

재무 보고서의 키워드 검출 기반 딥러닝 감성 분석 기법

Toward Sentiment Analysis Based on Deep Learning with Keyword Detection in a Financial Report

Dongsik Jo¹ · Daewhan Kim² · Yoojin Shin^{3*}

¹Assistant Professor, Department of Digital Contents Engineering, Wonkwang University, Iksan, 54538 Korea

²Senior Researcher, Creative Contents Research Division, Electronics Telecommunication Research Institute, 34129 Daejeon, Korea

^{3*}Assistant Professor, Division of Business Administration, Wonkwang University, Iksan, 54538 Korea

ABSTRACT

Recent advances in artificial intelligence have allowed for easier sentiment analysis (e.g. positive or negative forecast) of documents such as a finance reports. In this paper, we investigate a method to apply text mining techniques to extract in the financial report using deep learning, and propose an accounting model for the effects of sentiment values in financial information. For sentiment analysis with keyword detection in the financial report, we suggest the input layer with extracted keywords, hidden layers by learned weights, and the output layer in terms of sentiment scores. Our approaches can help more effective strategy for potential investors as a professional guideline using sentiment values.

Keywords : Deep learning, Keyword Detection, Financial report, Sentiment Analysis, Text mining

I. 서 론

딥러닝은 인간의 뉴런과 유사한 인공신경망 방식으로 정보를 처리하는 것으로 많은 양의 데이터를 깊은 깊이

(Deep depth) 신경망 구조로 가중치 파라미터 (Parameter)를 학습하여 구성한다. 최근, 이러한 인공 지능 (Artificial intelligence) 기술이 발전함에 따라 인간의 행동, 사물의 상태, 주변 환경의 상황 등 인식 및 예측 기술이 널리 적용되고 있고, 이 기술을 기반으로 한 패턴 인식, 컴퓨터 비전, 자율 주행 등 성능이 급격히 향상되고 있다[1]. 또한, 신문, 잡지, 보고서, 학술논문 등 문서(document)를 대상으로 인공지능 기술을 적용하고, 이를 활용하고자 하는 추세가 확대되고 있다[2]. 즉, 자연어에 관한 텍스트 마이닝(Text mining) 기술을 적용하여 문장의 의미를 분석하는 기술이 점진적으로 발전하고 있다[3,4].

본 논문에서는 텍스트로 구성된 재무 및 회계 정보(예. 주가의 반응, 수익률, 거래량 등)의 질적 특성이 자본시장에 미치는 영향을 파악하기 위한 방법으로 딥러닝 키워드 검출 및 감성분석 기법을 제시하고자 한다. 여기에서 언급된 감성 분석은 보고서에 포함된 텍스트 정보의 성격이 감성적으로 긍정적인 문구인지 혹은 부정적인 문구인지를 판단하는 것으로 일반적으로 콘텐츠에 대한 인간의 감성(기쁨, 슬픔 등) 분석과는 의미가 차이가 있다[5]. 보고서에 대한 감성 분석에 대한 대표적인 기존 연구 사례로 Na et al.의 연구는 감사보고서의 텍스트 마이닝을 기반으로 감성분석을 수행하였다. 하지만, 감성값을 전문가들의 합의에 의해서 결정하였으므로 자동화된 감성 분석에 대한 적용에는 한계가 있다고 할 수 있다[5]. 따라서, 본 논문에서는 이를 극복하고자 딥러닝 키워드 검출을 통한 입출력을 정의하고, 이러한 정보의 다각적 활용을 위한 파라미터 학습 및 의사 결정 구조에 대한 방법론을 제시한다. 요약하자면, 본 논문에서는 텍스트로 구성된 재무 및 회계정보에서 추출된 키워드를 입력으로 한 다층 퍼셉트론(Multi-layer perceptron) 구조의 뉴럴넷을 구성하고, 출력 노드들의 감성값을 이용하여 정보의 유용성을 판단하는 구조에 대한 기법을 제시한다.

본 논문에 제시된 적용 방법을 통해 상대적으로 전문성이 낮은 일반 투자자의 의사 결정에 영향과 도움을 줄

Received 15 March 2020, Revised 29 March 2020, Accepted 28 April 2020

* Corresponding Author Yoojin Shin(E-mail:yooji01@wku.ac.kr, Tel:+82-63-850-6288)

Assistant Professor, Division of Business Administration, Wonkwang University, Iksan, 54538 Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.5.670>

print ISSN: 2234-4772 online ISSN: 2288-4165

수 있을 것으로 예상된다. 예를 들면, 본 논문의 방법으로 데이터가 일반 투자자에게 제공이 된다면 재무 분석가가 제시하는 분석 보고서 및 예측 정보와 회계감사인이 제공하는 감사보고서 등에 대한 자발적인 판단에 도움을 줄 수 있을 것이다. 즉, 전문지식의 보유수준이 낮더라도 텍스트에 포함된 정보가 긍정적인지 혹은 부정적인지를 판단할 수 있는 정보를 제공할 것이다.

