하이퍼네트워크 연상메모리 기반의 이미지-텍스트 교차검색

(Image-Text Crossmodal Retrieval via Hypernetwork Memory)

지도교수 : 장병탁

이 논문을 공학학사 학위 논문으로 제출함.

2012년 6월 21일

서울대학교 공과대학
컴 퓨 터 공 학 부
천 효 선

2012년 8월

목차

1. 서론

2. 하이퍼네트워크 연상메모리 기반의 학습 모델
   2.1 하이퍼네트워크
   2.2 점진적 교차학습 모델

3. 멀티모달 하이퍼네트워크 학습 및 교차검색
   3.1 멀티모달 데이터 전처리
   3.2 멀티모달 하이퍼네트워크 학습
   3.3 텍스트 쿼리에 의한 이미지 검색

4. 실험 및 결과
   4.1 학습 후 쿼리에 의해 반환된 이미지
   4.2 학습에 따른 단어 수 및 이미지 패치 수 변화
   4.3 학습에 따른 특정 단어 관련 이미지 패치 수 변화

5. 결론

참고문헌

부록

하이퍼네트워크 연상메모리 기반의 이미지-텍스트 교차검색 (Image-Text Crossmodal Retrieval via Hypernetwork Memory)

천 효 선

초 록

인간은 여러 감각기관을 통해 들어온 감각정보들 사이의 연관관계를 파악하여 단어의 의미를 학습한다. 이러한 인간의 학습 과정을 반영한 단어 학습 모델은 다음의 두 가지 특성을 필요로 한다. 첫째, 다양한 형식으로 존재하는 데이터 간의 연관관계에 대한 학습이 가능한 멀티모달 연상모델이 필요하다. 둘째, 꾸준히 증가하는 데이터를 지속적으로 학습하기 위한 증분적 학습 기법이 필요하다. 본 연구에서는 이러한 두 특성을 만족시키는 하이퍼네트워크 연상메모리 기반의 이미지-텍스트 교차검색 기법을 제안한다. 제안한 검색 기법의 검증을 위해 본 논문에서는 어린이 학습용 만화영화 <Maisy>를 모델에 적용시켜보았다. 점진적 교차학습의 결과로, 텍스트 쿼리를 입력으로 주었을 때 쿼리와 연관된 이미지가 검색되는 것을 확인하였다.

1. 서론

태아들은 어머니의 뱃속에서 빠져나온 순간부터 무수히 많은 정보에 노출된다. 빛의 존재를 경험하면서 병원의 풍경과 같은 시각적 이미지들을 새로 체험하게 되고, 주위의 소리 혹은 어머니의 손길 등 다양한 감각정보들을 인지하기 위해 모든 감각기관들이 활발하게 작동하기 시작한다. 태아가 새로 받아들이게 되는 감각 정보들 속에는 언어도 여러 가지 형태로 포함되어 있다. 단어를 표현하는 소리 조합과 단어가 의미하는 대상의 시각적 형태 등 여러 감각 정보를 지속적으로 경험하면서 태아들은 언어를 점차 습득하게 된다. ‘엄마’라는 단어를 예로 들면, ‘엄마’라는 발음 정보와 어머니의 목소리, 어머니의 생김새, 몸짓, 피부색 등의 시각 정보, 어머니에게서 느껴지는 냄새와 체온 등 다양한 감각 정보를 꾸준히 받아들여야만 비로소 이 짧은 두 글자의 단어를 완벽히 받아들일 수 있다[1][2].

컴퓨터가 단어의 의미를 학습하는 데 있어서도 여러 형식의 데이터를 함께 학습할 필요가 있다. 컴퓨터상에는 이미 다양한 형식의 데이터가 존재하나 모두 독립적 형태로만 존재하고 있다. 컴퓨터가 이들 간의 연관관계를 학습할 수 있다면 효과적으로 단어의 의미를 표현 가능할 것이

다. 따라서 다양한 형식의 데이터 간의 연관관계에 대한 학습이 가능한 멀티모달 연상모델이 필요하다.

또한, 지속가능한 학습 모델이 필요하다[3]. 영유아기의 어린이들은 꾸준히 경험을 늘려가면서 새로운 지식을 얻고 이전의 잘못된 개념을 바로잡는다. 성인이 된 후에도 이전의 경험을 바탕으로 한 학습이 쉼 없이 이루어진다. 컴퓨터의 경우에도 꾸준한 학습이 필요한 것은 마찬가지다. 컴퓨터가 새로이 학습해야하는 데이터의 양과 종류는 빠른 속도로 증가하며, 학습 모델이 증가하는 데이터의 변화를 학습하기 위해서는 지속적인 학습이 가능한 점진적 기법이 필요하다.

이러한 두 가지 특성을 갖는 학습 모델로서 하이퍼네트워크(hypernetwork) 모델이 제시된 바 있다[4]. 하이퍼네트워크는 다수의 데이터들 간의 다대다 관계를 표현하기 위한 확장된 그래프 구조이다. 하이퍼네트워크는 데이터 사이의 연관관계를 표현할 수 있기 때문에 연상메모리로서 기능하게 된다. 하이퍼네트워크 상에서 정점(vertex)은 데이터를 의미하고, 하이퍼에지(hyperedge)는 두 개 이상의 정점들을 연결하여 연관관계를 나타내며, 그 연결 강도가 가중치를 통해 표현된다. 하이퍼에지가 다수의 정점들을 자유롭게 연결하는 것이 가능하기 때문에 하이퍼네트워크 모델을 통해 다양한 형식의 데이터 간의 연관관계를 쉽게 표현할 수 있다. 또한 하이퍼에지의 추가, 제거 및 수정이 용이하기 때문에 하이퍼네트워크를 기반으로 한 학습 모델은 지속적인 학습이 가능하다는 특징을 가진다.

본 연구에서는 하이퍼네트워크 연상메모리를 기반으로 이미지-텍스트 교차검색 기법을 제안한다. 하이퍼네트워크를 응용한 점진적 멀티모달 학습은 여러 연구를 통해 검증되어왔다[5][6][7]. 이러한 하이퍼네트워크 기반의 학습모델에 이미지 형식과 텍스트 형식의 데이터를 적용시켜 이들 간의 연관관계를 파악한 후 쿼리를 입력하여 이와 연관된 데이터들을 찾아낸다면 원하는 데이터를 효과적으로 얻을 수 있을 것이다.

