미지 또는 다른 텍스트가 존재하는지 여부를 캡션 추출을 위한 자질로 제안한다.

$$f_{ind\_img}(cap,img)=\begin{cases}0 & if\ \exists i\in I,\ d(cap,img)>d(cap,i)\\ 1 & otherwise\end{cases}$$

$$f_{ind\_cap}(cap,img)=\begin{cases}0 & if\ \exists t\in T,\ d(cap,img)>d(cap,t)\\ 1 & otherwise\end{cases}$$

i는 cap과 img가 존재하는 문서 내의 임의의 이미지이고 t는 임의의 텍스트단편이다. I는 문서 내 모든 이미지의 집합이며, T는 문서 내 모든 텍스트단편의 집합이다.

이미지 너비 대비 캡션의 너비 캡션은 특유의 형태로 존재한다. 특히 캡션의 너비는 이미지의 너비에 의존적이다. 일반적으로 웹 문서의 저자는 이미지의 너비를 넘지 않도록 캡션의 너비를 정하여 작성한다고 가정한다. 본 연구에서는 캡션후보의 가로 너비가 이미지의 너비에 비해 더 좁은지 아니면 더 넓은지 여부를 캡션 추출을 위한 자질로 제안한다.

$$f_{wide}(cap,img)=\begin{cases}1 & if\ w(cap)>w(img)\\ 0 & otherwise\end{cases}$$

$w(x)$는 x의 너비를 의미한다.

3.2 캡션과 본문의 어휘적 유사성을 이용한 자질

[2]와 [5]는 캡션을 추출하는 데에 이미지와 캡션의 의미적 연관성을 고려하지 못한 한계가 있었다. 예를 들어, 그림 2는 이미지와 캡션 간의 의미적 연관성을 파악하지 못하면 캡션을 잘못 추출할 수 있는 경우의 전형적인 예이다. 주어진 이미지들은 캡션이 없는 이미지들

1) 본 논문에서는 MsHtml 인터페이스 모음(http://msdn.microsoft.com/en-us /library/aa741317.aspx)을 사용하여 HTML 파서를 구현하였다. MsHtml의 navigate2 인터페이스를 사용하면 웹 페이지를 구성하고 있는 모든 요소들의 웹 브라우저 상의 실제 위치값을 구할 수 있다. MsHtml은 인터넷 익스플로러에서 파싱과 랜더링을 담당하고 있다.

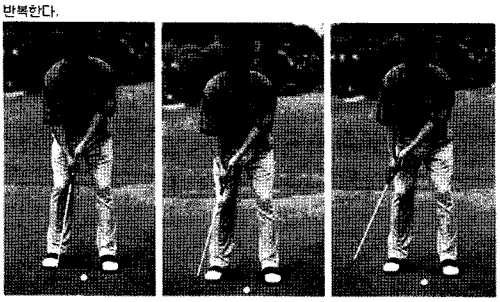

그림 2 캡션과 본문의 어휘적 유사성이 유용한 예

이다. 그러나 이미지들의 아래에 있는 "다음 글: 칩샷과 피치샷"이라는 텍스트를 캡션으로 잘못 추출할 가능성이 높다. 이 텍스트는 그림을 설명하고 있는 텍스트가 아닌, 단지 이 문서에 이어지는 다음 글이 있음을 알려주는 텍스트이다. 캡션과 이미지의 의미 관계를 고려하지 않는 한 이러한 모호성을 해결하기는 쉽지 않다. 이 절에서는 이러한 한계를 극복하기 위한 언어적 분석 방법을 제안한다.

웹 문서 내에서 캡션을 갖는 이미지는 일반적으로 저자가 설명하고자 하는 내용의 이해를 돕기 위해 삽입된다. 따라서 캡션을 갖는 이미지는 문서 본문과 의미적으로 밀접한 연관성을 갖는다고 할 수 있다. 다시 말해, 이미지가 속한 문서의 본문의 내용을 파악하는 것이 이미지의 의미를 파악하기 위한 간접적인 방법이 될 수 있다.

본 연구에서는 캡션후보와 이미지의 의미적 연관성을 파악하기 위한 방법으로서 이미지가 속한 문서의 본문과 캡션후보의 어휘적 유사성을 측정하는 방법을 제안한다. 캡션에 사용된 어휘와 본문에 사용된 어휘의 유사성이 높을수록 캡션후보와 이미지의 의미적 연관성이 높다고 할 수 있다. 그림 2의 예를 보면, "다음 글: 칩샷과 피치샷"이라는 텍스트에 포함된 "다음", "글", "칩샷", "피치샷" 등의 어휘는 본문 텍스트에서 등장하지 않고 있다. 그러므로 주어진 이미지와 관련된 본문과 해당 캡션후보는 어휘적 유사성이 낮고, 이미지와 캡션후보 간 의미적 연관성 또한 떨어진다고 볼 수 있다.