II. 재무 정보 딥러닝 감성 분석 기법 흐름도

최근, 대규모의 데이터를 통계적으로 분석하여 사물이나 데이터를 군집화(Clustering), 분류(Classification), 회귀(Regression) 딥러닝 알고리즘에 널리 활용되고 있다[6]. 본 논문은 재무 보고서를 기반으로 감성 분석(예. 재무 및 회계 정보가 긍정적인지 혹은 부정적인지 분석)을 수행하기 위한 딥러닝 키워드 검출 기반의 방식을 제시하고자 한다. 그림 1은 본 논문에서 제시한 재무 보고서의 딥러닝 키워드 검출 및 투자 예측 신경망 구성에 대한 개념도 및 전체적인 흐름을 보여주고 있다. 그림에서 제시한 것과 같이 재무 보고서 기초데이터의 추출이 선행되고, 이는 온라인 혹은 오프라인으로 수집이 될 수 있다. 온라인인 경우는 웹 페이지에서 데이터를 추출하는 웹 크롤링(Crawling) 과정을 통해 데이터를 추출할 수 있다. 오프라인(예. 프린팅 된 신문 등)의 경우는 재무 보고서의 키워드 추출을 위해 텍스트 검출 등 추가적인 과정이 필요하다. 즉, 그림 2와 같이 온라인 혹은 오프라인으로 보유한 재무 보고서에 대해 키워드 데이터 검출을 위한 딥러닝 알고리즘이 추가적으로 필요할 수 있다.

온라인에 비해 상대적으로 복잡한 오프라인 재무 보고서를 가지고 본 논문에서 제시한 기법을 순차적인 과정을 살펴보자면, 문서를 먼저 스캔한 뒤 텍스트 검출 과정을 거친다. 여기에서는 문서에서 특정 키워드(예. 재무 및 회계 정보와 관련된 키워드)를 찾는 키워드 검출 과정이 필요하다. 예를 들면, 이익, 손실, 증가, 기대, 개선, 부진, 불확실성, 성장세, 수혜, 상향 등과 같은 재무 및 회계와 관련된 특정 키워드를 추출하여야 한다. 즉, 이러한 키워드를 추출하기 위한 과정에서도 딥러닝 기반으로 수행하는 것이 추가적으로 필요하다.

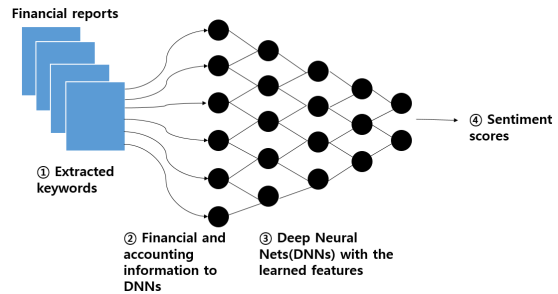


Fig. 1 Deep neural networks(DNNs) for sentiment analysis with keyword detection in a financial report: The Input layer consists of extracted keywords from the financial report, hidden layers are learned with parameters (or weights), the output layer shows scores in terms of sentiment analysis.

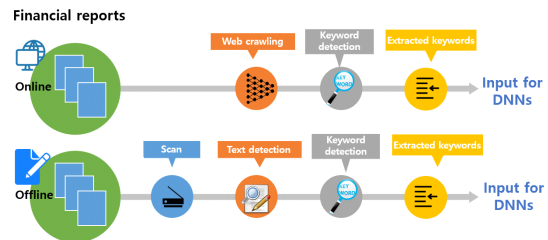


Fig. 2 How to configure deep neural networks(DNNs) for text mining and keyword detection in the financial report.

텍스트로 구성된 기업의 재무정보는 온라인을 통하여 접근하는 경우 비교적 쉽게 구축할 수 있다. 금융감독원 전자공시시스템(<http://dart.fss.or.kr>)에서는 상장기업의 사업보고서 및 감사보고서 자료를 제공하고 있으므로 웹 크롤링 과정을 통하여 텍스트 자료를 데이터베이스화할 수 있다. 이렇게 구성된 자료를 기반으로 딥러닝 감성분석을 수행하는 경우 감성값을 얻을 수 있다.

