

텍스트마이닝과 연관규칙을 이용한 외부감사 실시내용의 그룹별 핵심어 추출

성윤석* · 이동희** · 정옥***†

* 동국대학교-서울 경영대학 회계학과 박사과정생

** 동국대학교-서울 경영대학 경영학과 학부생

*** 동국대학교-서울 경영대학 경영학과 교수

Group-wise Keyword Extraction of the External Audit using Text Mining and Association Rules

Seong, Yoonseok* · Lee, Donghee** · Jung, Uk***†

* Ph.D. student, Dept. of Accounting, School of Business, Dongguk University-Seoul

** Undergraduate student, Dept. of Management, School of Business, Dongguk University-Seoul

*** Professor, Dept. of Management, School of Business, Dongguk University-Seoul

ABSTRACT

Purpose: In order to improve the audit quality of a company, an in-depth analysis is required to categorize the audit report in the form of a text document containing the details of the external audit. This study introduces a systematic methodology to extract keywords for each group that determines the differences between groups such as 'audit plan' and 'interim audit' using audit reports collected in the form of text documents.

Methods: The first step of the proposed methodology is to preprocess the document through text mining. In the second step, the documents are classified into groups using machine learning techniques and based on this, important vocabularies that have a dominant influence on the performance of classification are extracted. In the third step, the association rules for each group's documents are found. In the last step, the final keywords for each group representing the characteristics of each group are extracted by comparing the important vocabulary for classification with the important vocabulary representing the association rules of each group.

Results: This study quantitatively calculates the importance value of the vocabulary used in the audit report based on machine learning rather than the qualitative research method such as the existing literature search, expert evaluation, and Delphi technique. From the case study of this study, it was found that the extracted keywords describe the characteristics of each group well.

● Received 1 February 2022, 1st revised 17 February 2022, accepted 24 February 2022

† Corresponding Author(ukjung@dongguk.edu)

© 2022, Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited..

* 본 연구는 2022학년도 동국대학교 논문게재장려금 지원으로 이루어졌음.

Conclusion: This study is meaningful in that it has laid the foundation for quantitatively conducting follow-up studies related to key vocabulary in each stage of auditing.

Key Words: External Audit, Text Classification, Variable Importance, Association Rules, Group-wise Keywords

1. 서 론

2022년초 대한민국 주식시장은 상장사 오스텝임플란트 자금 담당 직원의 2000억 규모 횡령 사건으로 충격에 휩싸였다. 현대 외부감사 방법론은 위험-기반 접근법(Risk-based approach)을 기반으로 구축되어 있다(Heo and Jung, 2013). 위험-기반 접근법에 따라 내부회계관리제도(Internal accounting control system)의 프로세스별 위험을 평가하고, 그에 따라 감사자원을 효율적으로 배분할 때 감사품질(Audit quality)을 극대화 할 수 있다는 것이 방법론의 핵심이다. 상술한 횡령사건으로 큰 피해를 입게 된 정보이용자들은, 해당 기업의 외부감사인 내부회계관리제도상의 위험 평가에 충분한 자원 배분을 하였는지에 대한 의문을 갖게 되었다. 내부회계관리제도는 회사의 재무제표가 일반적으로 인정되는 회계처리기준에 따라 작성·공시되었는지에 대한 합리적 확신을 제공하기 위해 설계 및 운영되는 내부통제제도의 일부분으로서 회사의 경영진과 이사회를 포함한 모든 구성원들에 의해 지속적으로 실행되는 과정을 의미한다 (Kim, 2019). 특히, 개정된 주식회사 등의 외부감사에 관한 법률(이하, “외감법”)에 따라 해당 기업이 자금운영 프로세스를 포함한 내부회계관리제도에 대하여 감사를 받았으며, 적정의견이 표명 받았다는 것에 그러한 의문은 심화되었다. 이처럼 내부회계관리제도에 대한 감사인의 충실한 업무 수행 여부는 정보 이용자들의 주된 관심 사항이 되고 있다. 보다 구체적으로는, 예를 들면, 감사보고서에 첨부된 외부감사 실시내용을 분석하여 감사계획(Audit planning)과 중간감사(Interim audit)의 투입 비중을 추정함으로써 위험평가 투입 비중이 최종 감사품질에 어떠한 영향을 미치는 지에 대한 관심으로 이어질 수 있다(Kwon, et al., 2016). 따라서 외부감사 실시내용의 주요 감사실시내용에 포함된 문서(텍스트) 형태의 감사업무수행 정보를 연구의 목적에 따라 그룹별로 구분하는 등의 다각적이고 심층적인 분석이 요구되고 있다.