이미지-텍스트 교차검색 기법의 검증을 위해 본 연구에서는 어린이 학습용 만화영화 'Maisy'를 이용하여 실험을 진행하였다. 학습 데이터는 자막 형식을 취하고 있는 텍스트 데이터와 각각의 자막이 시작되는 순간을 캡쳐한 이미지 데이터의 쌍으로 구성된다. 특히, 이미지 파일은 여러 시각적 언어들로 구성되어 있으므로 최대 안정 외측 구역(MSER, maximally stable external regions)[8]기법과 크기불변 특징 변환(SIFT, scale-invariant feature transform)[9]기법, 그리고 K-means 클러스터링 기법을 이용한 전처리 과정을 거쳐서 다수 개의 이미지 패치들로 나누어서 사용하였다. 전처리과정을 거친 학습 데이터를 순차적으로 학습 모델에 적용시키면서 이미지-텍스트 교차검색이 효과적으로 일어나는지 확인하였다.

본 연구에서는 이미지-텍스트 교차검색을 위한 하이퍼네트워크 연상메모리 모델의 학습 효과를 알아보기 위해 세 가지 실험을 진행하였다. 첫 번째 실험으로 특정 단어가 입력되었을 때 어떤 이미지가 검색되는지를 알아보았고, 학습의 결과로 실제 단어와 연관된 이미지가 조회되는 것을 확인하였다. 두 번째 실험으로 학습의 진행에 따라 모델 내에 존재하는 단어의 수와 이미지

패치 수가 어떻게 변화하는지 알아보았고, 그 결과 모델 내의 단어 수 및 이미지 패치 수가 증가하는 것을 관찰하여 학습의 점진적 진행을 확인하였다. 마지막으로 학습이 이루어지면서 특정 단어와 연관된 이미지 패치의 수가 어떻게 변화하는지를 알아보았고, 그 결과 단어와 연관된 이미지 패치의 수가 증가하는 것을 관찰하여 점진적 교차학습이 이루어졌음을 확인하였다.

본 논문은 다음과 같이 구성되어있다. 2장에서는 하이퍼네트워크 연상메모리 기반의 점진적 교차학습 모델을 소개한다. 3장에서는 실험 방법에 대해 설명하고 4장에서 실험 결과를 분석한다. 5장에서 결론 및 향후 연구 과제를 제시한다.

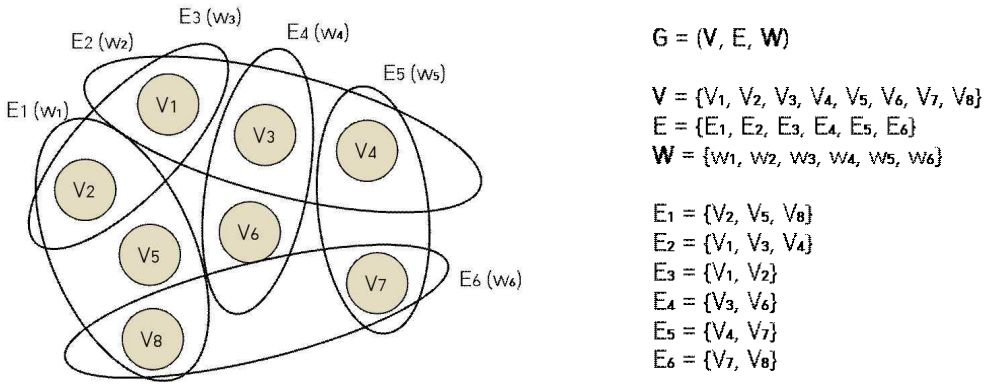
2. 하이퍼네트워크 연상메모리 기반의 학습 모델

이 장에서는 하이퍼네트워크 모델을 소개하고, 하이퍼네트워크 연상메모리 기반의 점진적 교차학습 알고리즘을 살펴본다.

2.1 하이퍼네트워크

하이퍼네트워크(hypernetwork)는 다수의 정점(vertex)들 간의 다대다 관계를 하이퍼에지(hyperedge)를 통해 표현하는 하이퍼그래프(hypergraph) 구조이다[4]. 하이퍼네트워크 G는 정점의 집합 V와 하이퍼에지의 집합 E, 하이퍼에지의 가중치 집합 W로 표현된다(식 1). 정점은 데이터를 의미하고, 하이퍼에지는 정점들 간의 연관관계를 나타낸다. 일반적인 그래프 구조에서는 에지(edge)가 오직 한 쌍의 정점만을 연결할 수 있으나, 하이퍼그래프를 구성하는 하이퍼에지(hyperedge)는 두 개 이상의 정점을 연결하여 그들 사이의 연관관계를 표현할 수 있다. 정점들 사이의 연관관계가 얼마나 강한지는 하이퍼에지의 가중치를 통해 표현된다. 하이퍼에지가 연결하는 정점들 사이의 관련성을 높을수록 높은 가중치를 가지게 된다. 그림 1은 하이퍼네트워크의 구조를 그림으로 나타낸 것이다.

$$G = (V, E, W) \qquad \text{(식 1)}$$



[그림 1] 8개의 정점과 6개의 하에퍼에지로 구성된 하이퍼그래프의 예

하이퍼네트워크는 여러 형식의 데이터 간의 연관관계 표현이 용이하기 때문에 연상메모리로서 기능하게 된다. 특히 하이퍼에지가 연결하는 정점들의 수에 제한이 없고, 정점들이 나타내는 데이터의 형식도 자유롭기 때문에, 사실상 무한한 종류의 데이터 간의 연관관계를 하이퍼네트워크 내부에 저장할 수 있다.

또한 하이퍼네트워크를 기반으로 한 학습 모델은 점진적 학습이 가능하다는 특징을 가진다. 하이퍼에지의 추가, 제거 및 가중치 수정을 통해 구조의 수정이 용이하기 때문에 새로운 데이터가 발생했을 때 잘 대처할 수 있다.

2.2 점진적 교차학습 모델

하이퍼네트워크 연상메모리 기반 학습은 인지, 생성, 예측, 보완 네 단계의 반복을 통해 이루어진다(알고리즘 1)[6]. 새로운 데이터가 관찰되면 새로운 정점을 만들어낸다(인지). 인지된 정점들을 랜덤하게 연결시켜 새로운 하이퍼에지를 만들어 내고 하이퍼그래프에 추가한다(생성). 새로 학습하는 데이터 간의 연관관계를 파악하는 것이다. 인지된 정점들은 기존 개념의 수정에도 사용된다. 기존에 가지고 있던 하이퍼네트워크 중에서 인지된 정점들과 관련 있는 하이퍼에지들을 유사성 검사를 통해 추출해낸 뒤(예측), 추출한 하이퍼에지의 가중치를 수정한다(보완).