III. 재무 정보 딥러닝 감성분석 모형

재무 및 회계 정보 자료에서 추출된 특정 키워드는 딥러닝에 입력 레이어(Input layer)로 제공이 되고, 사전에 학습된 hidden layer를 통해 -1에서 1까지 normalization 된 형태의 감성값으로 출력되게 된다.

출력으로 나온 감성값은 highly positive, positive, normal, negative, highly negative와 같은 단계로 간단한

형태로 분류되어 정보이용자의 의사 결정에 유용한 데이터로 출력이 되게 할 수 있다. 예를 들면, 재무 보고서에서 나온 감성값이 0.5보다 크다면 highly positive로 매핑이 되어 보고서 정보가 매우 긍정적인 정보로 판단될 수 있도록 하고, 0.3이 나왔다면 positive로 매핑이 되어 긍정적인 정보로 판단될 수 있게 되도록 할 수 있다.

이렇게 산출된 감성값은 기존 연구에서 제시한 감성값과 차이가 있을 것으로 예상된다. 즉, Na et al.의 연구에서는 감사보고서의 텍스트 마이닝을 기반으로 전문가들의 합의에 의하여 키워드를 분류하여 감성값을 산출하였다[5]. 그러나 본 연구는 딥러닝을 통하여 키워드 분류 과정을 수행하고자 한다. 이 경우 감성값을 도출하는 과정에서 객관성이 향상될 것으로 예상되며 도출된 감성값도 다를 수 있다고 예상된다. 이는 기존 연구에 의한 감성값과 딥러닝을 활용하여 도출한 감성값의 평균을 비교(t-test)하여 검증할 수 있을 것이다.

딥러닝을 통하여 산출된 감성값은 다양한 분야에 적용될 수 있다. 구체적으로, 감성값이 자본시장에 미치는 다각적인 영향을 분석하기 위하여 추가적으로 회귀분석을 수행할 수 있다. 즉, 회귀분석 모형에 본 논문에서 추출한 감성값을 반영할 수 있으며 회귀분석 모형은 아래와 같은 수식으로 표현할 수 있다.

$$Dep. Variables_{i,t} = \beta_0 + \beta_1 Senti_{i,t} + K + \varepsilon \quad (1)$$

여기에서 독립변수인 Dep. Variables는 텍스트로부터 얻어진 감성값이 유의한 영향을 미치는지를 파악하기 위한 대상에 해당되고, Senti는 본 논문에서 제시한 감성값이다. 그리고, K는 기업 규모, 재무 구조, 성장성, 당기성과, 영업 현금 흐름, 고객 기업의 영업 위험, 유형 자산 비중, 영업환경 복잡성, 손실여부, 대주주 지분율, 외국인 지분율, 대형 감사인 여부, 연도 및 산업고정 효과 등 기업의 환경과 관련된 통제변수이다. 상기의 모형에서 독립변수는 다양한 항목으로 설정될 수 있다. 예를 들어, 감성값이 주가에 미치는 영향을 분석하고자 한다면 종속변수를 주가로 설정한다. 주가 외에도 주식의 거래량, 변동성, 수익률 등등의 변수로 설정하여 텍스트 정보의 감성값이 자본시장에 미치는 영향이 유의한지 파악할 수 있다.

본 논문에서 제시한 방법에 의하면 보고서에서 재무 및 회계 정보를 추출하여 보수를 계산하는 입력 값인 감성값과 K값을 계산하는 것이 가능하다. 이러한 변수의

측정이 중요한 이유는 기업가치의 추정 및 주가예측 등에 활용될 수 있기 때문이다.

기업의 내재가치를 평가하기 위한 가치평가 모형은 효율적 시장가설을 기초로 한다. 효율적 시장가설 (Efficient Market Hypothesis) 하에서는 주가가 기업가치와 관련된 정보를 반영하고 있다고 가정하고 있으나 실제적으로 주가가 기업의 진정한 내재가치를 정확히 반영하지 못하고 있다. 따라서 기업가치를 정확히 추정할 수 있는 가치평가모형이 제시된다면 이러한 모형이 활용될 수 있는 범위는 상당할 것이다. 그러나 현재까지 제시된 가치평가 모형 하에서는 정확한 가치평가를 수행할 수는 없다. 지금까지 제시된 가치평가모형 중 가장 대표적인 가치평가모형은 Ohlson 모형으로 이는 다음과 같다[7].