본 연구에서는 이러한 문서 자료의 형태로 수집된 외부감사 실시내용의 주요 감사실시내용으로부터 기계학습(Machine learning) 기법을 이용하여 위험평가, 입증감사절차와 같은 그룹별로 문서를 자동 분류하고, 그룹별 차이를 결정짓는 그룹별 핵심어휘를 추출하기 위한 텍스트 마이닝(Text mining)과 연관규칙 분석(Association rules analysis)에 기반한 체계적 방법론을 소개한다. 이러한 방법론은 기존 연구(Gray et al. 2014; Chan and Vasarhelyi 2018; Na et al., 2019)에서 전통적 회계 분야에서도 인공지능(Artificial intelligence)에 기반한 도구들이 적용될 필요성을 강조한 바와 연결된다. 본 연구에서 제안하는 방법론의 핵심내용은 다음과 같다. 먼저 첫번째 단계로 텍스트 마이닝을 통한 문서의 전처리(Preprocessing)를 수행한다. 이 과정에서 불용어와 조사들은 제거한다. 두번째 단계로는 전문가 집단의 경험과 지식을 바탕으로 라벨(label)을 부여하여 훈련용 데이터셋을 준비한다. 그리고 기계학습 기법을 사용하여 훈련용 데이터셋의 문서들을 그룹별로 분류하는 분류모형(Classifier)을 개발하고 이를 바탕으로 분류의 성능에 지배적인 영향을 끼치는 중요 어휘들(A)을 추출해 낸다. 그리고 세번째 단계로 각 그룹별 문서들에 대한 연관규칙을 추출해 내고 이를 바탕으로 그룹별 중요한 어휘들(B)을 찾아낸다. 네번째 단계로는 분류를 위한 중요 어휘들(A)과 각 그룹의 연관규칙들을 대표하는 중요 어휘들(B)을 상호 비교하여 각 그룹별 특징을 대변하는 그룹별 최종 핵심어휘들(C)을 추출해낸다.

본 연구의 학술적 의의는 다음과 같다. 첫째, 본 연구는 기계학습 기법 중의 하나인 텍스트 마이닝과 연관규칙분석 기법을 적용하여 외부감사 실시내용상에 사용된 어휘들에 대한 특징별(그룹별) 중요도값을 계량적으로 산출하고 이를 토대로 감사보고서의 특징별 핵심어휘를 추출한다. 이는 향후 회계감사와 관련된 연구에서 기계학습 기법이 어떻게 적용될 수 있는지를 이해하는데 유용한 실증 분석의 틀을 제시한다고 볼 수 있다. 둘째, 외부감사 실시내용에서 사용된 다양한 어휘들을 텍스트 마이닝 기법인 백 오브 워즈(Bag of Words) 방법을 적용하여 효과적으로 형태소 분석을 실시하고 해당 어휘들 중 명사에 해당하는 어휘들을 중심으로 문서-어휘 행렬을 구축하여 외부감사 실시내용의 주요 감사실시내용의 그룹 분류에서 핵심적인 역할을 하는 명사들을 추출한다. 이렇게 추출된 명사 어휘들은 향후 자동화된 주요 감사실시내용 분류 결과의 재확인 및 감사업무의 단계별 핵심어휘와 관련된 후속 연구에도 효과적으로 참고할 수 있는 토대가 될 것으로 기대된다. 넷째, 외부감사 실시내용의 어떤 특성들이 유의미한 정보를 전달하는가에 대한 연구는 지속적으로 이루어져왔으나 이는 대부분 문헌 조사, 전문가 평가, 델파이(Delphi) 기법과 같은 정성적(Qualitative) 연구방법이 사용되어 왔다. 본 연구는 외부감사 실시내용의 주요 감사실시내용을 명사 어휘 단위로 구성하고 이들 명사 어휘의 중요도를 정량적인 수치로 측정하게 함으로써 향후 회계감사 관련 연구를 보다 정량적으로 수행 할 수 있는 기틀을 마련할 것으로 기대된다. 마지막으로, 본 연구 결과 식별된 핵심 어휘로 정보이용자의 외부감사 실시내용 이해가능성을 제고하고, 연구 결과로 규제당국이 공시제도 개선점을 파악할 수 있을 것이라 기대한다.

본 연구의 구성은 다음과 같다. 제 2장에서는 외부감사 실시내용을 바탕으로 감사품질을 측정하는 관련연구들과 기계학습에서의 텍스트 마이닝 관련 연구들이 어떻게 이루어져 왔는지 관련 연구들을 간략히 소개한 후 제 3장에서는 본 연구에서 사용할 방법론과 제안하는 방법론을 소개한다. 제 4장에서는 제안하는 방법론의 적용 과정과 분석 결과에 대해서 서술한다. 마지막으로 제 5장에서 본 연구의 결론과 시사점을 서술한다.

