

온라인 동향 분석을 위한 이벤트 문장 추출 방안

Event Sentence Extraction for Online Trend Analysis

윤보현
목원대학교

Bo-Hyun Yun(ybh@mokwon.ac.kr)

요약

기존의 이벤트 문장 추출에 관한 연구는 학습단계에서 3W 자질을 학습하지 않고, 추출단계에서 3W 자질의 존재여부에 따른 규칙만을 적용하여 이벤트 문장을 추출하였다. 본 논문에서는 온라인 동향 분석을 위해 학습단계에서 3W 자질을 추출하고 가중치를 계산하고, 추출단계에서 3W 자질을 반영하는 문장 가중치 기반 이벤트 문장 추출 방안을 제시한다. 실험결과, 자질필터링은 $TF \times IDF$ 가중치 기법을 사용한 상위 30% 자질만을 사용하는 것이 가장 우수한 결과를 보였다. 공공이슈 분야인 부동산 도메인에서 문장 가중치 기반 방법은 3W 자질 중 who와 when 자질이 가장 영향을 많이 미치는 것으로 나타났다. 아울러 다른 기계학습 방법과의 비교하여 공공이슈 분야인 부동산 도메인에서 문장 가중치 기반 이벤트 문장 추출 방법이 가장 좋은 성능을 보였다.

■ 중심어 : | 이벤트 문장 추출 | 온라인 동향 분석 | 자질추출 |

Abstract

A conventional event sentence extraction research doesn't learn the 3W features in the learning step and applies the rule on whether the 3W feature exists in the extraction step. This paper presents a sentence weight based event sentence extraction method that calculates the weight of the 3W features in the learning step and applies the weight of the 3W features in the extraction step. In the experimental result, we show that top 30% features by the $TF \times IDF$ weighting method is good in the feature filtering. In the real estate domain of the public issue, the performance of sentence weight based event sentence extraction method is improved by who and when of 3W features. Moreover, In the real estate domain of the public issue, the sentence weight based event sentence extraction method is better than the other machine learning based extraction method.

■ keyword : | Event Sentence Extraction | Online Trend Analysis | Feature Extraction |

I. 서론

웹 및 소셜 미디어 상의 다양한 이슈에 대해 정보의

왜곡이 없이 글로벌 온라인 동향의 실시간 분석 및 미래 예측의 핵심 기술 개발 필요하다. 글로벌 경쟁 시대에서 온라인 동향 분석 및 예측 기술은 기존 전문가 집

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원금을 받아 수행된 것임.
(No. 한국연구재단에서 부여한 과제번호 : 2010-0022201)

단에 의한 고비용, 저효율 수작업 위주 작업을 대체가 능한 비용 대비 가장 효과적이고 신속한 결과를 얻을 수 있는 장점이 있다.

본 논문의 목적은 웹상에 존재하는 많은 부동산 관련 뉴스기사로부터 특정 시기의 부동산 동향을 파악할 수 있는 이벤트 문장을 추출하는 것이다. 이벤트 문장이란 특정 주제에 대한 구체적인 내용 즉, 누가, 어디서, 언제, 무엇이, 어떠한가 등이 표현된 문장을 의미한다. 예를 들어, 부동산 관련 뉴스 기사에서 이벤트 문장의 예는 “2일 부동산114에 따르면 지난해 수도권 전체 소형 아파트 가격은 2010년 말보다 1.04% 올랐다.”이다. 본 논문에서는 이벤트 문장 추출을 위해 각 문장의 가중치를 계산하여 특정 임계치 값 이상인 문장만을 추출하고자 한다.

문장 추출에 관한 기존의 연구는 지식베이스 확장을 위한 핵심문장추출[14], 적응기법을 이용한 문장추출[11], 통계적 기법을 이용한 문장추출[1][2][7][9][10][13], TextRank 알고리즘을 이용한 문장추출[12][15]로 나누 질 수 있다.