G: hypergraph,
E: hyperedge, V: vertex, W: weight set
X_T: text data, X_I: image data
R: number of iterations for correction

$G \leftarrow \{\}$
For $n \leftarrow 1$ to N
    $V \leftarrow$ Percieve (X_T, X_I);
    For $i \leftarrow 1$ to R
        $E \leftarrow$ Generate (V);
        $G' \leftarrow$ Predict (G, V);
        $W' \leftarrow$ Correct (G', V, W);
    End
    $G \leftarrow G \cup E, \quad W \leftarrow W'$
End

[알고리즘 1] 학습 알고리즘. 학습은 인지, 생성, 예측, 보완 네 단계로 이루어진다.

예측 단계에서 기존 하이퍼네트워크의 보완을 위해서 인지된 정점들과 관련된 하이퍼에지를 유사성 검사를 통해 추출해내는데, 유사도가 미리 설정한 한계값 θ를 넘으면 서로 관련이 있다고 판단하는 것이다(알고리즘 2).

G' ← Predict (G, V);

G' ← {}
For n ← 1 to size(G)
    score ← similarity (V, G_n);
    If (score > θ)
        G' ← G' ∪ G_n
    End
End

[알고리즘 2] 학습 알고리즘 중 예측단계. 인지된 정점과 관련된 하이퍼에지를 유사성 검사를 통해 추출한다.

보완 단계에서의 하이퍼에지 가중치 수정도 유사도 값을 이용해 이루어진다(알고리즘 3). 인지된 정점들과 하이퍼에지 사이의 유사성이 높으면 학습의 강화로 인해 가중치가 높아진다.

W' ← Correct (G', V, W);

W' ← {}
For n ← 1 to size(G')
    score ← similarity (V, G_n');
    w ← $\alpha w + (1-\alpha) score$
    W' ← W' ∪ w
End

[알고리즘 3] 학습 알고리즘 중 보완단계. 유사도 값을 이용해 가중치를 수정한다.

예측 단계와 보완 단계는 하나로 결합될 수 있다(알고리즘 4). 또한 모든 하이퍼에지의 가중치 새로운 데이터가 관찰될 때마다 인지된 정점과의 관련성 여부와 관계없이 λ의 비율로 감소하게 하여 시간에 따른 망각효과를 더해준다.

W' ← Predict_and_Correct (G, V, W);

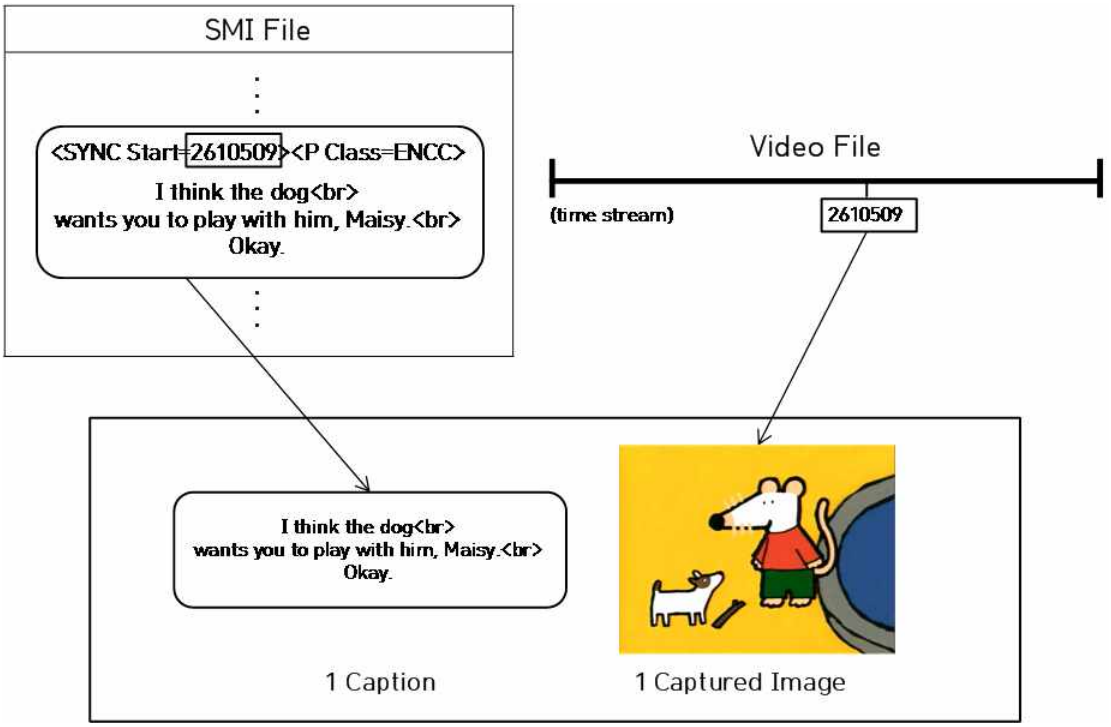
W' ← {}
For n ← 1 to size(G)
    score ← similarity (V, G_n);
    If (score > θ)
        w ← $\alpha w + (1-\alpha) score$
    End
    w ← λw
    W' ← W' ∪ w
End

[알고리즘 4] 학습 알고리즘 중 예측 단계와 보완단계의 결합. 시간에 따른 망각효과를 더해준다.

3. 멀티모달 하이퍼네트워크 학습 및 교차검색

3.1 멀티모달 데이터 전처리

본 연구에서는 실험 데이터로 어린이 학습용 만화영화 <Maisy>를 사용하였다. 학습 데이터는 텍스트와 이미지 두 가지 형식이다. 텍스트 데이터는 자막 파일 형태로 주어지고, 이미지 데이터는 이미지 파일 형태로 주어진다. 학습데이터 한 쌍은 자막 파일을 이루는 자막 하나와 그 자막이 시작되는 순간의 영상을 캡쳐한 이미지 데이터로 구성된다(그림 2).