2. 관련 연구

2.1 회계 및 회계감사 분야에서 기계학습 및 텍스트 마이닝의 활용

회계 및 회계감사 분야에서 기계학습 및 텍스트 마이닝은 주로 귀납 추론을 통한 예측(prediction) 또는 연역 추론을 위한 변수 생성(variable generation) 목적으로 활용되었다(Cho et al., 2020). Goltz and Mayo(2017)는 사전에 위험을 미리 감지하는 시스템을 도입하는데 기계학습과 텍스트 마이닝이 주된 역할을 하게 될 것이라고 주장하였다. 또한 Byrnes et al.(2018)은 향후의 감사업무가 기계학습의 도움으로 자동화 될 것이라고 주장하였다. Boskou et al.(2018)은 연간 재무제표의 문자들을 추출하여 유의하게 반복되는 핵심어휘들을 추출하고 유사한 어휘들을 그룹으로 묶어 내부 통제 기능의 지표(indicator)를 개발하는데 사용하였다. Kamaruddin et al.(2015)은 재무제표 문서 내에 존재하는 비정상 문장을 탐지하기 위해 재무제표 내의 텍스트 데이터와 수치 데이터를 수집하여 텍스트 마이닝 기법을 적용하였다. 현재까지의 관련 연구는 회계처리기준을 위반한 기업을 예측하기 위한 연구가 주를 이루었다. 이들은 주로 텍스트 마이닝을 통해 재무제표를 재작성해야 할 가능성이 높은 기업을 사전에 예측하는 것을 목적으로 하였다. 또한 Jung et al.(2019)은 한국어로 작성된 재무제표의 경우 한국어 형태소 분석기의 문제점을 지적하며 재정정보의 특성을 반영하여 회계용어가 사용된 텍스트를 분석하기 위한 형태소 분석기를 구축하는 연구를 수행하였다.

이러한 다양한 회계 및 회계감사 분야의 기계학습 및 텍스트 마이닝 관련 연구들 중에 회계감사업무의 단계적 업무 흐름을 구분하는 것을 목적으로 한 적은 없었다. 본 연구에서는 기존 연구와 다른 문서 분류의 목적을 지니며 각 부류별 핵심어휘를 기계학습 분류기와 연관규칙 분석법을 함께 이용하여 추출하는 것에 초점을 두어 기존 연구와 차별화 된다고 볼 수 있다.

2.2 감사품질과 감사의 단계별 분류

외감법에 따라 감사를 수행한 감사인은 동법 제18조(감사보고서의 작성)에 따라 감사보고서에 외부감사 실시내용을 첨부하게 된다. 외부감사 실시내용에는 감사업무의 수행에 관한 다양한 정보가 제공되고 있다. Blokdiik et al.(2006)은 감사품질(audit quality)을 투입의 질(input quality), 투입량(input intensity), 그리고 감사 기법(audit technology)의 함수라 보았으며, 이를 각각 직급별 배합, 총 투입 시간, 단계별 배합이라는 대응치로 측정하였다. 주요 감사 실시내용에 포함된 감사업무의 단계별(감사계획, 중간감사, 입증감사절차 등) 수행 일수와 투입 인원 정보는 감사 기법의 수준을 측정하기 위한 활용 도구인 단계별 배합을 추정할 수 있는 정보를 제공하고 있다. 이러한 단계별 배합에 관한 정보는 선행연구의 맥락에서 고품질 감사업무의 수행을 위한 필요 조건을 파악하는데 유용하다.

Kwon, et al.(2016)은 외부감사 실시내용이 최초로 공시된 2014년의 상장기업 감사보고서에 첨부된 외부감사 실시내용 1,817건을 수작업으로 입수·가공하여 위험평가단계(감사계획과 중간감사)의 투입 비중을 추정하였다. 해당 연구는 단계별 배합, 특히 위험평가 투입 비중이 총 감사시간과 감사보수에 미치는 영향을 분석하였다. 해당 연구는 위험평가 투입 비중이 클 수록 감사의 효율이 증가한다는 것을 실증하여, 위험-기반 접근법의 효과성을 확인하였다. 위험평가 투입 비중에 대한 관심은 상장기업에서 전체 외부감사기업으로 확대될 수 있고, 또한 2014년 이후 기간으로 연장될 수 있다. 주요 감사실시내용, 특히 현장감사 주요내용은 회계법인별로, 그리고 업무팀에 따라 서로 다른 양식으로 제공하고 있으며 단계별 배합에 따라 정형화된 분류 결과(감사계획, 중간감사, 입증감사절차)로 매핑(mapping)하는 작업이 필요하다. 확대된 대상·표본에 대해서 Kwon, et al. (2016)에서 행한 바와 같이 수작업으로 분류 과업을 수행하는 것은 요구되는 전문인력 및 필요 시간 등을 고려해 볼 때 현실적으로 매우 어려운 상황이다. 2016년에서 2020년까지 5년간 외부감사 대상의 감사보고서 수가 152,025건에 달한다는 점에서 이를 확인할 수 있다.(금융감독원. 2020년 외부감사대상 회사 및 감사인 지정 현황. 보도자료. 2021년 1월 27일).