지식베이스 확장을 위한 핵심문장추출 방법[14]은 Wikipedia 문서에서 본문을 추출하고, 품사 태깅 규칙에 의해 전문용어를 추출하고, 전문용어를 기준으로 의미 가중치가 높은 핵심 문장을 추출하였다. 적응기법을 이용한 문장추출 방법[11]은 대상 문서 집합에서 순차적으로 문서를 입력받아 중요 문장을 추출, 이전에 구축된 요약문장과 현재 추출된 문장을 비교하면서 요약에 필요한 문장을 선택하는 적응기법을 활용하여 중요 문장을 추출한다.

통계적 기법을 이용한 문장추출 방법은 문서의 타이틀, 문장의 위치, 중요 단어, 연관된 문장 등의 정보를 통해 문장과 단어의 가중치를 결정하여 문장의 가중치를 계산하고 특정 임계치 이상의 문장을 추출한다. TextRank 알고리즘을 이용한 문장추출 방법은 PageRank 개념을 문장추출에 적용한 방법이다. PageRank는 기본적으로 중요한 사이트는 다른 많은 사이트로부터 링크를 받는다는 점에 착안하여 문서 A에서 가지고 있는 B의 링크가 B 사이트를 추천하는 한 표로 해석하여 이를 기준으로 중요도를 평가한다.

TextRank는 PageRank에서의 사이트를 단어 혹은 문장으로 생각하고 문장의 중요도를 계산한다.

기존의 이벤트 문장 추출에 관한 연구[5][8]는 학습단계에서 3W 자질을 학습하지 않고, 추출단계에서 3W 자질의 존재여부에 따른 규칙만을 적용하여 이벤트 문장을 추출하였다. 본 논문에서는 학습단계에서 3W 자질을 추출하고 가중치를 계산하고 추출단계에서 3W 자질의 가중치를 반영하고자 한다.

본 논문에서는 공공이슈 도메인의 온라인 동향 분석을 위해 문장 가중치 기반 이벤트 문장 추출 방안을 제시한다. 학습단계에서 형태소분석과 개체명인식의 결과인 자질들을 추출하고 가중치를 계산한다. 추출단계에서는 학습단계에서 계산된 자질 가중치를 이용하여 새롭게 입력된 문장의 가중치를 계산하고 특정 임계치 이상의 문장만을 추출한다. 실험에서는 최적의 자질의 개수 찾는 실험, 최적의 자질 조합을 찾는 실험, 개체명 자질의 효과 실험, 다른 기계학습 방법과의 비교를 수행하였다. 본 논문에서 2장은 자질 유형에 대해서 기술하고, 3장에서 문장 가중치 기반 이벤트 문장 추출을 설명한다. 그리고 4장에서는 실험 및 평가와 5장에서 결론을 기술한다.

II. 자질 유형

본 논문에서는 네 가지 자질유형을 사용한다. 첫째, 동사 자질은 형태소 분석 결과에서 ‘PV(동사)’ 및 ‘NC(보통명사)+XSV(동사파생접미사)’로 태깅된 것들을 추출한다. 예를 들어, 상승하다는 ‘상승/NC+하/XSV+였/EP(선어말어미)+다/EF(종결어미)’로 분석되어 동사 자질로 추출된다.

둘째, 명사 자질은 어휘 정보 이용하기 위해 ‘NC’와 ‘NC+XSN(명사파생접미사)’로 ‘아파트/NC’라는 명사 등을 어휘로 추출한다. 또한 품사 정보를 이용하여 ‘NN(수사)+NN’와 ‘NN+NB(의존명사)’로 ‘1000NN+만원/NB’라는 자질을 추출하였다.

셋째, 명사구 자질은 같은 문장 내에서 인접해서 나타나는 명사 자질들을 bi-gram으로 추출한 것을 의미

한다. 예를 들어 ‘아파트 가격’, ‘단독 주택’ 등을 명사구 자질로 추출하였다.