$$P_t = BV_t + \alpha_1 X_t^a + \alpha_2 v_t \quad (2)$$

$$\alpha_1 = \frac{w}{(1+r_e - w)},$$

$$\alpha_2 = \frac{(1+r_e)}{(1+r_e - w)(1+r_e - \gamma)}$$

상기 모형에서 P는 주가, BV는 장부가치, X는 이익, v는 이익에 반영되지 않은 가치관련정보(비회계정보)를 의미하고 a는 가중치를 의미한다. 즉, 기업의 가치는 기업의 장부가치와 이익정보, 기타정보로 결정됨을 의미한다. 여기에서 장부가치와 이익정보는 재무제표를 통하여 쉽게 얻을 수 있으나 기타정보에 대한 정확한 파악은 쉽지 않을 수 있으나 본 연구결과를 통해 산정될 수 있겠다. 또한, 기업의 특성에 따라서 비재무적정보의 중요성이 큰 경우도 있으므로 본 연구의 방법은 기업에 대한 정보를 유추하는데 도움이 될 수가 있겠다[8]. 따라서, 보고서의 텍스트에 포함된 정보의 감성값이 기타 정보의 상당부분을 차지한다면 본 연구에서 제시된 방법론을 적용하여 가치평가의 정확성을 높일 수 있을 것으로 예상된다.

IV. 결 론

인공 지능(Artificial Intelligence) 기술이 발전함에 따라 보고서 등의 문서에 대한 감성 분석 기술이 향상되고 있다. 본 논문에서는 딥러닝 기술을 활용하여 재무 보고

서에 포함된 텍스트에 대하여 텍스트마이닝 기술을 적용하고, 문장의 의미를 분석하여 재무 및 회계 정보의 감성분석을 수행하는 기법을 제시하였다. 본 논문에서는 제시한 방법에 따라 산출된 정보는 일반투자자 혹은 전문가 모두에게 유용하게 활용될 수 있으며 특히 일반 투자자의 경우 유용성이 더욱 높을 것으로 예상된다. 즉, 전문적인 지식이 결여된 상황에서도 감성값을 활용하여 투자 의사결정에 가이드라인으로 활용할 수 있을 것으로 본다.

본 연구에서 제시된 방법론은 추후 연구로 확장될 가능성이 높다. 구체적으로, 딥러닝을 기반으로 하여 텍스트로 구성된 재무 및 회계정보에 대한 추출을 더욱 구체화할 수 있으며 감성값을 자동적으로 계산하는 알고리즘을 구현하고 신뢰도 테스트를 수행할 수 있다. 또한, 텍스트로 구성된 정보의 출처(정형화된 보고서 또는 뉴스 기사, 칼럼, 논문 등)에 따라 키워드의 신뢰도의 차이가 있을 것으로 예상되므로 감성값 계산 시 정보의 출처에 따라 딥러닝을 활용하여 가중치를 부여하는 방법 등에 대한 연구를 추가적으로 진행하고자 한다.

ACKNOWLEDGEMENT

This paper was supported by Wonkwang University in 2019.

REFERENCES

- [1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May. 2015.
- [2] S. A. Oliveira, B. Seguin, and F. Kaplan, "dhSegment: A generic deep-learning approach for document segmentation", in *Proceeding of the 16th International Conference on Frontiers in Handwriting Recognition*, Niagara Falls: NY, 2018.
- [3] N. I. Widiastuti, "Deep learning- Now and next in text mining and natural language processing", in *Proceeding of the Materials Science and Engineering*, Bandung:Indonesia, 2018.
- [4] J. K. Jang, K. H. Lee, and Z. Lee, "How the title of investment strategy report affects stock price forecast: using text mining method," *The Korean Journal of BigData*, vol.1, no.2, pp.21-34, 2016.
- [5] H. J. Na, K. C. Lee, S. U. Choi, and S. T. Kim. "An analysis on the validity of audit opinion using unstructured data from audit reports and audit fees and hours: emphasis on utilizing text mining and sentiment analysis," *Korean Accounting Review*, vol.44, no.4, pp. 175-214, 2019.
- [6] I. H. Sarker, A.S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone," *Journal of Big Data*, vol. 6, no.57, 2019.
- [7] J. Ohlson, "Earnings, book values, and dividends in equity valuation : an empirical perspective", *Contemporary Accounting Research*, Vol.11 No.2, pp.661-687, 1995.
- [8] J. Kim, "The Effects of Financial Information to the Firm Valuation for Information Technology Related Companies : Evidences from Software, Digital Content, Internet Related Companies listed in KOSDAQ", *The Journal of Society for e-Business Studies(JSEBS)*, Vol.17 No.3, pp.73-84, 2012.