2.3 텍스트 분류 (Text Classification) 및 단어 중요도(Word Importance)

문서분류(text classification)란 기계학습 알고리즘을 통해 문서를 여러 개의 부류중 하나로 분류하는 작업을 말한다. 이런 문서분류 기술은 방대한 텍스트 데이터가 넘쳐나는 오늘날 수작업으로 수행하기 불가능한 분류 작업을 자동화하는데 성공적으로 사용되고 있다. 예를 들면, 웹상의 다양한 문서들을 주제별로 분류하거나, 이메일에서 스팸 메일을 걸러내는 등 다양한 상황에서 적용된다(Aggarwal and Zhai, 2012). 이처럼 분류 작업을 수행하는 수리적 모형을 분류기(classifier)라고 하는데 이는 의사결정 트리(Decision Tree), 선형 판별 분석(Linear Discriminant Analysis: LDA), 로지스틱 회귀분석(Logistic Regression), 서포트벡터 머신(SVM), 신경망(Neural Networks), 랜덤 포레스트(Random Forest: RF) 등 다양한 종류가 존재한다. 이들은 각각 다른 알고리즘을 사용하고 대상이 되는 문서의 종류에 따라 그 정확도에도 차이가 나타난다.

본 연구에서는 여러 분류기 모형들 중에서도 분류의 정확성과 변수들의 중요도 측정에서 특히 강점이있는 랜덤 포레스트(Random forest: 이하 RF) 모형 (Breiman, 2001)을 사용한다. RF모형은 부트스트랩(bootstrap)방식을 이

용하여 다수의 표본을 생성하고, 의사결정나무(decision trees) 모형을 적용하여 그 결과를 종합하는 앙상블 학습(ensemble learning) 방법이다. 부트스트랩을 통한 무작위성은 의사결정나무들 간에 상관관계가 줄어들게 하고 이로 인해 예측오차가 줄어들게 된다. RF모형은 특히 설명변수의 수가 다수일 때 매우 훌륭한 예측력을 보이는 매우 안정적인 모형이다(Siroky, 2009). 이는 하나의 문서가 많은 어휘를 설명변수로 사용하여 표현되는 본 연구의 상황에서 특히 그 가치가 높아진다고 볼 수 있다. 그러나 부트스트랩 표본을 몇 개로 할 것인지, 각 마디에서 설명변수의 개수를 몇 개로 할 것인지, 결과 종합 시 다수의 분류 결과들을 어떻게 선형결합 할 것인지 등은 초매개변수를 이용하여 연구자가 선택해야한다.

또한 RF모형은 각 변수의 중요도를 학습과정을 통해 제공할 수 있다. 즉, 텍스트로 구성된 문서의 경우에는 하나의 문서를 표현하는 어휘들의 중요도를 계산할 수 있다는 것으로 해석될 수 있다. 이는 분기되는 각 의사결정나무 마디마다 특정 변수가 분류의 정확성에 미치는 영향력을 기초로 하여 측정되는 값이다. 어떤 변수들이 종속변수를 예측하는데 있어서 중요한 변수로 고려되어야 할 것인가를 알아보기 위해서 OOB(out-of-bag) 자료로 실제 관측값과 예측된 값과의 차이를 이용해서 오차를 구하는 방법을 사용한다(Hastie et al., 2001; Strobl et al., 2009). 이때 그 변수의 영향력이 클수록 중요도 지수가 커진다(Strobl et al., 2009).

2.4 연관규칙 분석

연관규칙(Association rules) 분석은 데이터 안에 존재하는 여러 항목 간의 상호 관계를 분석하는 것으로 장바구니 분석이라고도 한다(Zhao and Bhowmick, 2003; Kim and Cho, 2021). 이때 손님의 장바구니에 속한 다수의 항목들은 하나의 트랜잭션(Transaction)을 구성하게 되며 트랜잭션에 존재하는 아이템 리스트(Item list)들의 발생 규칙을 수치화하는 것이다. 텍스트로 구성된 문서의 경우 각 문서가 트랜잭션이 되며 문서에 포함된 특정 어휘 간 연관성을 찾아내면 되는데 연관성을 측정하는 지표로는 지지도(Support), 신뢰도(Confidence)와 향상도(Lift) 값가 주로 사용된다(Kim et al., 2021). 지지도는 전체 문서 중 단어 A와 단어 B가 동시에 발생하는 확률을 의미한다. 신뢰도는 단어 A를 포함한 문서 중에서 단어 A와 단어 B가 함께 발생할 확률을 의미한다. 지지도와 신뢰도는 1에 가까울수록 연관성이 높다고 할 수 있으며 주로 너무 많은 연관규칙들이 나타나는 것을 피하기 위하여 일정 수준 이상인 지지도와 신뢰도를 가지는 연관규칙들을 소수로 걸러 내는데 사용된다. 향상도는 단어 B가 발생할 확률에 비해 단어 A가 발생하였을 때 단어 B의 발생 확률의 비율이다. 향상도가 1이면 두 단어의 발생 연관이 서로 관련이 없는 우연적인 결과이고, 1보다 크면 두 단어의 발생 연관이 우수하며, 1보다 작으면 두 단어의 발생 연관은 우연적 기회보다 좋지 않음을 의미한다(Kim and Cho, 2021).