3W 자질은 개체명 인식결과에서 Who, When, Where 자질을 추출하는 것이다. Who 자질은 개체명 태그 정보 중 ‘PERSON’과 ‘ORGANIZATION’의 하위 목록들을 모두 포함하는 것이다. 예를 들어 ‘부동산 114/ORGANIZATION’ 등이 Who 자질이다.

When 자질은 개체명 태그 정보 중 DATE, TIME의 하위 목록들을 모두 포함하는 것이다. ‘DATE’의 하위 자질은 DT_OTHERS, DT_DURATION, DT_DAY, DT_MONTH, DT_YEAR 등이 있다. ‘TIME’의 하위 자질은 TL_OTHERS, TL_DURATION, TL_HOUR, TL_MINUTE, TL_SECOND이 있다. 예를 들어, ‘오후 /TL_OTHERS’와 ‘15일/DT_DAY’이 When 자질이다.

Where 자질은 개체명 태그 정보 중 LOCATION의 하위 목록들을 모두 포함하는 것이다. ‘LOCATION’의 하위 자질은 LC_COUNTRY, LC_COUNTY, LC_CAPITALCITY, LC_RIVER, LC_MOUNTAIN, LC_SPACE 등이 있다. 예를 들어, ‘경북 /LC_PROVINCE’와 ‘필리핀 LC_COUNTRY’ 등이 Where 자질이다.

Where 자질에서 개체명 태그 정보 중 장소 자질에 적합한 목록들 포함한다. ‘ARTIFACTS’ 태그이다. ‘ARTIFACTS’는 AF_BUILDING, AF_ROAD를 포함하며, 예를 들어 ‘황금동 신천지타운’은 <LCP_COUNTY :황금/NC+동/NC> <AF_BUILDING:신/XP(접두사)+천지/NC+타운/NC>로 개체명인식되며 신천지타운도 장소 자질로 추출된다.

III. 문장 가중치 기반 이벤트 문장 추출

3.1 시스템 구성도

이벤트 문장 추출 시스템 구성도는 [그림 1]과 같다. 학습 단계에서는 공공이슈 분야인 부동산 관련 뉴스 기사를 형태소분석과 개체명인식을 통해 언어처리를 수행한다. 개체명 인식된 결과에서 동사자질, 명사자질, 명사구자질, 3W 자질을 추출한다. 추출된 자질에 대해

서 $TF \times IDF$ 가중치를 적용하여 가중치를 계산한다.

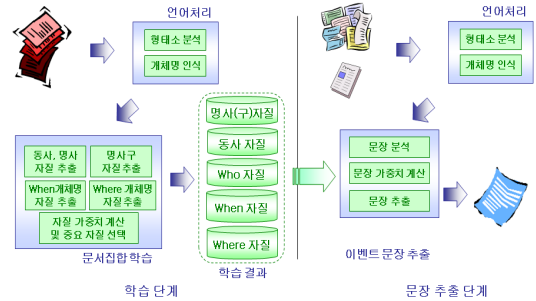


그림 1. 이벤트 문장 추출 시스템 구성도

이벤트 문장추출 단계에서는 새로운 문서가 입력되면 학습단계와 같이 언어처리를 수행한다. 문장분석을 통해 학습단계의 자질 정보를 이용하여 문장 가중치를 계산한다. 계산된 문장 가중치를 바탕으로 특정 임계치 이상이면 이벤트로 추출한다.

3.2 문장 가중치 계산 방법

본 논문에서 사용하는 문장 가중치 계산방법은 [8]에서 제안된 식(3)에 3W 자질을 반영한 식(4)에 의해 계산한다. 식(1)과 (2)는 문장에 포함되어 있는 명사 자질과 명사구 자질을 문장 가중치 계산에 반영하기 위해, i 번째 문장에서 동사- j 와 함께 나타나는 명사 자질들에 대한 가중치 합의 평균을 나타내는 $Cov_{i,j}$ 값과 i 번째 문장에서 동사- j 와 함께 나타나는 명사구 자질들에 대한 가중치 합의 평균을 나타내는 $Cov_{i,j}$ 값을 의미한다.