캡션후보와 본문의 어휘적 유사성 측정에는 다양한 유사도 측정 방법을 사용할 수 있다. 그러나 본 논문에서는 캡션과 본문의 어휘적 유사성이 이미지 캡션 추출에 유용할 것이라는 가설을 검증하는 데에 초점을 맞추고 있다. 따라서 간단히 유사도를 측정할 수 있는 벡터 공간 모델 관점의 코사인 유사도 척도와 언어 모델 관점의 접근법을 사용하는 방법을 제시하고, 실험을 통해 각 방법으로 얻은 자질이 캡션 추출 성능에 기여하는 정도를 비교하도록 한다.

언어 모델 관점의 접근법은 이미지가 속한 문서의 본문으로부터 만들어진 언어 모델에 의해 캡션후보가 생성되었을 확률을 구하는 방법이다. 이 확률이 높을수록 그 캡션후보와 해당 본문의 어휘 유사성이 높다고 할 수 있다. 구체적인 절차는 다음과 같다. 먼저 웹 문서 내의 각 이미지에 대해 해당 본문을 결정한다. 이때 본문은 이미지가 속한 문서 내에 존재하면서 그 이미지의 캡션후보가 아닌 모든 텍스트를 포함한다. 그리고 그 본문으로부터 언어 모델을 만든다. 언어 모델은 유니그램으로 가정하고, 명사와 동사, 형용사만을 고려한다. 마지막으로, 해당 이미지에 대응되는 캡션후보들이 언어 모델로부터 생성될 확률을 구한다.

$$P(C \mid M_d) = \prod_{t \in C} P(t \mid M_d) = \prod_{t \in C} \frac{tf_{(t,d)}}{dl_d}$$

M_d는 각 이미지와 대응되는 본문 d로부터 만들어진 언어 모델이고, C는 캡션후보며, $tf_{(t,d)}$는 캡션후보 C에 속한 한 단어 t가 본문 d에서 출현한 회수이고, dl_d는 본문 d 내에 존재하는 단어들의 출현 회수 총합이다.

이 때, 캡션후보의 길이가 길수록 확률이 작아지는 경향이 있으므로 길이 정규화를 해준다. 결국, 이렇게 구해진 값은 캡션후보와 본문의 유사한 정도를 보여주는 척도가 된다. 최종 수식은 다음과 같다.

$$\frac{1}{|C|} \sum_{t \in C} \log \frac{tf_{(t,d)}}{dl_d}$$

벡터 공간 모델 관점의 접근법은 캡션후보와 이미지와 대응되는 본문을 각각 하나의 벡터로 간주하고, 두 벡터 간의 코사인(cosine) 값을 통해 유사성을 측정하는 방법이다. 이러한 코사인 유사도는 두 텍스트 간의 길이의 차이가 클 때 유용한 유사도 측정 방식이다. 따라서, 상대적으로 길이가 긴 본문과 길이가 짧은 캡션후보 간의 유사도 측정에 좋은 척도이다.

문서 내의 캡션후보를 포함한 모든 텍스트에 나타난 명사, 동사, 형용사 목록으로 벡터 공간을 구성한다. 그리고 각 캡션후보와 본문을 벡터로 표현한다. 두 벡터, v_1, v_2 간의 코사인 유사도를 계산하는 수식은 다음과 같다.

$$\cos\theta = \frac{\mathrm{V}_1 \cdot \mathrm{V}_2}{|\mathrm{V}_1||\mathrm{V}_2|}$$

예를 들어, 하나의 캡션후보가 "칩샷과 피치샷"이라면 그 캡션후보에 속한 단어는 "칩샷", "피치샷"이므로 벡터는 (0,0,...,1,...,1,...)로 표현된다. 그리고 비교하고자 하

는 본문에 "칩샷"이 두 번, "피치샷"이 세 번 출현하였다면 그 본문의 벡터는 (...,2,...,3,...)로 표현된다. 마지막으로 이 두 벡터의 코사인 유사도를 구한다.

4. 실험 및 평가

이 장에서는 본 연구에서 제안하는 캡션 추출을 위한 자질들에 대한 실험에 대해 설명한다. 4.1절에서는 실험을 위해 구현된 캡션 추출 시스템을 설명하고, 4.2절에서는 실험 데이터를 설명하며, 4.3절에서는 평가 척도를 설명한다. 4.4절에서는 관련 연구에서 제안된 자질들로부터 기저 방법 설정을 위한 실험 결과를 설명하며, 4.5절에서는 기저 성능과 제안하는 자질들의 성능의 비교 평가를 보여준다. 마지막으로 4.6절에서는 오류를 분석한다.