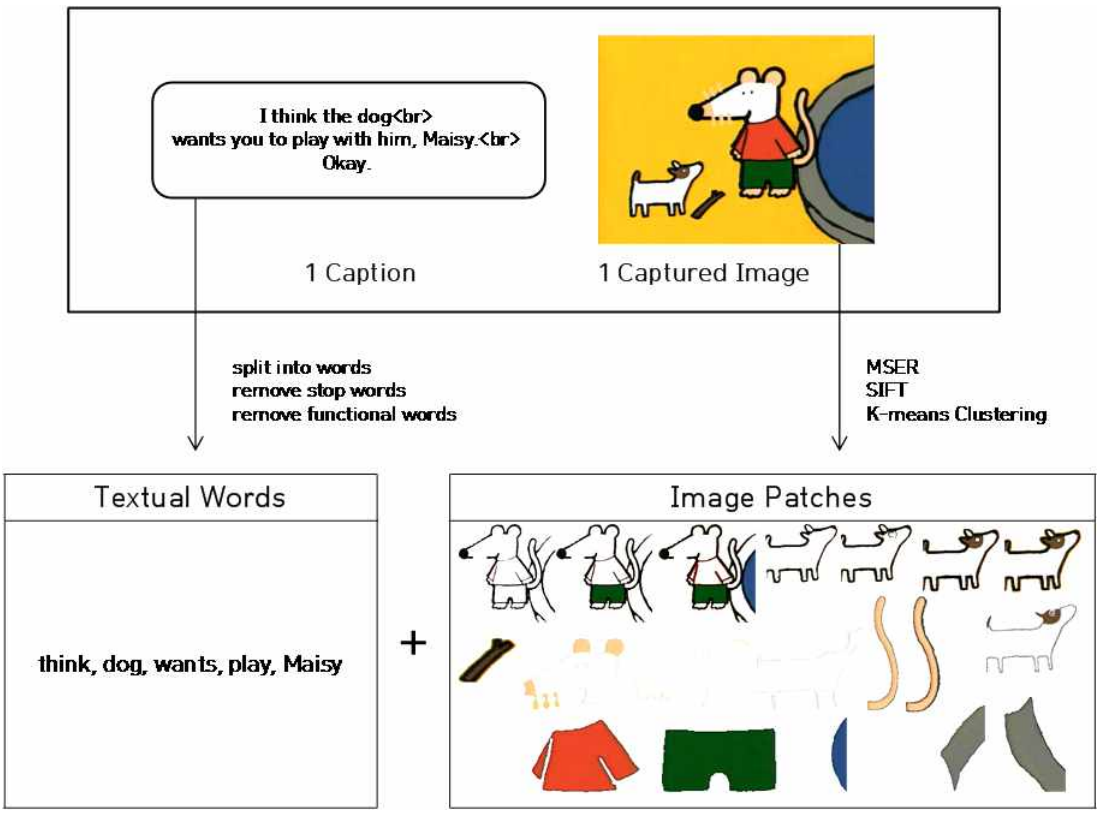
[그림 2] 학습 데이터 한 쌍의 구성. 한 쌍의 학습데이터는 자막 형식의 텍스트 데이터와 각각의 자막이 시작되는 순간의 영상을 캡쳐한 이미지로 이루어진다.

자막 파일 및 이미지 파일 형태로 주어지는 학습 데이터를 실험에 사용하기 위해서는 전처리 과정을 거쳐야한다(그림 3). 문장 단위로 주어지는 텍스트 데이터의 경우에는 단어 단위로 나누어져야 한다. 문장을 단어의 집합으로 나눈 후, 의미가 없는 단어인 불용어(stop word) 및 문법어(functional word)들을 제거한다. 데이터에서 제거된 불용어 및 문법어들의 목록은 부록에서 확인할 수 있다.

이미지 데이터 처리는 더 복잡하다. 일반적인 이미지 파일은 여러 시각적 언어들이 결합되어 있다. 학습을 위해서는 통합적 이미지를 개별 요소들로 분리해낼 필요가 있다. 본 연구에서는 이미지 데이터 처리를 위해 두 가지 기법이 함께 사용되었다. 최대 안정 외측 구역(MSER,

maximally stable external regions)[8]기법과 크기불변 특징 변환(SIFT, scale-invariant feature transform)[9]기법이 그것이다. 본 연구에서는 MSER 기법을 이용해 그림에서 의미가 있을만한 부분들을 분리해낸 뒤, 각 구역에 대해 SIFT 기법을 적용하여 두드러진 특성을 추출하여 이미지 패치를 만든다. 또한 이미지를 이루는 구역들의 SIFT 특성 간에 중복이 존재하기 때문에, 특성들을 보다 효율적으로 나타내기 위해 K-means 클러스터링 기법이 사용되었다. K-means 클러스터링 과정을 거치면 이미지 패치들은 K개의 SIFT 특성으로 나타내어진다. 실험에서는 K값을 100으로 설정하여 각각의 이미지 패치들이 100개의 SIFT 특성들로 표현되도록 하였다.

전처리 과정을 거친 데이터는 총 323개의 단어와 9490개의 이미지 패치로 이루어지게 된다.

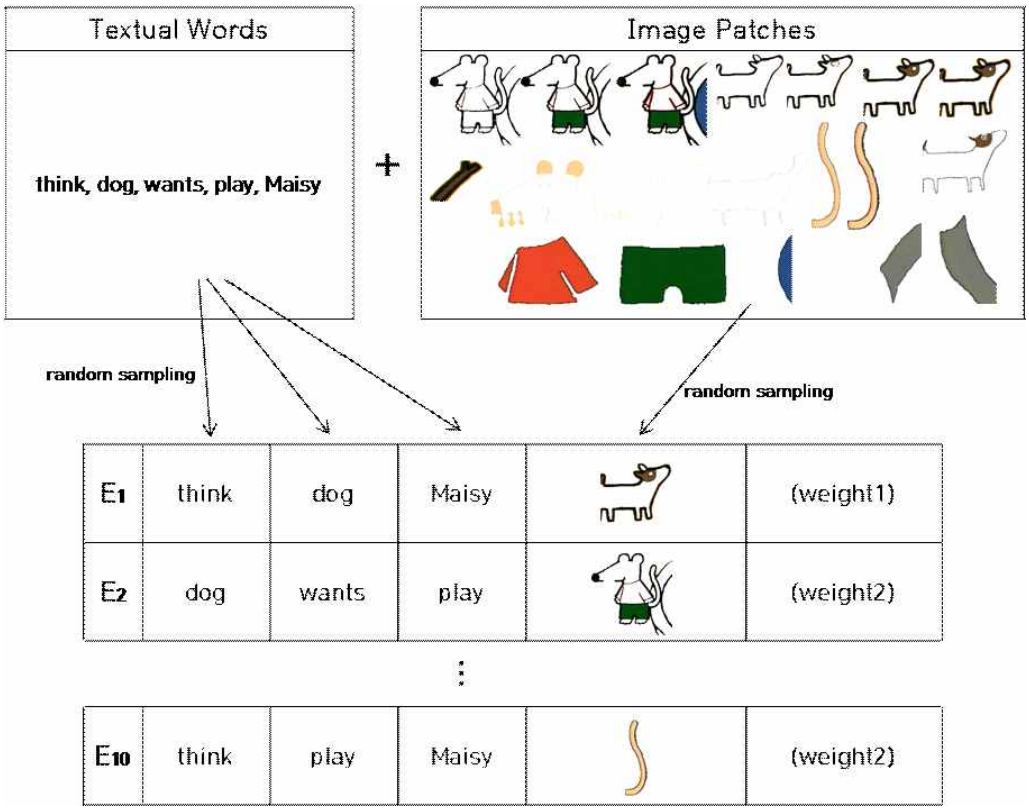


[그림 3] 전처리 과정을 거친 학습 데이터 한 쌍. 전처리과정을 통해 자막 파일 형식의 텍스트 데이터와 이미지 파일 형식의 이미지 데이터는 각각 단어와 이미지 패치 단위로 쪼개어진다.

