

온톨로지를 이용한 뉴스 비디오의 자동 분류 기법

김하은^o 이동호

한양대학교 컴퓨터공학과

{hekim^o, dhlee72}@cse.hanyang.ac.kr

Full-automatic Classification Technique of News Video using Domain Ontologies

Ha-Eun Kim^o, Dong-Ho Lee

Knowledge and Data Engineering Laboratory

Dept. of Computer Science & Engineering Hanyang University

요 약

본 논문은 온톨로지를 이용하여 뉴스 비디오를 분야별로 자동으로 분류하는 효율적인 기법을 제안한다. 이를 위해서 뉴스 비디오를 파싱하여 키프레임(Key frame), 샷(Shot), 씬(Scene)으로 나누고 키프레임과 샷에서 특징 정보를 추출한다. 추출된 특징 정보를 이용하여 샷의 키워드 집합을 만들고 이를 이용하여 씬의 키워드 집합을 만든다. 그리고 씬의 키워드 집합을 어휘 온톨로지와 뉴스 온톨로지에 매칭(추론)하여, 씬의 분야를 결정한다. 또한 이렇게 결정된 분야를 기반으로 서로 유사한 씬들을 자동으로 그룹화하는 방법을 제안한다.

1. 서 론

최근 비디오 자동 분석에 관한 많은 연구가 진행되고 있다. 이는 비디오 서비스를 위해 필요한 필수 조건인 고성능의 컴퓨터 환경과 빠른 인터넷 환경이 구축되면서 비디오 서비스에 대한 일반 사용자의 수요가 폭발적으로 증가되기 때문이다. 그중에서도 뉴스 비디오는 영상 자체가 상당히 구조적으로 구성되어 있기 때문에 여러 연구에서 비디오 검색을 위한 다양한 방법들을 제안하고 있다. 일반적으로 뉴스는 시간 순으로 앵커가 말하는 부분과 뒤이어서 나오는 비 앵커구간의 기자가 설명하는 화면으로 규칙적으로 구성되어 있다. 따라서 이를 하나의 뉴스 단위로 설정하여 여러 조각(이하 씬(Scene))으로 나누어 놓을 수 있다. [그림 1]은 이러한 뉴스 비디오의 구조를 보여주고 있다. 뉴스 비디오의 구조를 효과적으로 분석하기 위해서는 앵커 구간의 정확한 추출이 선행되어야 한다. 앵커 구간의 추출을 위해서 여러 가지 방법이 제시되었다[1]. 그러나 앵커 구간의 정확한 추출법만 제시되고 씬의 의미에 대해서는 연구가 활발히 진행되고 있지 못하다.

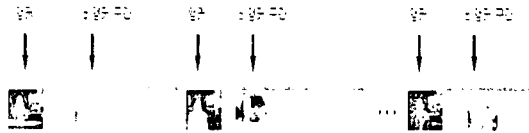


그림 1 뉴스 비디오의 일반적 구성

뉴스 비디오의 씬들은 기사간의 유사도를 가지고 있다. 이는 일반적인 비디오는 가지지 못하는 특성이다. 따라서 이 특성을 이용하여 뉴스 비디오에 대해서 효과적으로 계층을 구분 짓는 것이 가능하다. 이를 이용하기 위해서는 뉴스 비디오를 위한 새로운 계층형 구조와 이를 구축해 내기 위한 기법이 필요하다.

본 논문에서는 뉴스 비디오 씬들의 의미적 유사도를 측정하여 자동으로 분류하는 기법을 제안한다. 또한 뉴스 비디오의 씬들 간의 의미적 관계를 정의할 수 있는 온톨로지를 제안하고 제안된 온톨로지를 이용하여 기사간의 의미적 유사도를 측정하여 새로운 계층구조도 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 연구와 관련된 지식을 기술하고 3장에서는 뉴스 비디오를 파

싱하여 샷과 씬으로 검출 하는 방법을 기술하고 4장에서는 온톨로지를 이용하여 새로운 상위 계층을 구축하는 방법을 제안한다. 마지막으로 5장에서 결론을 맺는다.