4.1 캡션 추출 시스템 구성

실험에 사용된 캡션 추출 시스템은 [2]에서 제안된 캡션 추출 과정을 참고하여 구현되었다. 본 캡션 추출 시스템의 특징은 웹 브라우저에 보여지는 요소들의 정확한 위치 값을 알아내기 위해 인터넷 익스플로러에서 사용하는 MsHtml API를 사용하여 HTML 문서를 분석하였다는 점이다. 또한 학습 모델로 [2]에서는 선형 모델을 사용한 반면, 본 연구에서 제안하는 시스템은 최대 엔트로피(Maximum Entropy) 모델[17]을 사용하였다.[2] 최대 엔트로피 모델은 선형 모델에 비해 다양한 자질들에 대해 빠른 학습 속도를 보이며, 자질 별 가중치를 자동으로 계산해내는 장점을 지닌다.

캡션 추출 시스템은 그림 3과 같이 구성된다.

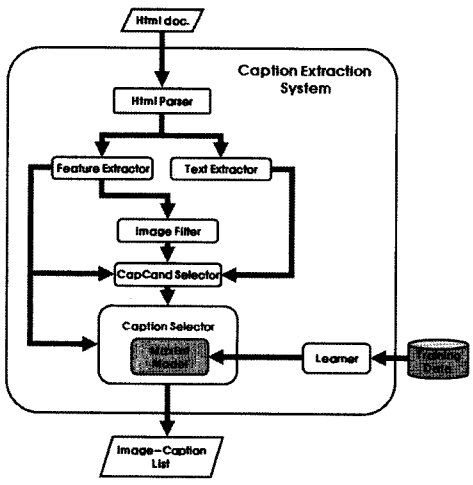

그림 3 캡션 추출 시스템 구성도

2) 본 연구에서는 Zhang Le의 최대 엔트로피 모델 툴킷(http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html)을 사용하여 캡션 선정기를 구현하였다.

HTML 파서(HTML Parser)는 HTML 문서를 분석하여 각 요소 별로 위치 값, 태그명, 속한 텍스트 등의 정보를 추출한다.

텍스트 추출기(Text Extractor)는 HTML 파싱 결과를 바탕으로 브라우저 화면에 보여지는 텍스트를 추출한 후, 캡션이 될 수 있는 단위로 단편화(fragmentation)를 수행한다. 이 때, 캡션이 될 수 있는 단위는 HTML 요소 단위로 정하였으며, 트리 상에서 단말에 가장 가까운 <p>와 <td> 요소에 속한 텍스트를 하나의 단편(fragment)으로 정한다. <p>는 단락의 구분을 나타낼 때 사용되는 태그이며, <td>는 테이블 셀의 구분을 나타낼 때 사용되는 태그이다. 우리는 많은 HTML 문서 작성자가 <p>와 <td> 태그를 통해 문서 내의 텍스트를 의미적, 구조적으로 구분해 놓는다는 점을 관찰하였으며, 따라서 이 태그 정보를 이용하여 텍스트의 단편화를 수행하였다.

자질 추출기(Feature Extractor)는 HTML 파서의 분석 결과를 바탕으로 각 이미지와 각 텍스트 단편에 대해서 각종 자질의 값을 추출한다.

이미지 여과기(Image Filter)는 문서 내에서 학습 및 테스트 대상이 될 이미지들을 선별한다. 이 때, 이미지 크기, 이미지 가로세로 비율, 이미지 파일명에 포함된 단어 등을 기준으로 하여 아이콘, 배너와 같이 캡션을 갖지 않을 만한 이미지는 여과한다.

캡션후보 선정기(CapCand Selector)에서는 각 이미지에 대해 캡션후보를 선정한다. 이 때, 이미지를 기준으로 상, 하, 좌, 우 각 방향의 가장 가까운 하나의 텍스트 단편만을 후보로 선정한다.

캡션 선정기(Caption Selector)는 교사 학습 기법을 통해 사전 학습된 최대 엔트로피 모델을 사용한다. 최대 엔트로피 모델은 새로운 테스트 데이터(이미지-캡션후보 쌍)에 대해 그것이 정답 쌍일 확률을 출력한다. 캡션 선정기는 각 이미지에 대한 캡션후보들 중 정답 쌍일 확률이 가장 높은 캡션후보를 캡션으로 선정한다. 이 때, 임계값을 적용하여 임계값을 넘는 캡션후보가 없는 이미지에 대해서는 캡션이 없는 이미지라고 출력한다.

4.2 실험 데이터

본 연구의 실험은 이미지 캡션 정답이 부착된 HTML 문서 집합을 사용하여 수행하였다. 문서 집합은 8,660개의 한글 HTML 문서로 구축하였다. 캡션 정답은 모든 이미지를 사람이 직접 보고 각 이미지 별 캡션을 판단하여 부착하였다. 전체 문서 집합 내에서 이미지 여과 후, 이미지의 개수는 총 64,250개였으며, 캡션을 갖는 이미지의 개수는 7,863개였다.