3.2 멀티모달 하이퍼네트워크 학습

한 쌍의 학습 데이터 당 10개의 새로운 하이퍼에지가 생성되도록 하였다. 하이퍼에지는 한 쌍의 학습 데이터를 이루는 단어 집합과 이미지 패치 집합 중 3개의 단어와 1개의 이미지 패치를 무작위로 뽑아 연결한다(그림 4). 학습과 관련된 다른 파라미터들은 표 1과 같이 설정하였다.

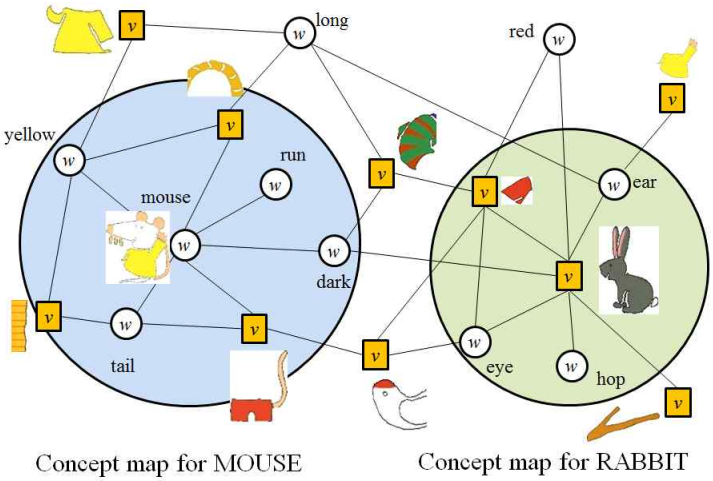
학습이 제대로 이루어진다면 단어와 이미지 패치들은 그림 5와 같이 단어와 이미지패치들이 하이퍼네트워크에 자리 잡게 된다.



[그림 4] 하이퍼에지의 생성. 한 쌍의 학습 데이터 당 10개의 하이퍼에지가 생성된다. 하이퍼에지 하나는 3개의 단어와 1개의 이미지패치를 무작위로 선택하여 연결한다.

Parameters	Values
R (number of iterations for correction)	5
θ (similarity threshold)	0.7
λ (decay ratio)	0.99
α (constant for balancing previous weight value and new similarity score)	0.9

[표 1] 파라미터 설정



[그림 5] 학습 후 단어 'mouse'와 'rabbit'의 개념지도

3.3 텍스트 쿼리에 의한 이미지 검색

이미지-텍스트 연관관계 학습의 결과로 검색이 효과적으로 이루어지는지 알아보기 위해, 특정 단어가 입력으로 주어졌을 때 이미지를 반환시켜 본다. 본 연구에서 사용한 이미지-텍스트 교차 검색 기법을 알고리즘 5에 나타내었다. 입력 단어가 다른 데이터와 연결되어 있는 하이퍼에지를 하이퍼네트워크로부터 추출해낸 뒤, 이들 하이퍼에지를 가중치가 높은 순서로 정렬한다. 하이퍼에지를 입력 단어와 연관성 높은 순서로 배열하기 위해서다. 정렬이 끝나면 하이퍼에지가 만들어진 원래 이미지들을 반환한다.

ImageSet ← Retrieval (Query, G, W);
$G' \leftarrow \{\}$
$W' \leftarrow \{\}$
For $n \leftarrow 1$ to size(G)
    If (Query $\in$ G_n)
        $G' \leftarrow G' \cup G_n$
        $W' \leftarrow W' \cup W_n$
    End
End

ImageSet ← {}
$G' \leftarrow$ SortByWeight (G', W', descend)
For $n \leftarrow 1$ to size(G')
    ImageSet ← ImageSet $\cup$ G_n'(Image)
End

[알고리즘 5] 텍스트 쿼리에 의한 이미지 검색 알고리즘

4. 실험 결과

이미지-텍스트 교차검색을 위한 하이퍼네트워크 연상메모리 모델의 학습 효과를 다음 세 가지 실험을 통해 알아보았다. 첫 번째 실험으로 학습 후 특정 단어가 입력되었을 때 실제로 단어와 연관된 이미지가 조회되는지를 알아보았다. 두 번째 실험으로 학습이 점진적으로 진행되어감에 따라 모델 내에 존재하는 단어의 수와 이미지 패치 수가 어떻게 변화하는지 알아보았다. 마지막으로 학습이 이루어지면서 특정 단어와 연관된 이미지 패치의 수가 어떻게 변화하는지 알아보았다.