$$Cov_{i,j} = \frac{\sum_{k=1}^{C_{i,noun}} (W_{n_k} \times Co_{v_j,n_k})}{C_{i,noun}} \quad (1)$$

$$Cov_{i,j} = \frac{\sum_{k=1}^{C_{i,np}} (W_{np_i} \times Co_{v_j,np_i})}{C_{i,np}} \quad (2)$$

$$W_i = \sum_{j=1}^{C_{i,verb}} (W_{v_j} \times (\alpha \cdot Cov_{i,j} + \beta \cdot Cov_{i,j})) / C_{i,verb} \quad (3)$$

위의 $C_{i,noun}$, $C_{i,np}$, $C_{i,verb}$ 수식들에서는 각각 문장-i 내에 출현한 명사, 명사구, 동사 자질의 수를 나타내고, W_{n_k} , W_{np_i} , W_{v_j} 는 학습의 결과로 얻은 각 자질의 가중치를 나타낸다. 또한 Co_{v_j, n_k} , Co_{v_j, np_i} 는 각각 동사-j와 명사-k의 공기빈도, 동사-j와 명사구-i의 공기빈도를 나타내고, α 와 β 는 명사와 명사구 자질이 문장추출에 기여하는 정도에 따라 조정되는 상수 값이다. 각 문장의 가중치가 계산되면 이를 이용해 단일 문서 내에서 문장들을 정렬해 문장추출 시에 조건항목으로 사용한다.

$$W_i = \sum_{j=1}^{C_{i,verb}} (W_{v_j} \times (\alpha \cdot Covn_{i,j} + \beta \cdot Covp_{i,j}) + \gamma(W_{who} + W_{when} + W_{where})) / C_{i,verb} \quad (4)$$

식(4)에서 γ 는 3W 자질이 문장추출에 기여하는 정도에 따라 조정되는 상수 값이다. W_{who} , W_{when} , W_{where} 는 각 개체명 자질의 가중치를 의미한다. [그림 2]는 제안한 방법으로 문장의 가중치를 계산한 결과를 보여주고 있다. 예를 들어, “2일 부동산114에 따르면 지난해 수도권 전체 소형 아파트 가격은 2010년 말보다 1.04% 올랐다.”는 1.570497의 높은 가중치가 부여되었음을 알 수 있다.

IV. 실험 및 평가

4.1 실험자료 및 인자

본 논문에서 사용하는 실험 자료는 [표 1]과 같다. 자동차 사고 도메인에서 문서 120개를 수집하여 수동 분류한 결과, 792개의 문장중에서 이벤트인 문장은 289개이고, 이벤트가 아닌 문장은 513개였다. 부동산 도메인에서 문서 120개를 수집하여 분류한 결과, 1,209개의 문장중에서 이벤트인 문장은 233개이고, 이벤트가 아닌 문장은 987개였다.

부동산 경기 침체로 수도권 아파트 가격 하락세가 장기화되고 있지만 60㎡전용면적 기준) 이하 소형 아파트는 되레 오른 것으로 나타났다. [0.139658]
2일 부동산114에 따르면 지난해 수도권 전체 소형 아파트 가격은 2010년 말보다 1.04% 올랐다. [1.570497]
같은 기간 60㎡초과~85㎡이하의 중소형 아파트와 85㎡초과 규모의 중대형 아파트는 0.40%와 1.47%가 각각 떨어졌다. [0.616911]
60㎡이하 아파트의 가격 상승세는 경기지역이 주도했다. [0.038694]
경기지역에서 1년 동안 2.63%가 오르면서 서울(0.01%)과 인천(1.40%)에서 떨어진 것을 상쇄하고, 수도권 전체 소형 아파트값을 상승세로 바꿔놨다. [0.274619]