본 연구에서 연관규칙 분석을 활용하는 주된 목적은 다음과 같다. 외부감사 실시내용을 감사흐름의 단계별로 구분하는데 주된 역할을 하는 핵심어휘를 기계학습을 통한 분류기로 추출한다고 할지라도 이들 핵심어휘들이 감사흐름의 어느 단계를 특징짓는지는 명확하지 않다. 만약 각 감사흐름 단계에 따른 문서들을 대상으로 빈도를 파악하여 핵심어휘를 구분하는 방법을 취한다면 이는 다양한 단계의 문서 전체적으로 빈도가 높은 단어들이 추출됨에 따라 다른 감사흐름 단계에서 중복된 단어들이 추출될 가능성이 높다. 따라서 본 연구에서는 문서 분류의 정확도를 높이는 핵심어휘들을 다시 각 감사흐름 단계별 연관규칙을 생성하는 주요 어휘들과 비교하여 각 단계별 핵심어휘를 추출하고자 한다.

3. 외부감사 실시내용 그룹별 핵심어 추출 방법론

본 연구에서 수행한 방법론은 (1단계) 문서 전처리, (2단계) 문서 분류 및 어휘별 중요도 산출, (3단계) 연관규칙 추출, (4단계) 그룹별 핵심키워드 생성의 네 단계로 구성되며 Figure 1.과 같다.

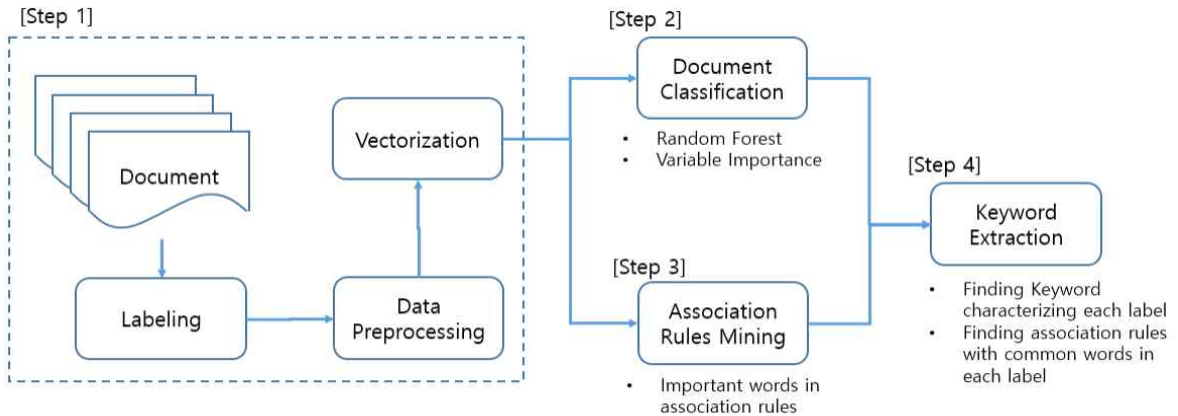


Figure 1. Group-wise Keyword Extraction Framework

1단계로는 문서 형태의 수집된 주요 감사실시내용을 전문가의 확인을 거쳐 일정 비율에 대해 라벨을 할당한다 (labeling). 그리고 수집된 문서에 대해 형태소 분석(part-of-speech tagging; POS tagging)을 실시하여 품사가 명사인 단어들만을 추출한다. 이후 불필요한 구두점, 숫자, 공백을 제거한다. 그리고 유의미한 분석을 위해 출현빈도가 일정 수준 미만이면서 문자열의 길이가 2개 미만인 단어는 제외한다. 단어가 모인 문서집합을 BOW 방식으로 단어 빈도(term frequency)가 행렬의 값이 되는 문서-단어 행렬을 구성한다.

2번째 단계로는 기계학습의 분류기를 통해 문서-단어 행렬을 이용한 문서분류를 수행한다. 이때 문서분류를 수행하는데 중요한 역할을 하는 단어들을 그 중요도 순으로 추출한다. 즉, 분류기를 통한 분류의 결과가 정확하다면 분류에서 중요한 역할을 하는 단어들을 파악함으로써 각 라벨별 문서들의 특징을 추측해 볼 수 있다. 이는 문서들을 사용자가 재확인하는 과정에서도 업무의 효율성에 큰 도움을 줄 수 있을 것이다. 본 연구에서는 분류의 정확성과 중요도 측정에서 특히 강점이 있는 RF 모델을 사용한다. 하지만 분류에서의 변수 중요도를 산출할 수 있는 다른 방법들, 예를 들면 정규화(regularization)를 이용한 로지스틱 회귀(logistic regression) 등도 가능하다.

3번째 단계로는 각 라벨별 문서집합을 대상으로 연관규칙 분석을 수행하여 각 라벨에서 출현하는 핵심적인 단어들의 연관규칙을 찾아내고 이를 바탕으로 해당 라벨을 대표하는 중요 단어들을 파악한다. 이때 중요 단어들은 해당 그룹을 대표하는 규칙들을 구성하는 단어들을 의미한다.