2. 배경지식 및 관련 연구

2.1 샷 영역 검출

일반적으로 사용되는 샷 영역 검출 기법은 크게 5가지의 분류가 있는데 각각 픽셀기반 검출법, 통계기반 검출법, 변환기반 검출법, 특징기반 검출법, 히스토그램기반 검출법이다. 이들 기법들을 단독적으로 사용하거나 혹은 복합적으로 사용하여 샷 영역을 검출 할 수 있다. 픽셀기반 검출법[3]은 픽셀의 차이값(PDV, Pixel Difference Value)을 이용한 방법으로 연속되어진 비디오에서 특정한 위치에 존재하는 픽셀의 색의 차이값으로 검출하는 방법이다 이 방법은 알고리즘이 단순하고 속도가 빠른 반면에 빛에 민감하다는 단점이 있다. 통계기반 검출법은 픽셀기반 검출법의 단점을 보완하기 위해 제안된 것으로 명암에 대한 통계값을 이용하여 검출하는 방법이다. 변환기반 검출법은 MPEG등과 같이 압축된 영상에서 압축을 해제하지 않고 압축된 상태에서 움직임 벡터(motion vector)를 이용하여 샷을 검출하는 방법으로 실시간으로 검출이 가능한 것이 가장 큰 장점이다. 특징기반 검출법은 비디오의 엷지의 특성을 이용하여 검출하는 방법으로 비디오 내의 엷지를 비교하여 그 차이를 이용하여 샷을 검출하는 방법이다. 이 검출 방법은 가장 정확하게 샷을 검출 할 수 있다는 장점이 있지만 엷지에 대한 계산량이 많아 속도가 지나치게 느리다는 단점이 있다. 마지막으로 히스토그램 기반 검출법은 HSV(Hue, Saturation, Value) 히스토그램을 이용하여 샷 영역을 검출하는 방법이다. 가장 많이 사용되며 시간과 성능비가 가장 좋은 검출 기법이다.

2.2 키 프레임 추출

샷 영역 검출기법이 끝난 후에 하는 것으로 하나의 샷 영역에서 대표되는 하나 혹은 복수개의 프레임은 추출하는 과정이다. 단순한 추출기법으로는 하나의 샷 영역의 첫 프레임과 마지막 프레임은 추출하는 기법이 있으며 그 밖에 시각적인 화면의 복잡도에 의한 추출 방법, 샷의 활성도를 이용하는 방법, 샷의 모션을 이용하는 방법이 있다.

2.3 씬 구축

씬 구축은 중간단계인 그룹을 이용하여 구축하는 방법과 바로 씬으로 구축하는 방법이 있다. 그룹을 이용하여 구축하는 방법을 최근 연구에서 많이 사용하는데 이는 그룹 계층 없이 바로 씬을 구축할 경우에 “윈도우 효과(Window Effect)”같은 현상이 발생되어 정확도가 떨어지기 때문이다[2]. 탐색트리를 이용한 방법과 시각적인 내용에 의한 방법, 의미적인 내용에 의한 방법이 존재 한다. 탐색트리를 이용한 방법은 비디오를 복수개의 부분장면으로 나누는 기법으로 동일한 샷의 개수로 나누는 방법이다. 시각적인 내용에 의한 방법은 샷 영역이 가지는 시각적인 화면의 유사도(물리적)를 측정하여 묶는 방법이 가장 많이 사용된다. 의미적인 내용에 의한 방법은 화면이 가지고 있는 의미 정보를 가지고 의미의 유사도(논리적)를 측정하여 묶는 방법으로 최근에 많은 연구가 진행 되고 있다.

3. 뉴스 비디오 파싱

본 장에서는 본 연구에서 사용한 뉴스 파싱 기법에 관해서 기술한다. 3.1절과 3.2절, 3.3절은 기존에 연구된 알고리즘을 수정하여 사용한다[2].

3.1 샷 영역 검출 및 키 프레임 추출법

샷 영역 검출은 기존의 많은 연구에서 사용하는 히스토그램 기반의 검출법을 사용한다. 이 방법은 현재 프레임과 이전 프레임의 히스토그램 간의 변화량을 측정하여 그 변화량이 큰 지점을 장면이 바뀌는 지점으로 찾는 방법이다[2]. 키 프레임의 추출법은 단순히 시작 프레임만을 선택하는 방법을 이용한다.