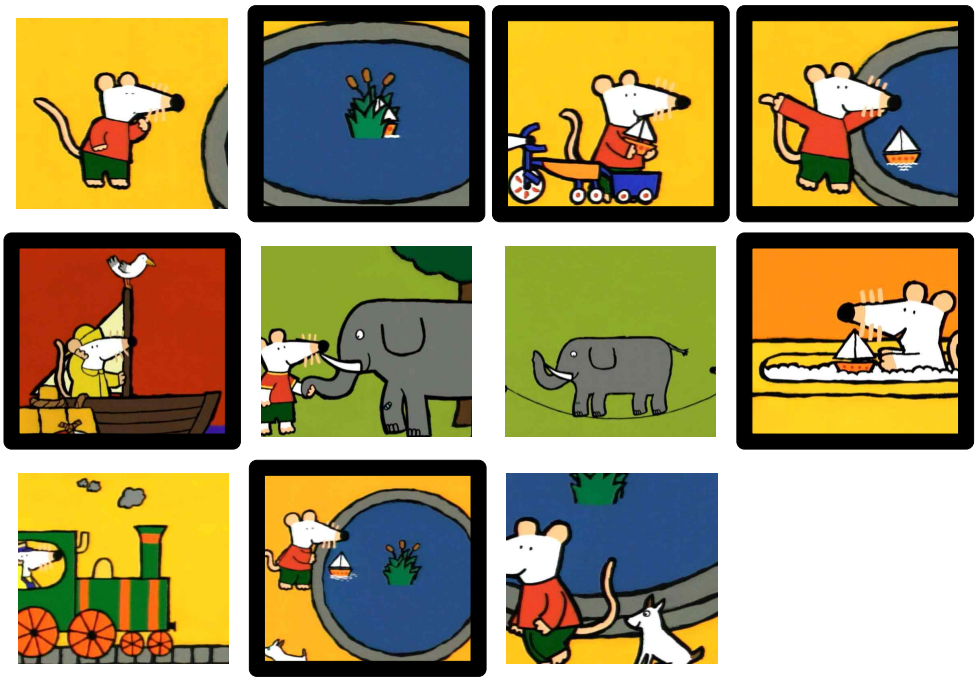
4.1 학습 후 쿼리에 의해 검색된 이미지

첫 번째 실험으로 학습 후에 특정 단어가 입력으로 주어졌을 때, 하이퍼에지 상에서 해당 단어와 연관된 이미지를 반환시켜보았다. 실제로 관련이 있는 이미지가 조회되는지 관찰하여 학습의 효과를 확인하기 위한 것이다. 단어 'maisy'와 'bird', 'boat'를 입력으로 주었을 때 모델로부터 반환된 12개의 이미지를 각각 그림 6과 그림 7, 그림 8에서 확인할 수 있다('boat'의 경우에는 학습 데이터 11쌍에서만 등장했기 때문에 11개의 이미지만 반환되었다.) 대부분 입력 단어가 나타내는 대상이 포함된 이미지들이 검색되어 학습이 잘 이루어졌음을 알 수 있다.

[그림 6] 단어 'maisy'가 주어졌을 때 조회된 이미지

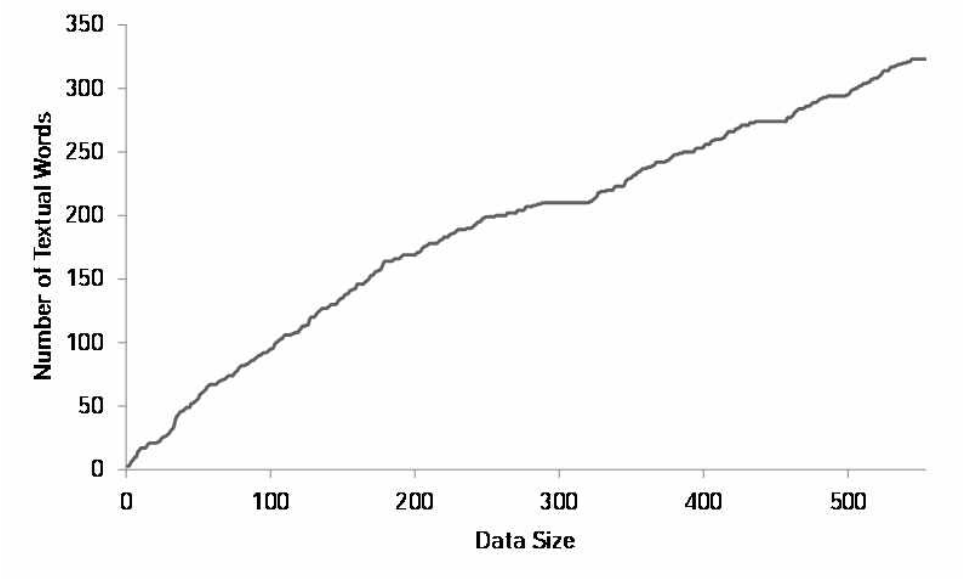
‘bird’	

[그림 7] 단어 ‘bird’가 주어졌을 때 조회된 이미지

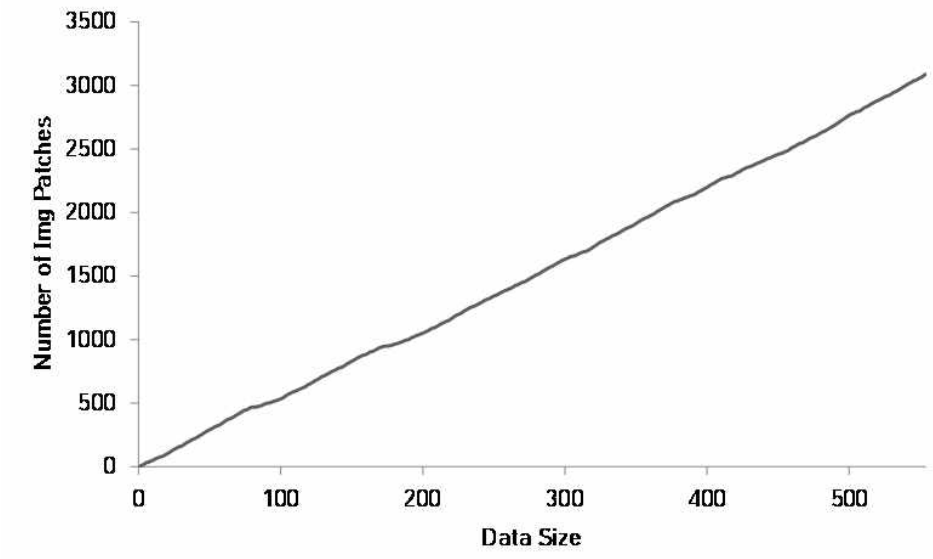
‘boat’	

[그림 8] 단어 ‘boat’가 주어졌을 때 조회된 이미지

4.2 학습에 따른 단어 수 및 이미지 패치 수 변화



[그림 9] 학습 진행에 따른 단어 수 변화. 모델 내에 존재하는 단어의 수가 증가하였다.



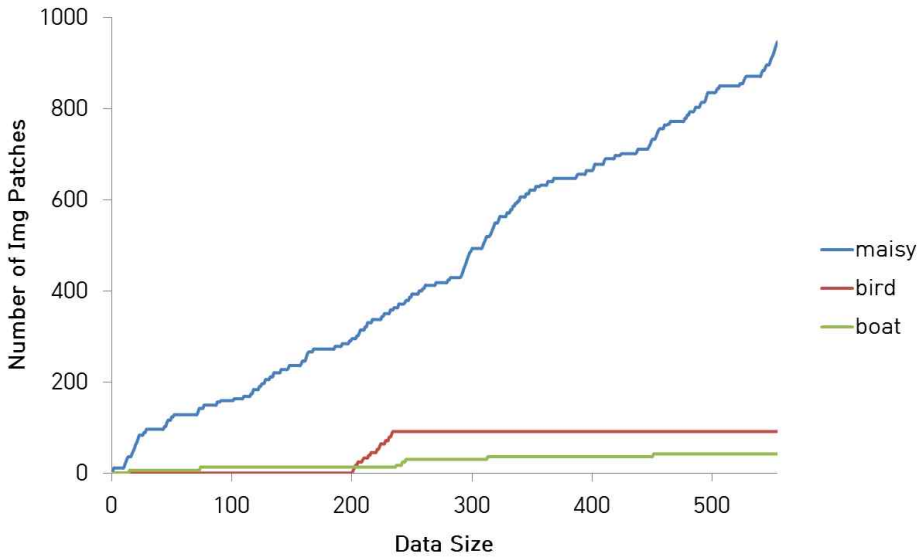
[그림 10] 학습 진행에 따른 이미지 패치 수 변화. 모델 내에 존재하는 이미지 패치의 수가 증가하였다.