마지막 4번째 단계에서 2단계와 3단계의 결과물을 상호 비교하여 각 라벨들을 구분 짓는 핵심적인 단어들을 찾아낸다. 이때 각 라벨에서 단독으로 사용되는 단어들은 해당 라벨의 고유한 핵심어휘로 파악된다. 만약 2단계의 중요 단어들이 3단계의 연관규칙에서 발굴된 복수개의 라벨별 단어 모음에서 동시에 모두 출현한다면 각 해당 라벨에서 어떤 규칙으로 어떤 다른 단어들과 함께 쓰이고 있는지를 조사함으로써 동일한 단어이지만 각 라벨별 연관규칙의 차이를 파악하게 된다.

4. 분석 과정 및 결과

4.1 데이터 수집 및 전처리

본 연구는 외부감사법에 의한 법정 감사업무 중 보고기간이 2017년에서 2019년까지인 건들을 대상으로 한다. 본 연구에서 요구되는 외부감사 실시내용의 텍스트 자료는 금융감독원 전자공시시스템(DART)에서 HTML 형식으로 직접 입수하였다. 연구 표본은 기업-연도에 따른 94,659 건의 감사보고서에 첨부된 외부감사 실시내용에 해당하며 분석 대상 주요 감사업무 수행 내용에 포함된 현장감사 주요내용은 206,585건이다. 본 연구는 입수된 문서에서 974 개의 표본을 임의로 추출한 후, 연구자의 전문성을 활용하여 훈련용 데이터 생성을 위한 초기 분류(Label 1(위험평가) 416건과 Label 2(입증감사절차) 558건)를 수행 하였다. 감사 업무의 흐름은 감사계획, 거래유형 테스트, 계정잔액 테스트, 감사 종료 및 보고로 나눌 수 있으며, 이는 위험-기반 접근법의 관점에서 다시 위험평가(감사계획과 거래유형테스트)와 입증감사절차(계정잔액 테스트, 감사 종료 및 보고)로 묶을 수 있다는 관점에서 결정한 분류이다. 본 연구의 입수 문서는 길이가 짧고, 중복이 다수 포함되어 있으나 임의 추출 과정에서 이를 고려하지 않고, 전체 집단에서 표본 추출하였다. 분류된 문서에 대해 지도학습 기반의 형태소 분석기 Komoran과 비지도학습 기반의 형태소 분석기 Soynlp를 사용해 형태소 분석(part-of-speech tagging; Pos tagging)을 실시하였으며 불용어를 제거한 뒤 Table 1와 같이 품사가 명사인 단어만 추출하였다. 최종적으로 명사를 추출한 문서집합을 문서-단어 행렬(Document-term matrix; 이후 DTM)과 TF-IDF 행렬(Term frequency inverse document frequency matrix; 이후 TFIDF)로 변환하였다.

Table 1. Example of Documents and Results of Morphology Analysis

Documents	Label	Komoran	Soynlp
중요 계정잔액 및 공시내용에 대한 질문 분석적 검토 문서검토 재계산 등	2	계정 잔액 공시 내용 질문 분석 검토 문서 검토 계산	중요 계정잔액 공시 내용에 질문 분석적 검토 문서 검토 재계산
기말감사	2	기말 감사	기말 감사
내부통제 이해 및 평가	1	내부통제 이해 평가	내부통제 이해 평가
내부통제 검토 및 중간감사	1	내부통제 검토 중간 감사	내부통제 검토 중간감사
계정잔액 및 주식에 대한 입증감사절차 수행 및 내부회계관리제도 평가	2	계정 잔액 주식 입증 감사 절차 수행 내부 회계 관리제 평가	계정잔액 주식 입증감사 절차 수행 내부회계관리제도 평가
...

4.2 문서 분류 및 분류기 내 중요 단어 추출

형태소 분석기 및 텍스트 분류기의 선정을 위해 4.1절에서 구축한 DTM과 TFIDF를 기반으로 RF 모형을 이용하여 문서 분류를 수행하였다. 성능을 우수성을 비교하기 위하여 텍스트 분류에서 자주 쓰이는 Naive Bayes 분류기도 함께 적용하였다. 이후 정확도(Accuracy)를 지표로 활용하여 5겹 교차검증(5-fold cross validation)을 적용하여 형태소 분석기 및 텍스트 분류기의 성능을 평가하였다. Figure 2에서 보여지는 바와 같이 문서 분류의 평균 정확도는 Komoran-DTM-RF의 조합(KDR)으로 구성했을 때 0.9682로 가장 높은 평균값을 보였으며 다른 조합들에 비해 상

대적으로 낮은 분산을 보였다. 이는 후속 단계에서 RF 모델을 문서 분류기로 사용하는 것이 충분한 수준의 신뢰성을 보이는 것으로 볼 수 있다.

가장 높은 수준의 분류 정확도를 보인 Komoran 기반의 문서-단어 행렬을 이용한 RF 모형(KDR 모형)을 이용하여 분류에서 큰 영향력을 미친 중요한 단어들을 그 중요도의 순서대로 정렬하여 상위 20개 단어를 확인하면 다음 Figure 3와 같다. 순서대로 보면 ‘중간’, ‘내부통제’, ‘평가’, ‘기말’, ‘이해’, ‘입증’, ‘계정’, ‘절차’, ‘감사’ 등의 순으로 나타났다.