3.2 씬 구축을 위한 특징 정보 추출

샷 i 의 활동성은 기존 연구에서 제안된 방법을 사용하여 다음과 같이 계산한다[2].

$$Act_i = \frac{1}{N_i - 1} \sum_{k=1}^{N_i - 1} Diff(k, k-1)$$

$$Diff(k, k-1) = Dist(Hist(k), Hist(k-1))$$

Act_i 는 샷 i 의 활동성(Activity)값이고 N_i 는 샷 i 의 총 프레임의 개수다. $Diff_{k,k-1}$ 는 프레임 k 와 $k-1$ 사이의 히스토그램의 값의 차이이고 $Dist()$ 는 히스토그램들 간의 유사도 함수이다. 본 논문에서는 교차거리(Intersection Distance)를 사용한다. 히스토그램은 H 의 정규화를 거친 100개의 벡터 값과 S 의 100개의 벡터 값을 사용한다. V 값은 사용하지 않는데 이는 V 가 빛에 민감하기 때문에 결과에 좋지 않은 영향을 주기 때문이다[2].

키 프레임은 가장 단순한 방법인 첫 번째 프레임만 사용한다.

$$Hist(b_i)$$

b_i 는 샷 i 의 시작 프레임이다.

본 논문에서는 특징 정보로 뉴스 비디오의 자막(Caption)정보도 같이 사용하기 때문에 자막 정보도 추출 한다.

$$Cap_i = CaptionofFrame_{b_i}$$

Cap_i 는 샷 i 의 자막 정보이고 샷 i 의 시작프레임에 자막이 있을 경우에만 추출한다. 자막의 추출 방법은 오티지 기반의 OCR을 이용한다[4].

샷은 다음과 같이 정의 된다.

$$shot_i = shot_i(b_i, Act_i, Hist(b_i), Cap_i)$$

3.3 씬 구축

뉴스 비디오에서의 씬구축은 앵커샷을 찾는 것을 의미한다. 기존의 비디오는 윈도우 효과(Window Effect)가 존재하기 때문에 이 현상을 막기 위해서 그룹을 구축한 후에 씬을 구축 하지만 뉴스 비디오는 이와 같은 윈도우 효과가 없기 때문에 한

번에 바로 씬으로 만들어 낼 수 있다. 이는 뉴스 비디오가 가지는 구조적인 특징 때문이다[1]. 뉴스 비디오의 샷을 씬으로 구축하기 위해서 다음과 같이 유사도를 계산하여 구축한다. 두 샷을 지칭하는 i, j 는 항상 j 가 보다 크다.

1. 샷의 화면 유사도인 ShotColorSim를 구한다. 유사도를 구할 때는 전체 화면의 왼쪽 1/3영역의 유사도만 구한다. 왼쪽 1/3 영역만 사용하는 이유는 계산량을 줄일 수 있고 뉴스 화면의 오른쪽의 항상 바뀌는 뉴스 상자에 의해 좋지 않은 결과가 나오는 현상을 방지 할 수 있기 때문이다.

(a) 2개의 프레임간의 유사도를 계산한다.

$$FrameColorSim_{b_i, b_j}, FrameColorSim_{e_i, e_j}$$

$FrameColorSim_{x,y}$ 는 다음과 같이 정의 된다.

$$FrameColorSim_{x,y} = 1 - Diff_{x,y}$$

x, y 는 각각 샷의 프레임이고 $x < y$ 이다.

(b) 샷 i 와 샷 j 의 유사도를 정의 하기위해 프레임간의 비교값 중에서 가장 큰값을 취한다.

$$ShotColorSim_{i,j} =$$

$$\max(FrameColorSim_{b_i, b_j}, FrameColorSim_{e_i, e_j})$$

2. 샷의 활동성의 유사도인 ShotActSim를 구한다.

$$ShotActSim_{i,j} = \text{avr}(Act_i, Act_j)$$

3. 두 값을 이용하여 최종적인 유사도를 구한다.

$$ShotSim_{i,j} = ShotColorSim_{i,j} + ShotActSim_{i,j}$$

위의 유사도를 가지고 유사한 샷들을 가지고 씬을 구축하게 되면 특정 샷들이 하나의 씬으로 모이게 되는데 이것이 바로 앵커 샷의 집합이다. 왜냐하면 하나의 뉴스 비디오에서는 유사도에 따른 샷의 집합에서 가장 많은 개체수를 갖는 것이 바로 앵커 샷이기 때문이다. 이렇게 모여진 앵커 샷들의 시작부분을 바로 하나의 씬의 시작부분으로 선택하게 되면 씬 구축이 끝난다.