4.3 평가 척도

캡션 추출의 결과를 평가하기 위한 척도로서 캡션 추

출 정확률(Caption Extraction Precision, CEP)과 캡션 추출 재현율(Caption Extraction Recall, CER), 캡션 추출 F-measure(Caption Extraction F-measure, CEF)를 사용하였다. 이 평가 척도는 [5]에서 사용한 평가 척도와 유사하다.

먼저, 표 1과 같이 실제 정답과 시스템의 결과에 따라 전체 이미지를 구분하고 각각의 개수를 센다. 그리고 나서 그림 4의 수식에 따라 각 척도의 값을 구한다. 본 연구의 문제는 각 이미지에 대해 하나의 캡션을 추출하는 것이므로, 이미지 단위로 성능을 측정하는 것이 타당하다. 그리고, 실제 캡션을 갖는 이미지에 대해 시스템이 캡션을 추출하였을 때, 정답 캡션과 추출된 캡션의 문자열이 완전히 일치할 경우에만 맞은 것으로 간주하는 완전 일치(exact matching) 방식으로 평가하였다.

캡션 추출 정확률은 시스템이 추출한 캡션이 얼마나 정확한지를 평가하는 척도이며, 캡션 추출 재현율은 실제 모든 정답 캡션 중에 시스템이 정확한 캡션을 얼마나 추출해내었는지를 평가하는 척도이다. 캡션 추출 F-measure는 캡션 추출 정확률과 캡션 추출 재현율을 결합한 척도이다.

표 1 이미지 구분

		실제 정답		
		캡션을 갖는 이미지		캡션을 갖지 않는 이미지
시스템 결과	캡션이 추출된 이미지	CCo	CCx	CNx
	캡션이 추출되지 않은 이미지	NCx		NNo

$$CEP = \frac{CCo}{CCo + CCx + CNx}$$

$$CER = \frac{CCo}{CCo + CCx + NCx}$$

$$CEF = \frac{2 \cdot CEP \cdot CER}{CEP + CER}$$

그림 4 평가 척도

4.4 기저 방법의 설정

본 연구에서는 [2]에서 제안된 8가지 자질을 기반으로 기저 방법을 설정하였다. 우선 8가지 자질 중 한글 코퍼스에 적합하지 않은 자질인 이미지 파일명과 캡션후보에 동시에 출현하는 단어의 존재 여부를 제거하고, 나머지 7가지 자질 중 실험을 통해 기저 방법을 설정하였다. 표 2는 7가지 자질에서 하나씩 제외하여 캡션 추출한 실험 결과이다. 캡션에 자주 출현하는 단어가 나타났는지 여부는 학습 데이터의 정답 캡션을 조사하여 "그림", "사진", "모습", "풍경" 등과 같이 캡션에 많이 등장하는 단어를 결정하여 사용하였다. 캡션에 자주 사용되는 HTML 태그가 나타났는지 여부 역시, 학습 데이터의 정답 캡션의 수식 태그 분포를 조사하여 빈도가 높게 나타난<font>와 <a> 태그로 결정하여 사용하였다. 실험 결과, 성능 기여가 없는 자질인 이미지 파일명에 숫자 존재 여부, 이미지 파일명에 캡션 추정의 긍정적 단서가 되는 용어/부정적 단서가 되는 용어 존재 여부를 제외한 5가지 자질을 기저 방법으로 정하였다.

4.5 실험 결과 및 분석

본 연구에서 제안하는 자질들을 기저 방법의 자질들과 결합하여 캡션 추출의 성능을 측정함으로써 기존에 알려진 자질 외에 본 연구에서 제안하는 자질들이 캡션 추출에 유용함을 보이고자 하였다. 모든 결과는 10-fold cross validation을 수행한 결과이고, 임계값을 조절하여 캡션 추출 F-measure가 가장 큰 지점을 찾아 캡션 추출 정확률과 캡션 추출 재현율을 제시하였다. 실험에 사용된 자질의 목록은 표 3과 같다.