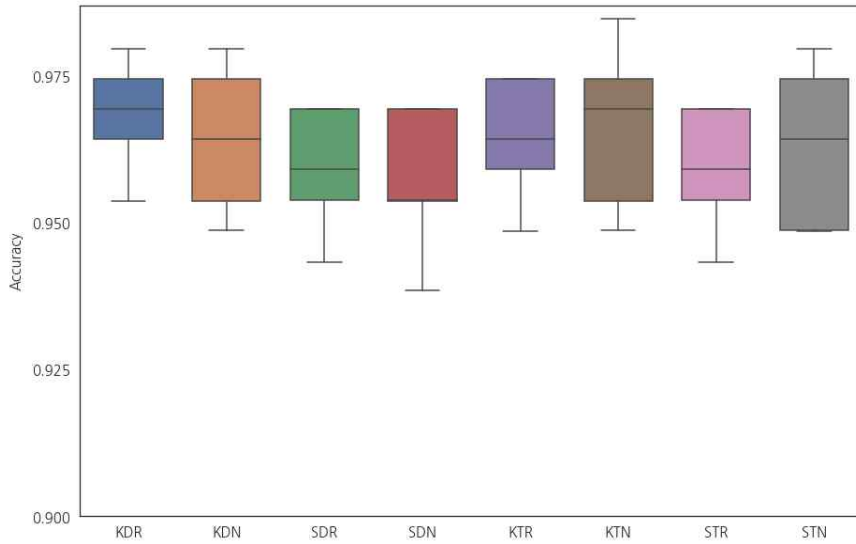


Figure 2. Classification accuracy comparison (K=Komoran; S=Soynlp; D=DTM; T=TFIDF; R=RF; N=Naive Bayes)

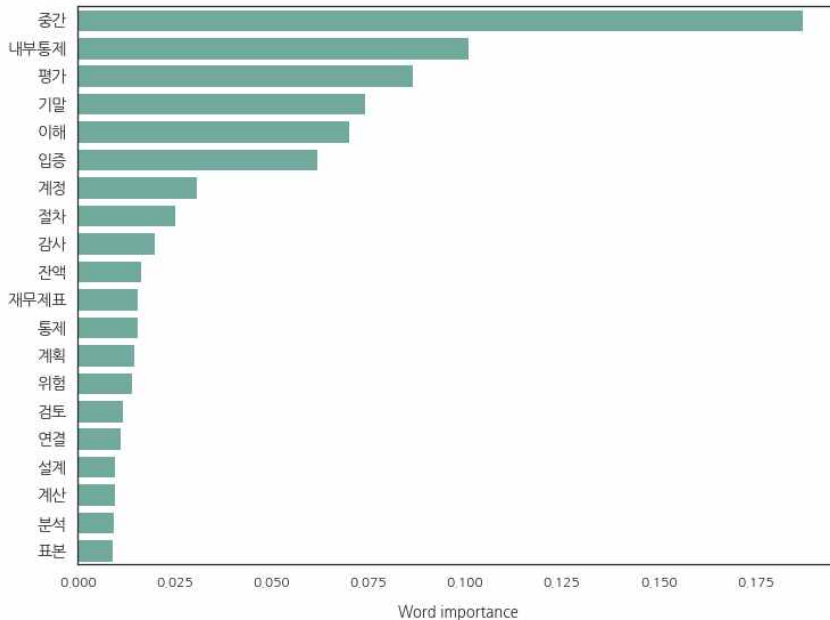


Figure 3. Top 20 important words for document classification

4.3 라벨 그룹별 연관규칙 추출

이번에는 각 그룹(라벨)별로 어떠한 단어들이 어떤 연관규칙을 가지고 나타나는지 확인하기 위하여 연관규칙 분석을 시행하였다. 빈발항목집합을 생성시 최소 지지도는 0.05, 최대 길이는 3으로 설정하였다. 연관규칙 생성시 최소 신뢰도는 0.5로 설정하였다. 그 결과 라벨 1 그룹은 총 2,789개, 라벨 2 그룹은 총 1,481개의 연관규칙이 발견되었다. 각 라벨 그룹별 향상도가 높은 상위 8개의 연관규칙은 Table 2와 같다.