그림 2. 문장가중치 계산의 예제

표 1. 실험 데이터

도메인	크기	문장수	이벤트	이벤트 X
자동차 사고	90.5Kb	792	289	513
부동산	127.7Kb	1,209	233	987

분류 시스템의 평가인자는 식 (5)에서 식(8)까지 네 가지를 사용하였다. [표 2]는 이진분류에 대한 Contingency table이다. 이를 바탕으로 식 (5)에서 식 (8)까지 평가인자가 유도된다. 정확률(Precision)은 시스템이 이벤트라고 분류한 총 개수로 시스템이 정확하게 분류한 이벤트 문장의 개수를 나눈 것이다. 재현율(Recall)은 사람이 이벤트라고 분류한 문장의 개수로 시스템이 정확하게 분류한 이벤트 문장의 개수를 나눈 것이다. 분류 정확도(Classification Accuracy)는 사람이 이벤트와 이벤트가 아니라고 분류한 총 개수로 시스템이 정확하게 분류한 이벤트와 이벤트가 아니라고 분류한 개수를 나눈 것이다. 일반적으로 분류시스템은 결합된 평가 인자인 F-measure와 분류 정확도로 성능을 평가하였다[6].

표 2. 이진분류의 Contingency Table

		시스템	
		Pos	Neg
사람	Pos	True Positive	False Negative
	Neg	False Positive	True Negative

$$\text{정확률} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{재현율} = \frac{TP}{TP+FN} \quad (6)$$

$$F\text{-measure} = \frac{2 \times \text{정확률} \times \text{재현율}}{\text{정확률} + \text{재현율}} \quad (7)$$

$$\text{분류정확도} = \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

자질 개수에 따른 분류는 [표 3]에서 보이고 있다. Information Gain과 $TF \times IDF$ 방법을 자질을 랭킹하여 상위 15%부터 100%까지 분류하였다. 실험한 결과는 [그림 3]과 [그림 4]에서 보이고 있다. 자동차 사고 도메인과 부동산 도메인 모두에서 상위 30% 자질만을 사용하는 것이 최적의 결과를 보이고 있다. 또한 자질 필터링 방법으로 $TF \times IDF$ 방법이 Information Gain 보다 나은 결과를 보이고 있음을 알 수 있다. 따라서 본 논문에서 자질 필터링을 위해 $TF \times IDF$ 방법을 사용하여 상위 30% 자질만은 이용하였다.

표 3. 자질 개수에 따른 분류표

실험	자질	실험 문서
①	상위 자질 15%	테스트 문서(40건)
②	상위 자질 30%	테스트 문서(40건)
③	상위 자질 50%	테스트 문서(40건)
④	상위 자질 70%	테스트 문서(40건)
⑤	상위 자질 100%	테스트 문서(40건)

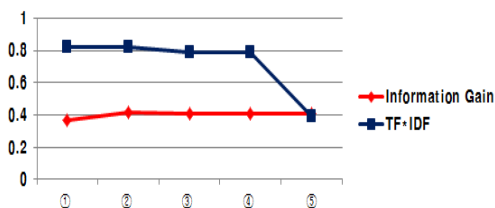


그림 3. 자동차사고 도메인에서 자질 개수별 성능 비교

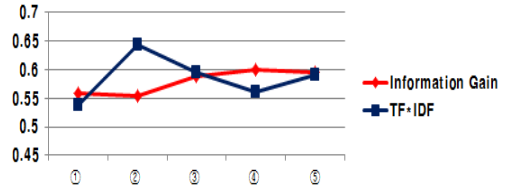


그림 4. 부동산 도메인에서 자질 개수별 성능 비교

[표 4]에서 문장 가중치 수식의 조합을 보이고 있다. ①번은 식 (3)을 의미한다. ②는 식 (4)에서 W_{who} 를 제외한 식이다. ③은 식 (4)에서 W_{when} 를 제외한 식이다. ④는 식 (4)에서 W_{where} 를 제외한 식이다. ⑤은 식 (4)를 의미한다.