제안하는 자질의 실험에 앞서, 캡션과 본문의 어휘 유사성 측정 방법의 비교 실험과 각 이미지 별 본문의 범위 설정에 따른 성능 비교 실험을 수행하였다. 표 4는 캡션과 본문의 어휘 유사성을 측정하는 접근법에 따른 캡션 추출 성능 비교 결과를 보여준다. 언어 모델 관점의 접근법과 벡터 공간 모델 관점의 접근법의 사용이 캡션 추출 성능에서 큰 차이를 보이지는 않았지만, 벡터 공간 모델 관점의 접근법이 약간 더 좋은 성능을 나타

표 2 기저 방법 설정 실험 결과

자질	CEP	CER	CEF
[2]에서 제안된 7가지 자질	0.181	0.206	0.193
7가지 자질 - 캡션에 자주 출현하는 단어 등장 여부	0.172	0.187	0.179
7가지 자질 - 캡션에 자주 수식되는 HTML 태그 수식 여부	0.164	0.184	0.173
7가지 자질 - 이미지 파일명에 캡션 추정의 부정적 단서가 되는 용어/긍정적 단서가 되는 용어 존재 여부	0.188	0.203	0.195
7가지 자질 - 이미지 파일명에 숫자 존재 여부	0.172	0.219	0.192
7가지 자질 - 이미지 포맷(gif/jpg)	0.062	0.084	0.071
7가지 자질 - 캡션 길이	0.155	0.201	0.175
7가지 자질 - 이미지 크기	0.164	0.211	0.185

표 3 실험에 사용된 자질 목록

분류	자질	자질번호
기저 방법의 자질	캡션에 자주 출현하는 단어가 나타났는지 여부	1
	캡션에 자주 사용되는 HTML 태그가 나타났는지 여부	2
	이미지 포맷	3
	캡션 길이	4
	이미지 크기	5
이미지와 캡션의 위치적 연관성을 이용한 자질	캡션과 이미지 간 거리	6
	이미지 대비 캡션의 상대적 위치	7
	캡션의 위치 독립성 1: 캡션과 대상 이미지 간의 거리보다 더 가까운 다른 텍스트 존재 여부	8
	캡션의 위치 독립성 2: 캡션과 대상 이미지 간의 거리보다 더 가까운 다른 이미지 존재 여부	9
	이미지 너비 대비 캡션의 너비	10
캡션과 본문의 어휘적 유사성을 이용한 자질	캡션과 본문의 어휘적 유사성(언어 모델 관점의 접근법 사용)	11
	캡션과 본문의 어휘적 유사성(벡터 공간 모델 관점의 접근법 사용)	12

표 4 캡션과 본문의 어휘적 유사성 측정 방법 비교 실험

자질	CEP	CER	CEF
기저 방법 + 11	0.205	0.286	0.239
기저 방법 + 12	0.215	0.280	0.243

내었다. 또한, 표 5는 본문의 범위 설정에 따른 캡션 추출 성능 비교 결과를 보여준다. 본문의 범위를 어떻게 정하느냐에 따라 제안하는 캡션과 본문의 어휘 유사성 자질의 유용한 정도가 달라질 수 있다. 실험 결과, 각 이미지로부터 200픽셀 이상으로 넓게 본문의 범위를 설정하는 것이 캡션 추출 성능에 더 좋은 영향을 주었다. 이것은 이미지와 관련된 어휘가 이미지로부터의 거리와는 크게 관계없이 문서 전체적으로 분포하고 있기 때문으로 분석된다.

표 6은 제안하는 자질들의 그룹별 성능 비교 결과를 보여준다. 이 실험에서 캡션과 본문의 어휘적 유사성 자질은 이미지로부터 200픽셀 이내 존재하는 텍스트를 본문으로 설정하였고 벡터 공간 모델 관점의 접근법에 의한 유사도를 사용하였다. 실험 결과를 통해, [2]에서 제안된 자질들과 이미지와 캡션의 위치적 연관성을 이용한 자질들을 같이 사용했을 때, 기저 성능에 비해 캡션 F-measure에서 2배 이상의 성능 향상을 가져왔음을 알 수 있다. 본문과의 어휘적 유사성 자질 역시 위치적 연관성을 이용한 자질들만큼은 아니지만 성능 향상을 보였다. 그리고 제안하는 모든 자질들을 결합하여 캡션 추출할 경우 가장 높은 성능을 보여주었다. 또한, 본 논문에서 제안하는 자질들만으로 캡션 추출할 경우에도 기저 성능보다 높은 성능을 보여주었다. 이러한 결과는 본 연구에서 제안하는 이미지와 캡션의 위치적 연관성을 이용한 자질들과 캡션과 본문의 어휘적 유사성을 이용한 자질의 유용성을 입증하는 결과라고 할 수 있다.