두 번째 실험으로 학습의 진행에 따라 모델 내에 존재하는 단어의 수가 어떻게 변화하는지 알아보았다. 그 결과 학습 데이터의 추가에 따라서 모델 내 단어 수도 증가하는 것을 확인하였다

(그림 9). 또한 학습의 진행에 따라 모델 내에 존재하는 이미지 패치의 수가 어떻게 변화하는지 알아보았고, 그 결과 학습 데이터의 추가에 따라서 모델 내 이미지 패치의 수도 증가하는 것을 관찰하였다(그림 10). 학습의 진행에 따라 모델 내부의 데이터 수도 같이 증가하여 학습의 점진적 진행을 확인할 수 있었다.

이미지 패치의 수가 거의 선형으로 증가하는 것에 비해, 단어 수는 증가와 유지를 반복하는 양상을 보였다. 이는 이미지 패치가 서로 다른 학습 데이터 사이에서 공유되지 않지만 단어는 여러 번 반복되어 나타날 수 있기 때문에 나타나는 현상이다. 한 단어에 대해 다양한 시각적 이미지를 학습하는 것이다.

4.3 학습에 따른 특정 단어 관련 이미지 패치 수 변화



[그림 11] 학습 진행에 따른 특정 단어 관련 이미지 패치 수 변화.
특정 단어와 관련된 이미지 패치의 수가 증가하였다.

학습의 진행에 따라 모델 내에서 세 단어 'maisy'와 'bird', 'boat'와 연관된 이미지 패치의 수가 어떻게 변화하는지 관찰하여, 한 단어에 대해 학습이 어떻게 이루어지는지 알아보았다. 결과는 그림 11에서 볼 수 있다. 주인공의 이름을 가리키는 단어 'maisy'는 이 만화에서 빈번하게 등장하여 학습 빈도가 상대적으로 높았다. 그 결과 단어 'maisy'는 'bird'나 'boat'와 같은 다른 단어들에 비해 단어와 연관된 이미지 패치의 수가 지속적으로 증가한 것을 볼 수 있다. 단어 'bird' 및 단어 'boat'와 연관된 이미지 패치의 수는 특정 시점에서만 잠시 증가하다가 일정하게 유지되는 경향을 보였다. 특히 단어 'bird'는 200번째 학습 부근에서만 집중적으로 학습된 모습을 관찰할 수 있다.

또한 그림 11의 결과를 이미지 조회 실험의 결과와 비교해보았다. 단어 ‘maisy’를 쿼리로 하였을 때 조회된 이미지들은 그림 6에서 확인할 수 있고, 단어 ‘bird’와 ‘boat’를 쿼리로 하였을 때 조회된 이미지들은 각각 그림 7과 그림 8에서 확인할 수 있다. 검색된 이미지 중 단어가 가리키는 대상이 직접적으로 등장하는 이미지의 수를 세어보았을 때, 만화의 주인공을 가리키는 단어인 ‘maisy’는 단어 ‘boat’에 비해 학습 빈도가 매우 높았음에도 불구하고 그 수가 비슷했다. 빈번하게 학습된 단어 ‘maisy’는 단어가 직접적으로 가리키는 대상뿐만 아니라, 단어와 연관된 다른 이미지에 대해서도 학습이 이루어진 것으로 보인다. 특정 시점에서만 집중적으로 학습된 단어 ‘bird’의 경우에는 단어의 대상이 직접 등장하는 이미지가 상대적으로 많이 검색되었다.

5. 결론

본 연구에서는 하이퍼네트워크 연상메모리를 기반으로 이미지-텍스트 교차검색 기법을 제안하였다. 여러 연구를 통해 검증된 하이퍼네트워크 기반의 학습모델에 이미지 형식과 텍스트 형식의 데이터를 적용시켜 이들 간의 연관관계를 파악한 후, 쿼리를 입력하여 이와 연관된 데이터들을 찾아내는 것이다. 제안한 기법을 검증하기 위해 어린이 학습용 만화영화 <Maisy>를 이용하여 실험을 진행하였다. 텍스트와 이미지 두 가지 형식의 데이터를 모델에 순차적으로 학습시켜 보았고, 세 가지 실험을 통해 학습이 효과적으로 이루어짐을 확인하였다. 첫 번째 실험으로 특정 단어가 주어졌을 때 어떤 이미지가 검색되는지를 알아보았고, 그 결과 실제 단어와 관련 있는 이미지들이 조회됨을 확인하여 학습이 잘 이루어졌음을 알 수 있었다. 두 번째 실험으로는 학습 데이터의 추가에 따라 모델 내에 존재하는 단어의 수와 이미지 패치 수가 어떻게 변화하는지 알아보았고, 그 결과 모델 내의 단어 수 및 이미지 패치 수가 지속적으로 증가하는 것을 관찰하여 학습의 점진적 진행을 확인하였다. 특히, 단어의 수는 이미지 패치의 수보다 느리게 증가하여 한 단어에 대해 여러 시각적 이미지를 학습하는 것을 확인하였다. 마지막으로 학습이 이루어지면서 특정 단어와 연관된 이미지 패치의 수가 어떻게 변화해 가는지를 알아보았고, 그 결과 학습 빈도가 높은 단어의 경우 연관된 이미지 패치의 수가 지속적으로 증가하는 것을 관찰하였다. 또한 이들 단어에 의해 검색된 이미지를 비교해봤을 때, 짧은 시간에 집중적으로 학습된 단어의 경우 단어의 대상이 직접적으로 등장하는 이미지가 많이 검색되었고, 지속적으로 빈번하게 학습된 단어의 경우 단어의 대상이 직접적으로 등장하는 이미지뿐만 아니라, 단어와 관련있는 다른 이미지들도 검색되는 것을 볼 수 있었다.