3.4 키워드 집합 구성

3.2절에서는 씬 구축을 위해서 추출한 특징정보 중에서 자막 정보를 이용하여 키워드 집합을 구성하는 방법을 제안한다. 자막정보를 온톨로지 추론을 위한 키워드(keyword)로 변환을 하게 되는데 이는 다음의 순서로 진행된다.

1. 띄어쓰기를 기준으로 키워드로 분해한다.

2. 숫자가 있는 키워드는 삭제한다.

3. 형태소 분석을 하여 명사만 추출한다. 합성 명사나 고유명사는 삭제하지 않는다.

4. 단일 글자는 삭제한다.

5. 하나의 씬에 복수개의 샷이 존재할 경우에는 각각의 샷에 대해서 분리된 키워드로 하나의 키워드 집합을 만든다.

Scene _{i} 에 대해서 Shot _{i} , Shot _{j} , Shot _{k} 가 존재 하고 각각의 샷은 다음과 같은 키워드 집합을 가진다.

$$Shot_i \in \{A, B, C, D\}$$

$$Shot_j \in \{B, C, E\}$$

$$Shot_k \in \{E, F\}$$

위의 순서에 따라 만들어진 씬의 키워드셋 Scene _{i} 는 다음과 같다. 단, {B, C, E}는 명사, {A}는 단일글자, {D, F}는 형용사를 뜻한다.

$$Scene_i \in \{B, C, E\}$$

4. 온톨로지를 이용한 자동 분류

본 연구에서는 2개의 온톨로지를 가지고 추론을 수행한다. 키워드에 대한 온톨로지인 어휘 온톨로지와 뉴스의 의미를 가지는 '뉴스 온톨로지를 이용한다. 어휘 온톨로지와 뉴스 온톨로지는 규칙(rule)에 의해서 관계가 기술 되어 있다.

4.1 어휘 온톨로지

본 연구에서 사용하는 어휘 온톨로지는 [5]의 연구에서 제시된 어휘 온톨로지를 사용한다. 어휘 온톨로지는 그 크기가 방대하고 어휘에 대한 주관적인 입장이 개입되어있기 때문에 많은 연구자들에 의해서 객관적으로 만들어진 온톨로지를 이용하는 것이 바람직하다.

어휘 온톨로지는 [그림 2]와 같이 한국어휘에 대해서 어휘가 가지는 계층적 위치(Category)를 기준으로 어휘를 분류해 놓은 것이다.

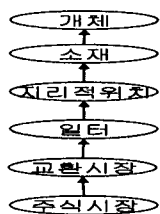


그림 2 어휘 온톨로지 중 "주식시장"

4.2 뉴스 온톨로지

본 연구에서는 [그림 3]와 같은 뉴스 온톨로지를 구축하여 사용한다.

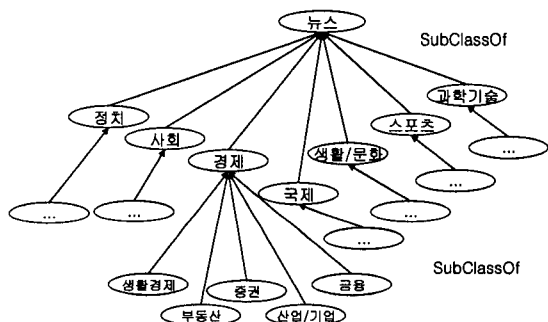


그림 3 뉴스 온톨로지 중 경제 분야의 예

위와 같이 뉴스 온톨로지의 상위 개념과 중간개념간의 관계는 sub Class of 관계이고 [그림 4]와 같이 뉴스 온톨로지와의 어휘 온톨로지의 관계는 참조 관계이다.