추가적으로, 제안하는 각 자질 별로 캡션 추출 성능에 기여하는 정도를 보기 위해, 모든 자질을 사용하였을 때의 성능과 각 자질을 하나씩 제외하였을 때의 성능을 비교 실험하였다. 표 7이 그 결과를 보여준다. 이미지 대비 캡션의 상대적 위치가 성능 향상에 가장 큰 기여를 하고 있는 자질임을 알 수 있다. 이는 캡션이 일반적으로 이미지 아래 방향에 위치하는 경우가 많기 때문으

표 5 본문 범위 설정에 따른 성능 비교 (벡터 공간 모델 관점의 접근법 사용)

이미지에 대응되는 본문의 범위	CEP	CER	CEF
기저 방법 + 12(각 이미지로부터 100픽셀)	0.213	0.258	0.233
기저 방법 + 12(각 이미지로부터 200픽셀)	0.219	0.272	0.243
기저 방법 + 12(각 이미지로부터 300픽셀)	0.214	0.278	0.242
기저 방법 + 12(이미지가 속한 문서 전체)	0.215	0.280	0.243

표 6 제안하는 자질 실험 결과

자질	CEP	CER	CEF
기저 방법(1 + 2 + 3 + 4 + 5)	0.173	0.219	0.193
기저 방법 + 위치적 연관성(6 + 7 + 8 + 9 + 10)	0.443	0.459	0.451
기저 방법 + 어휘적 유사성(12)	0.215	0.280	0.243
기저 방법 + 위치적 연관성(6 + 7 + 8 + 9 + 10) + 어휘적 유사성(12)	0.467	0.482	0.475
위치적 연관성(6 + 7 + 8 + 9 + 10) + 어휘적 유사성(12)	0.327	0.381	0.352

표 7 제안하는 자질 별 영향력 비교 실험 결과

자질	CEP	CER	CEF
모든 자질 (기저 방법 + 6 + 7 + 8 + 9 + 10 + 12)	0.467	0.482	0.475
모든 자질 -6	0.459	0.473	0.466
모든 자질 -7	0.365	0.425	0.393
모든 자질 -8	0.446	0.461	0.454
모든 자질 -9	0.460	0.461	0.461
모든 자질 -10	0.460	0.474	0.467
모든 자질 -12	0.443	0.459	0.451

로 분석된다.

본 논문에서 제안하는 캡션과 본문의 어휘적 유사성을 이용한 자질이 캡션 추출에 현저한 성능 향상을 보여주지 못한 이유는 다음과 같이 분석된다.

첫째로 단순히 어휘만을 이용하여 유사도를 측정하는 방법의 한계점이다. 본 유사도 측정 방법을 이용하면 본문과 캡션에 함께 나타나는 어휘를 사용하였을 때에만 유사도가 높게 측정되고 본문의 어휘와 유사한 의미이지만 다른 어휘를 사용한 경우에는 유사도가 낮게 측정될 수 있는 문제가 있다. 따라서 본문에 사용된 어휘의 동의어나 유의어를 사용하고 있는 캡션의 의미적 유사성이 반영될 수 있도록 유의어 사전을 활용하여 유사도를 측정하는 방법이 더 유용할 수 있다.

둘째로 웹 문서의 특성 상, 본문을 통해 이미지와 캡션의 의미적 연관성을 파악하는 방법은 일반적인 문서에서는 의미가 있지만, 몇몇 문서에서는 오류를 보여 주었다. 웹 문서에는 텍스트를 통해 내용을 전달하는 것을 목적으로 하면서 부수적으로 이미지를 삽입하는 경우뿐만 아니라, 이미지를 보여주는 것을 목적으로 하여 본문이 존재하지 않는 경우 또한 많다. 앨범 형태가 대표적인 경우라 할 수 있다. 그림 5가 이러한 앨범 형태의 한 예를 보여준다. 이와 같이 본문이 존재하지 않는 문서의 경우, 본 실험에서는 캡션과 본문의 어휘 유사성 자질에 대해 임의로 중간 점수를 할당한다. 이로 인해 이미지와 캡션의 실제 의미적 연관성은 높더라도, 본문과의 어휘적 유사성 자질에서 과소평가되는 현상을 보인다.

그림 5 본문이 존재하지 않는 웹 문서의 예

4.6 오류 분석

전체 테스트 집합의 5%의 테스트 결과를 추출하여 그 중, 캡션 추출 오류가 발생한 이미지들을 대상으로 오류를 분석하였다. 분석 결과, 다음 세 가지의 오류 유형이 다수를 차지하였다.

첫번째로, 몇몇 자질에 부합하지 않아 임계값을 넘지 못한 경우가 45%를 차지하였는데, 이 유형은 실제 캡션이 있는 이미지에 대해서 시스템이 캡션을 추출하지 않은 경우로써 캡션 추출 재현율을 하락시키는 주 요인이 된다. 예를 들면, 길이가 길면서 폭이 넓은 캡션, 대상 이미지보다 더 가까운 다른 이미지가 존재하는 캡션 등이 있다.