본 연구를 통해 하이퍼네트워크 연상메모리 기반의 점진적 교차학습 모델이 텍스트-이미지 검색에도 활용될 수 있는 가능성을 보았다. 그러나 이를 위해서는 속도 문제에 대한 개선이 필요할 것이다. 본 연구에서 알아본 모델은 학습 데이터의 양이 증가할수록 학습 속도가 느려진다는 단점이 있다. 인간은 성인이 된 후에도 상대적으로 빠른 학습 속도를 유지하는 것을 생각한다면, 그리고 학습 데이터의 양이 지속적으로 증가하는 것을 고려한다면, 속도 문제에 대한 개선은 필수적일 것이다. 또한 본 연구에서는 텍스트 쿼리에 의한 이미지 검색 실험만을 수행하였는데, 이미지 쿼리에 의한 데이터 검색 실험을 통해 검증이 된다면 이미지-텍스트 양방향 검색이 가능해질 것이다. 더 나아가, 이미지와 텍스트 두 형식의 데이터뿐만 아니라 더 다양한 형식의 데이터에 대한 학습이 이루어지는 것을 확인한다면 일반적인 데이터 교차검색에도 하이퍼네트워크 연상메모리 모델이 효과적으로 적용될 수 있을 것이다.

감사의 글

논문에 도움을 주신 서울대 컴퓨터공학부 바이오지능(biointelligence) 연구실의 하정우 선배님께 감사의 말씀을 드립니다.

참고문헌

[1] P. Kuhl, “Early Language Learning and Literacy: Neuroscience Implications for Education”, *Mind, Brain, and Education*, vol. 5, issue 3, pp. 128−142, 2011.

[2] M. Coene, K. Schauwers, S. Gillis, J. Rooryck, P. Govaerts, “Genetic predisposition and sensory experience in language development: Evidence from cochlear-implanted children”, *Language and Cognitive Processes*, vol. 26, issue 8, 2011.

[3] J. Evans, J. Saffran, “Statistical Learning in Children With Specific Language Impairment”, *Journal of Speech, Language, and Hearing Research*, vol. 52, pp. 321-335, 2009.

[4] B.-T. Zhang, “Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory,” *IEEE Computational Intelligence Magazine*, vol. 3, no. 3, pp. 49-63, 2008.

[5] J.-W. Ha, B.-H. Kim, H.-W. Kim, W.C. Yoon, J.-H. Eom, and B.-T. Zhang, “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)*, pp. 274-277, 2009

[6] B.-T. Zhang, J.-W. Ha, and M. Kang, “Sparse population code models of word learning in concept drift”, *In Proceedings of Annual Meeting of the Cognitive Science Society (CogSci 2012)*, 2012. (to appear)

[7] J.-W. Ha, B.-J. Lee, B.-T. Zhang, “Text-to-Image Retrieval Based on Incremental Association via Multimodal Hypernetworks”, *IEEE Systems, Man, & Cybernetics Society (SMC 2012),* 2012. (submitted)

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” *Image and Vision Computing*, vol. 22, no. 10, pp. 761-767, 2004.

[9] D. G. Lowe, Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 2, no. 60, pp. 91-110, 2004.

부록:
불용어(Stop Words) 및 문법어(Functional Words) 목록

1. Greeting

'hi'	'bye'	'thank'	'ba-bye'
'hello'	'bye-bye'	'good-bye'	

2. Onomatopoeia

'oh'	'shh'	'whoops'	'bing'
'ah'	'yeah'	'whoa-ho'	'whee'
'whoa'	'phew'	'hoooray'	'uh-uh'
'mm'	'whoops'	'hooray'	'ohhh'
'mmm'	'ho'	'ho-ho'	'ahhh'
'mmmm'	'ohh'	'eww'	'hurray'
'wow'	'uh'	'whew'	'mmm-mmm'
'ha'	'ahoy'	'whoo-hoo'	'who-o-oa'
'ooooh'	'yum'	'whoo'	'choo-choo'
'ooh'	'huh'	'eh'	'whoopee'
'uh-oh'	'uh-uh-uh'	'heh'	'aww'
'uh-huh'	'mm-hmm'	'whoo-whoo-whoo'	'ooh-ay'
'ahh'	'ha-ha'	'boo'	'sure'
'wha'	'aha'	'hmm'	'whoop'
'o-oh'	'ow'	'hey'	'bravo'
'ee-yew'	'yea'	'goodjob'	'hurrey'
'oh-ho'	'oops'	'oh-ho-ho'	

3. Article

'a'	'an'	'the'

4. Be Verb

'be'	'being'	'is'	'was'
'been'	'am'	'are'	'were'

5. Do & Have

'do'	'did'	'have'	'has'
'does'	'done'		

6. Conjunction

'if'	'or'	'so'	'else'
'and'	'yet'	'whether'	'because'
'nor'			

7. Auxiliary Verb

'will'	'could'	'maybe'	'shall'
'would'	'may'	'must'	'should'
'can'	'might'		

8. Personal Pronoun

'i'	'yourself'	'herself'	'our'
'my'	'yourselves'	'hers'	'ours'
'mine'	'he'	'they'	'ourselves'
'me'	'his'	'their'	'us'
'myself'	'him'	'theirs'	'it'
'you'	'himself'	'them'	'its'
'yours'	'she'	'themselves'	'itself'
'your'	'her'	'we'	

9. Reciprocal Pronoun

'other'	'each'	'another'

10. Demonstrative Pronoun

'this'	'that'	'those'	'there'
'these'			

11. Indefinite pronoun & Quantifier

'yes'	'everything'	'none'	'each'
'ok'	'something'	'very'	'none'
'okay'	'anything'	'many'	'such'
'no'	'nothing'	'much'	'either'
'not'	'everybody'	'more'	'neither'
'never'	'somebody'	'most'	'always'
'some'	'anybody'	'almost'	'lot'
'any'	'nobody'	'than'	'lots'
'thing'	'everyone'	'all'	'too'
'things'	'someone'	'both'	'well'

12. Interrogative Pronoun

'who'	'when'	'what'	'why'
'whom'	'whenever'	'whatever'	'which'
'whomever'	'where'	'how'	'whichever'
'whose'	'wherever'	'however'	'that'
'whosever'			

13. Preposition

'about'	'below'	'into'	'through'
'above'	'beneath'	'like'	'throughout'
'across'	'beside'	'near'	'though'
'after'	'between'	'nearly'	'although'
'again'	'beyond'	'next'	'till'
'against'	'but'	'now'	'to'
'ago'	'by'	'of'	'toward'
'ahead'	'despite'	'off'	'towards'
'along'	'down'	'on'	'under'
'among'	'during'	'onto'	'underneath'
'already'	'even'	'out'	'until'
'around'	'except'	'outside'	'up'
'as'	'here'	'over'	'upon'
'at'	'for'	'past'	'with'
'away'	'from'	'since'	'within'
'before'	'in'	'then'	'without'
'behind'	'inside'		

14. Special Case

'let' 'able' 'd' 'just'
'please' 'where'