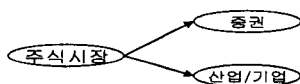


그림 4 어휘 온톨로지와 뉴스 온톨로지의 관계 "주식시장"

4.3 온톨로지 추론

본 논문에서는 씬의 키워드 집합으로 어휘 온톨로지를 매칭하는 방법을 이용한다. 어휘 온톨로지의 어휘들은 뉴스 온톨로지의 분야별로 참조 관계를 가지고 있다. 예를 들어 "주식시장"이라는 어휘는 [그림 6]과 같이 뉴스 온톨로지에서의 "경제"의 하위 개념인 "증권" 및 "산업/기업"과 참조 관계를 가지고 있다. 즉 "주식시장"이라는 키워드는 뉴스 온톨로지의 "경제"에 매칭된다.

참조 관계에 의해서 매칭되어지는 분야가 키워드 집합의 65% 이상일 경우엔 해당 분야가 키워드 집합의 분야이다. 즉 키워드 집합 $Scene_i \in \{A, B, C\}$ 에 대해서 $A \Rightarrow$ "증권" \rightarrow "경제", $A \Rightarrow$ "산업/기업" \rightarrow "경제", $B \Rightarrow$ "증권" \rightarrow "경제", $B \Rightarrow$ "증권" \rightarrow "경제"의 관계면 키워드 집합 $Scene_i$ 는 "증권" \rightarrow "경제"로 분류된다.

참조 관계에 의해서 매칭되어지는 분야가 키워드 집합의 65%미만인면서 분야의 상위 개념으로 65%이상일 경우엔 상위

개념이 키워드 집합의 분야이다. 즉 키워드 집합 $Scene_i \in \{A, B, C\}$ 에 대해서 $A \Rightarrow$ "증권" \rightarrow "경제", $B \Rightarrow$ "산업/기업" \rightarrow "경제", $B \Rightarrow$ "금융" \rightarrow "경제", $C \Rightarrow$ "부동산" \rightarrow "경제"의 관계이면 키워드 집합 $Scene_j$ 는 "경제"로 분류된다. 그 밖에 경우엔 가장 많은 분야의 상위 개념 2개가 분야가 된다. 예를 들어, 키워드 집합 $Scene_k \in \{A, B, C\}$ 에 대해서 $A \Rightarrow$ "증권" \rightarrow "경제", $B \Rightarrow$ "금융" \rightarrow "경제", $B \Rightarrow$ "국회" \rightarrow "정치", $C \Rightarrow$ "증권" \rightarrow "경제", $C \Rightarrow$ "정보기술" \rightarrow "기술/과학"의 관계이면 키워드 집합 $Scene_k$ 는 "정치"와 "경제"로 분류된다.

4.4 그룹 구축

3.5절에서 추론된 씬의 의미적 분야가 동일한 씬을 모아서 하나의 그룹으로 구축한다. 하나의 뉴스 비디오에는 "정치, 사회, 경제, 국제, 생활/문화, 스포츠, 과학기술"의 총 7개의 그룹이 존재 하게 된다.

5. 결론

본 논문에서는 온톨로지를 이용하여 뉴스 비디오의 의미적 분류 방법을 제안했다. 뉴스 비디오를 의미적으로 분류하기 위하여 뉴스 비디오에서 특징 정보로 뉴스가 가지는 자막정보를 사용 하였고 해당 특징 정보를 OCR을 이용하여 문자열로 변환 하였다. 이렇게 변환된 문자열을 키워드로 가공하여 어휘 온톨로지와 매칭을 하여 뉴스 비디오의 씬에 대해서 의미적 분류를 하였다. 또한 뉴스 비디오의 경우 대부분의 표준 어휘를 사용하기 때문에 표준 어휘에 대한 어휘 온톨로지를 구축하였다. 향후 연구에서는 OCR로 추출한 특징정보와 화면가지는 의미적 요소를 추출하여 두 가지 모두를 이용한 뉴스 자동 분류 기법을 연구할 예정이다.

참고문헌

[1] Seung-Chul Jun and Sung Han Park, "An Automatic News Video Semantic Parsing Algorithm," ITC-CSCC2001, 2001.7
 [2] Yong Rui, Thomas S. Huang, and Sharad Mehrotra, "Constructing Table-of-Content for Videos", ACM Multimedia Systems, 7(5), 1999
 [3] HongJiang Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video. ACM Multimedia Systems, 1(1), 1993
 [4] Seung-Chul Jun, Myeong-Gyu Kang and Sung-Han Park, "A Hangeul Recognition Method Using Directional Edges in Open Captions," ITC-CSCC 2002, July, 2002
 [5] www.wordreference.com