두번째 유형은 문서 레이아웃을 인식하지 못해 발생하는 오류로서 19%를 차지하였다. 사람은 웹 문서를 보았을 때 문서 내 레이아웃이 자연스럽게 인식되면서, 본문, 소제목, 이미지 캡션 등을 파악할 수 있지만, 학습 모델은 문서 내 레이아웃을 인지할 수 없어서 생기는 오류라 할 수 있다. 대표적인 예로 그림 6과 같이 이미지 아래로 이어지는 본문의 소제목을 캡션으로 잘못 추출한 경우가 있다.

10%를 차지한 마지막 유형은 캡션 자체가 모호한 경우이다. 이 유형은 사람이 보아도 어느 것이 캡션인지 결정하기 어려운 경우이다. 그림 7이 한 예인데, 이미지 위의 캡션후보와 아래의 캡션후보가 모호하다. 이러한 경우, 실험 데이터 구축자의 판단 하에 단 하나만을 캡션 정답으로 부착하였기 때문에 시스템이 정답을 정확히 추출해내기에는 한계가 있다.

5. 결 론

본 연구는 웹 문서 내의 이미지의 캡션을 추출하는

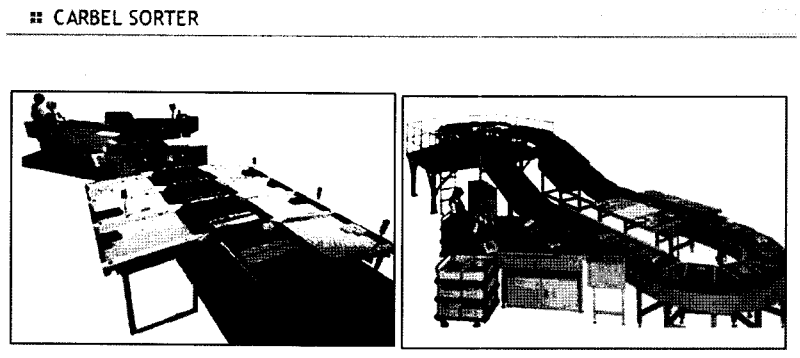

그림 6 문서 레이아웃을 인식하지 못해 캡션을 잘못 추출한 경우

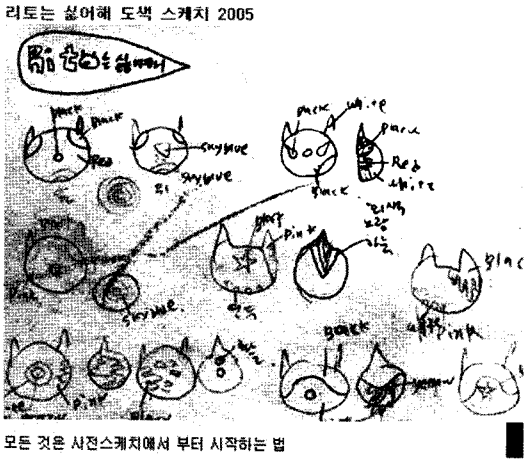

그림 7 캡션 자체가 모호하여 잘못 추출한 경우

데에 유용한 자질로서 이미지와 캡션 간의 위치적 연관성과 캡션과 본문의 어휘적 유사성을 이용한 자질들을 제안하였다.

이미지와 캡션 간의 거리 및 캡션의 이미지에 상대적인 위치, 캡션의 위치 독립성 등의 자질을 제안하여 이미지와 캡션 간의 위치적 연관성을 이용하는 것이 캡션 추출에 큰 도움을 줄 수 있음을 실험으로 증명하였다.

또한, 이미지와 캡션의 의미적 연관성을 간접적으로 파악할 수 있는 방법의 하나로서 본문과 캡션의 어휘적 유사성의 이용이 캡션 추출에 역시 유용한 자질임을 실험으로 증명하였다.

그리고, 어휘적 유사성을 측정하기 위해 본 연구에서는 언어 모델 관점의 접근법과 벡터 공간 모델 관점의 접근법을 사용하여 비교 실험하였으며, 각 이미지에 대응하는 본문의 범위를 다양하게 바꾸어서 실험하여 캡션 추출에 적합한 접근법과 본문의 범위를 찾아내었다. 또한, 제안하는 각 자질 별로 캡션 추출에 기여하는 정도를 실험하여 캡션의 이미지에 상대적인 위치 자질이 가장 유용한 자질임을 밝혔다.

본 연구는 이미지와 캡션 각각을 독립적으로 고려한 자질뿐만 아니라 이미지와 캡션의 연관성을 이용한 자질이 중요함을 밝혔다는 데에 의의를 둘 수 있다.

본 연구에서는 캡션과 본문의 유사도 측정 시 어휘적 유사성만을 고려하였다. 향후에는 구문구조 또는 의미구조 분석 등의 심층적인 자연어처리 기술을 적용한 유사도를 활용하는 연구가 필요하다. 이를 위해서는 저품질 문서, 문법적 오류가 많은 웹 문서에 견고하며 캡션의 구문적 특성에 견고한 유사도 측정 연구 역시 필요하다고 판단된다.

