

중심 화자를 선택하는 방법에 대해서 추가적인 조사를 진행하였다. 현재 따로 화자의 중요도를 측정하는 방법이 없기 때문에 화자 별로 말하는 대화 내용의 중요도와 화자의 길이를 복합적으로 고려하여 중심 화자를 선택하기로 하였다.

지난 시간 교수님의 피드백으로 text-rank와 rex-rank기법은 tf-idf라는 단어 임베딩 기법을 사용하는데 이 방법은 너무 오래된 방법이므로 정확도에 문제가 있을 수 있기에 추가적인 방법에 대해 조사를 진행하였다. 현재 RVA기법과 동시 출현 행렬 등의 기법이 있는데, RVA 기법은 단어에 대해서 벡터를 랜덤하게 생성하는 방식이기 때문에 단어의 유사도를 고려해야 하는 rex rank에는 적합하지 않는 방법이었다. 동시 출현 행렬은 통계 기반의 방법으로 문장의 순서와 문맥 정보를 고려하지 않는다는 단점이 있었다.

최근 많이 사용되는 머신러닝 기반의 단어 임베딩 기법으로는 word2vec나 glove등의 방법이 있다. 이러한 임베딩 기법을 기존의 rex-rank에 적용시키면 기존의 rex-rank의 성능이 향상될 것으로 기대된다.

TABLE I. RESULTS COMPARISON OF EXTRACTING THREE KEYWORDS

Algorithm	Result			
	<i>t</i>	<i>P</i>	<i>R</i>	<i>F</i>
TextRank	—	0.27075	0.31340	0.29052
Weighted TextRank	0.3	0.31735	0.37000	0.34166
	0.5	0.33027	0.38728	0.35651
	0.7	0.33299	0.39024	0.35935

TABLE II. RESULTS COMPARISON OF EXTRACTING FOUR KEYWORDS

Algorithm	Result			
	<i>t</i>	<i>P</i>	<i>R</i>	<i>F</i>
TextRank	—	0.25442	0.39068	0.30816
Weighted TextRank	0.3	0.30238	0.46510	0.36649
	0.5	0.30748	0.47354	0.37286
	0.7	0.30901	0.47660	0.37493

TABLE III. RESULTS COMPARISON OF EXTRACTING FIVE KEYWORDS

Algorithm	Result			
	<i>t</i>	<i>P</i>	<i>R</i>	<i>F</i>
TextRank	—	0.25656	0.48452	0.33548
Weighted TextRank	0.3	0.28575	0.53687	0.37298
	0.5	0.28616	0.53772	0.37353
	0.7	0.28575	0.53636	0.37286

참조 자료1 에따르면 word2vec을 text-rank 기법에 적용하였더니 다음과 같은 성능 향상이 있었다.

(P: Precision, R: Recall, F: F-score)

따라서 tf-idf대신 word2vec을 rex-rank에 적용하면 중심 문장 추출을 더 정확하게 수행할 수 있을 것으로 판단된다.

문자의 중요도를 구하는 다른 방법에 대한 조사를 하였는데, 참조 자료2 에서는 글에서 자주 등장하는 단어를 추출하여 단어 사이의 내부 경계값과 외부 경계값을 이용하여 단어의 중요도를 추출한다. 이를 사용하여 화자가 사용하는 단어의 중요도를 기반으로 중심

화자를 추출하는 작업을 수행할 수 있을 것으로 판단되었다.

이 두 방법중 어떠한 기법이 프로젝트에 적합한 방법인지 추가적인 연구와 회의가 필요할 듯한다. 추가적으로 전체 화자중 몇 명의 화자를 중심화자로 선택할 것인지의 문제에 대해서 더 고민을 해야 할 것이다.

참조 자료1: Yujun Wen, Hui Yuan and Pengzhou Zhang, "Research on keyword extraction based on Word2Vec weighted TextRank," 2016 2nd IEEE International Conference on

*Computer and Communications (ICCC)*, Chengdu, 2016, pp. 2109-2113, doi: 10.1109/CompComm.2016.7925072.

참조자료 2: Kim, H.-J., Cho, S., & Kang, P. (2014, February 15). KR-WordRank : An Unsupervised Korean Word Extraction Method Based on WordRank. *Journal of Korean Institute of Industrial Engineers*. Korean Institute of Industrial Engineers. <https://doi.org/10.7232/jkiie.2014.40.1.018>

작성자	일자 2023-04-07	확인자	일자 2023-04-07
	서명 신원철		서명 김유성