표 4. 문장 가중치 수식 조합

번호	문장 가중치 수식
①	식 (3)
②	식 (4)에서 W_{who} 를 제외한 식
③	식 (4)에서 W_{when} 를 제외한 식
④	식 (4)에서 W_{where} 를 제외한 식
⑤	식 (4)

표 5. 문장 가중치 수식 조합 실험 결과

조 건	평가	자동차사고	부동산
		테스트	테스트
①	F-measure	0.529801	0.506849
	분류정확도	0.736059	0.804878
②	F-measure	0.795455	0.524064
	분류정확도	0.866171	0.758807
③	F-measure	0.632911	0.582278
	분류정확도	0.784386	0.821138
④	F-measure	0.658228	0.821474
	분류정확도	0.799256	0.92398
⑤	F-measure	0.666667	0.589744
	분류정확도	0.802973	0.826558

[표 5]에서 자동차 사고 도메인에서는 ②번 즉, 식 (4)에서 W_{who} 를 제외한 식이 가장 좋은 결과를 보이고 있다. 다시 말하면, 자동차 사고 도메인에서는 3W 자질 중에서 when과 where 자질이 이벤트 문장 추출에 영향을 미친다는 것을 의미한다. 부동산 도메인에서 ④번 즉, 식 (4)에서 W_{where} 를 제외한 식이 가장 우수한 결과를 보이고 있다. 이것은 공공이슈 분야인 부동산 도

메인에서는 who와 when 자질이 많은 영향을 미치는 것임을 의미한다. where 자질의 경우, 많은 문장에서 언급이 되기 때문에 문장간 변별력에 좋은 영향을 미치지 못하는 것으로 나타났다.

[표 6]에서 문장 가중치 방법과 다른 기계학습 방법들과 비교한 결과를 보이고 있다. SVM(Support Vector Machine), BN(Bayesian Network), DT(Decision Tree) 알고리즘을 비교실험하기 위해 데이터마닝 웨카(WEKA)[16]를 사용하여 실험하였다. 웨카는 Java 언어로 개발된 데이터마닝 프로그램이다. 본 논문에서 사용하는 SVM 방법의 커널함수는 최적의 결과를 보이는 linear 커널함수이다.

표 6. 다른 기계학습 알고리즘과 비교

알고리즘	평가척도	자동차 사고	부동산
		테스트	테스트
SW	정확률	0.882	0.765
	재현율	0.824	0.889
	F-measure	0.852	0.821
	분류정확도	0.896	0.923
SVM	정확률	0.947	0.736
	재현율	0.899	0.636
	F-measure	0.922	0.678
	분류정확도	0.941	0.883
BN	정확률	0.881	0.681
	재현율	0.908	0.569
	F-measure	0.894	0.620
	분류정확도	0.920	0.866
DT	정확률	0.898	0.784
	재현율	0.840	0.639
	F-measure	0.865	0.703
	분류정확도	0.900	0.895

자동차 사고 도메인에서는 SVM 알고리즘[3][4]이 가장 나은 성능을 보이고 있다. 그러나 문장 가중치(Sentence Weighting, SW) 방법과는 F-measure와 분류정확도가 7%차이를 보이고 있어 크게 차이가 나지 않는다. 공공이슈 분야인 부동산 도메인에서는 문장 가중치 방법이 F-measure와 분류정확도 측면에서 가장 우수한 결과를 보이고 있다. SVM 알고리즘과는 F-measure 15%, BN 알고리즘과는 F-measure 20%, DT 알고리즘과는 F-measure 12% 차이를 보이고 있어 크게 차이가 나고 있다. 따라서 공공이슈 분야에서는 문장 가중치 기반 이벤트 문장추출 방법이 적합함을 알

수 있다.