또한, 이렇게 본문을 이용하는 방법은 본문이 존재하지 않는 웹 문서에서는 적용이 어렵기 때문에 위치적 연관성을 이용한 자질만큼 뚜렷한 성능 향상을 가져오지는 못했다. 따라서 이미지처리 기법을 이용하여 이미지의 내용과 캡션의 내용의 의미적인 연관성을 직접 구할 수 있는 연구도 향후 연구라 할 수 있다.

참 고 문 헌

[1] N.C.Rowe and B.Frew, "Automatic caption localization for photographs on World Wide Web pages," Information Processing and Management, Vol.34, No.1, pp. 95-107, 1998.

[2] N.C.Rowe, "MARIE-4: A High-Recall Self-Impro-

ving Web Crawler That Finds Images Using Captions," IEEE Intelligent Systems, 17(4), pp. 8-14, 2002a.
[3] N.C.Rowe, "Virtual Multimedia Libraries Built from the Web," Proceedings of Joint Conference on Digital Libraries (JCDL), 2002b.
[4] N.C.Rowe, "Exploiting captions for Web data mining," Web mining: applications and techniques pp. 119-144, 2005.
[5] Maderlechner et al., "Finding Captions in PDF-Documents for Semantic Annotations of Images," SSPR&SPR 2006, pp. 422-430, 2006.
[6] 황지익 외, "텍스트 정보와 시각 특징 정보를 이용한 효과적인 웹 이미지 캡션 추출 방법", 한국컴퓨터종합학술대회 논문집 Vol.33, No.1(B), pp. 346-348, 2006.
[7] 조수선 외, "기계학습 기반의 웹 이미지 분류", 한국정보처리학회 논문지 B, Vol.9-B, No.06, pp. 759-764, 2002.
[8] 조수선, "SOM 기반 웹 이미지 분류에서 고수준 텍스트 특징들의 효과", 한국정보처리학회논문지 B, Vol. 13-B, No.02, pp. 121-126, 2006.
[9] S.Mukherjea et al., "Automatically Determining Semantics for World Wide Web Multimedia Information Retrieval," Journal of Visual Languages and Computing, 10, pp. 585-606, 1999.
[10] M.Cascia et al., "Combining Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web," Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, 1998.
[11] S.Sclaroff et al., "ImageRover: A Content-Based Image Browser for the World Wide Web," Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 2-9, 1997.
[12] Z.Chen, et al., "Web Mining for Web Image Retrieval," Journal of the American Society for Information Science and Technology, Vol.52, No.10, pp. 831-839, 2001.
[13] C.Frankel, et al., "WebSeer: An Image Search Engine for the World Wide Web," Proceedings of IEEE Computer Vision and Pattern Recognition Conference, 1997.
[14] J.Smith et al., "WebSeek: An Image and Video Search Engine for the World Wide Web," in IS&T/SPIE Proceedings of Storage and Retrieval for Image and Video Database V, pp. 84-95, 1997.
[15] C.D.Manning et al., "Foundations of Statistical Natural Language Processing," The MIT Press, 2003.
[16] J.M.Ponte et al., "A language modeling approach to information retrieval," Proceedings of the ACM SIGIR, pp. 275-281, 1998.
[17] A.L.Berger, V.J.D.Pietra, and S.A.D.Pietra, A Maximum Entropy Approach to Natural Language Processing, Computational Linguistics Vol.22, No.1, pp. 39-71, 1996.

이 형 규
2005년 고려대학교 컴퓨터학과 학사 2007년~현재 고려대학교 컴퓨터·전파통신공학과 석박사통합과정. 관심분야는 자연어처리, 정보검색, 정보추출

김 민 정
2005년 고려대학교 컴퓨터학과 학사. 2007년 고려대학교 컴퓨터학과 석사. 2007년~현재 고려대학교 컴퓨터·전파통신공학과 박사과정. 관심분야는 자연어처리, 기계번역, 기계학습

홍 금 원
2000년 고려대학교 컴퓨터학과 학사. 2002년 고려대학교 컴퓨터학과 석사. 2007년~현재 고려대학교 컴퓨터·전파통신공학과 박사과정. 관심분야는 자연어처리, 한국어 자동 띄어쓰기, 기계번역, 언어모델

임 해 창
1981년 University of Missouri-Columbia 전산학과 학사. 1983년 University of Missouri-Columbia 전산학과 석사. 1990년 University of Texas at Austin 전산학과 박사. 1991년~현재 고려대학교 컴퓨터통신공학부 교수. 관심분야는 자연어처리, 정보검색, 한국어정보처리