VI. 결론

본 논문에서는 공공이슈 도메인의 온라인 동향 분석을 위해 문장 가중치 기반 이벤트 문장 추출 방안을 제시하였다. 실험결과, 자동차 사고 도메인과 부동산 도메인 모두에서 상위 30% 자질만을 사용하는 것이 최적의 결과를 보였다. 공공이슈 분야인 부동산 도메인에서는 who와 when 자질이 많은 영향을 미친다는 것을 알 수 있었다. 부동산 도메인에서 where 자질의 경우, 많은 문장에서 언급이 되기 때문에 문장간 변별력에 좋은 영향을 미치지 못하는 것으로 나타났다. 자동차 사고 도메인에서는 SVM 알고리즘이 가장 나은 성능을 보이고 있다. 결론적으로 공공이슈 분야인 부동산 도메인에는 문장 가중치 방법이 가장 나은 결과를 보였고, 사건사고 분야에는 기계학습 방법인 SVM 방법이 가장 나은 성능을 보였다.

참고 문헌

- [1] H. P. Edmundson, "New Methods in Automatic Extracting," Advances in Automatic Text Summarization, The MIT Press, pp.23-42, 1999.
- [2] J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell, "Summarizing Text Documents: Sentence Selection and Evaluation Metrics," In Proceedings of ACM-SIGIR'99, pp.121-128, 1999.
- [3] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.3096>, 2000.
- [4] T. Jachims, "Text categorization with support vector machines: Learning with many relevant features," In C. Ne'dellec & C. Rouveirol (Eds.),

Proceedings of the 10th European conference on machine learning, Springer, Heidelberg, DE, Chemnitz, DE., pp.137-142, 1998.

[5] M. Naughton, N. Stokes, and J. Carthy, "Sentence-level event classification in unstructured texts," Information Retrieval, Vol.13, No.2, pp.132-156, 2010.

[6] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, Vol.45, No.4, pp.427-437, 2009.

[7] 김동욱, 강수용, 김한준, 이병정, 장재영, "사용자 프로파일을 통한 개인화 문장 추출 시스템 설계", 한국인터넷정보학회 학술발표대회 논문집, pp.213-217, 2010.

[8] 김태현, 임수중, 윤보현, 박상규, "정보 추출을 위한 이벤트 문장 추출", 제14회 한글 및 한국어 정보처리 학술대회, 2002.

[9] 나종열, 신지애, 최기선, "VNA 집합을 이용한 뉴스기사의 중요문장 추출", 2007년도 제19회 한글 및 한국어 정보처리 학술대회, 한국정보과학회, 2007.

[10] 류제, 한광록, 손석원, 임기욱, "단어의 공기 관계 그래프를 이용한 문서의 핵심 문장 추출에 관한 연구", 정보처리학회지, 제7권, 제11호, pp.3427-3437, 2000.

[11] 임정민, 강인수, 배재학, 이종혁, "다중문서 요약에서 적응 기법을 이용한 문장 추출", 제16회 한글.언어.인지 학술대회, 2004.

[12] 이재걸, 박성배, 이상조, "2단계 문장 추출 방법을 이용한 회의록 요약", 2010년도 추계학술대회 학술발표논문집, 제20권, 제2호, 한국지능시스템학회, 2010.

[13] 정운철, 고영중, 서정연, "2단계 문장 추출방법을 이용한 자동 문서 요약", 한국정보과학회 04 봄 학술발표논문집(B), 한국정보과학회, 2004.

[14] 최준호, 김판구, "지식베이스 확장을 위한 핵심 문장 추출", 2010년도 춘계학술발표논문집, 한국

멀티미디어학회, 2010.

[15] 홍진표, 차정원, "TextRank 알고리즘을 이용한 한국어 중요 문장 추출", 2009 한국컴퓨터종합학술대회, 한국정보과학회, 2009.

[16] <http://www.cs.waikato.ac.nz/ml/weka/>

저 자 소 개

윤 보 현(Bo-Hyun Yun)

종신회원



- 1995년 : 고려대학교 컴퓨터학과 (이학석사)
 - 1999년 : 고려대학교 컴퓨터학과 (이학박사)
 - 1999년 ~ 2003년 : 한국전자통신연구원 팀장
 - 2003년 ~ 현재 : 목원대학교 컴퓨터교육과 교수
- <관심분야> : 자연어처리, 정보검색, Semantic Web