Table 2. Top 8 association rules for each label group sorted by lift

Label	Antecedent	Consequent	Support	Confidence	Lift
1	(수행, 시사)	(추적)	0.063	1.000	16.000
	(추적)	(수행, 시사)	0.063	1.000	16.000
	(절차, 시사)	(추적)	0.063	1.000	16.000
	(추적)	(절차, 시사)	0.063	1.000	16.000
	(설계, 수행)	(추적)	0.060	1.000	16.000
	(설계, 시사)	(추적)	0.060	1.000	16.000
	(입증, 시사)	(추적)	0.060	1.000	16.000
	(시사, 공시)	(추적)	0.058	1.000	16.000
2	(테스트)	(분석, 수행)	0.056	0.838	7.541
	(분석, 수행)	(테스트)	0.056	0.500	7.541
	(표본, 문서)	(점검)	0.131	0.948	6.696
	(점검)	(표본, 문서)	0.131	0.924	6.696
	(점검)	(표본, 계산)	0.129	0.911	6.605
	(표본, 계산)	(점검)	0.129	0.935	6.605
	(재무제표, 분석)	(검토)	0.054	0.789	6.575
	(검증, 공시)	(점검)	0.115	0.928	6.551

예를 들어 라벨 1 그룹에서 ‘수행’, ‘시사’, ‘추적’이 동시에 발견될 확률은 0.063으로 작게 나왔지만 선행키워드 ‘수행’과 ‘시사’가 포함된 주요 감사실시내용에서 후행키워드 ‘추적’이 발견될 확률인 신뢰도는 1로 나타났다. 신뢰도가 1이라는 것은 ‘수행’과 ‘시사’가 포함된 주요 감사실시내용에서 후행키워드 ‘추적’이 반드시 나타났음을 의미한다. 또한 연관성을 의미하는 향상도의 경우 ‘수행’, ‘시사’라는 선행키워드와 ‘추적’이라는 후행키워드의 향상도는 16으로 동시에 발견될 확률이 ‘추적’ 혼자 발견될 확률에 비해 그 비율이 상당히 큼을 알 수 있다. 다시 말하면, 해당 연관규칙의 신뢰도가 ‘추적’ 혼자 발견될 확률의 16배에 달한다는 것으로 매우 연관성이 강하다고 볼 수 있다.

4.4 주요 감사실시내용 그룹별 핵심 어휘 추출

연관규칙 분석을 통해 추출된 각 라벨 그룹별 연관규칙들을 구성하는 단어들은 각 그룹을 대표하는 중요 단어로 고려될 수 있으나 이렇게 파악된 라벨 그룹별 중요 단어들은 모든 라벨 그룹에서 공통적으로 빈번히 사용되는 단어들이 존재하여 해당 단어가 각 그룹의 특징을 구분하는 핵심단어라고 받아들이기에는 무리가 있다. 따라서 4.2절의 문서분류에서 큰 영향력을 발휘했던 단어들을 4.3절의 라벨 그룹별 중요 단어들과 매핑하여 그룹별 특징을 반영한 핵심단어를 찾을 필요가 있다. Table 3은 문서분류에서 영향력이 큰 단어들이 각 그룹별 연관규칙을 대표하는 단어들과 어떻게 매핑 되었는지를 보여준다. 예를 들어, Rank 1에 해당하는 ‘중간’이라는 단어는 RF 모형의 변수 중요도에서 가장 높은 중요도를 가지는 단어로서 이가 라벨 1 그룹에서 추출한 연관규칙에는 나타났으나 라벨 2 그룹에서 추출한 연관규칙에는 나타나지 않았음을 의미한다. 타 라벨 그룹과 구별되는 라벨 1 그룹만의 핵심단어로는 ‘중간’, ‘내부통제’, ‘평가’, ‘이해’, ‘통제’, ‘계획’, ‘위험’, ‘설계’와 같은 단어들이 추출되었으며, 라벨 2 그룹에는 ‘기말’, ‘계정’, ‘잔액’, ‘재무제표’, ‘연결’, ‘계산’, ‘분석’, ‘표본’ 등이 연관규칙에 출현한 문서분별력이 높은 단어들로 파악되었다. 이러한 결과는 라벨 1과 라벨 2의 핵심단어가 모두 업무 단계의 특징을 잘 표현하고 있다는 점을 다음과 같이 알 수 있다. 위험평가 단계에서는 회사, 거래를 이해하고 내부통제의 효과성을 테스트한다. 그 결과 고유위험과 통제 위험을 평가하고, 입증감사절차를 계획하는 것이 위험-기반 접근법에서 위험평가의 목표이다. 한편, 입증감사절차 단계에서는 계정별 기말 잔액을 검증하기 위해 평가된 위험에 따라 분석적 절차, 계산 검증, 샘플링을 통한 증빙 확인 등의 감사증거 입수 절차를 수행한다. 이러한 점을 볼 때, 각 그룹별로 독특하게 사용되어 추출된 핵심단어들은 각 그룹의 특징을 잘 반영하고 있다고 볼 수 있다.

Table 3. Extracted group-wise keywords (*: Common words for multiple labels)

Rank	Keyword	Label 1	Label 2	Rank	Keyword	Label 1	Label 2
1	중간	O		11	재무제표		O
2	내부통제	O		12	통제	O	
3	평가	O		13	계획	O	
4	기말		O	14	위험	O	
5	이해	O		15	검토(*)	O	O
6	입증(*)	O	O	16	연결		O
7	계정		O	17	설계	O	
8	절차(*)	O	O	18	계산		O
9	감사(*)	O	O	19	분석		O
10	잔액		O	20	표본		O

반면에 ‘입증’, ‘절차’, ‘감사’, ‘검토’와 같은 단어들에 대해서는 두 라벨 그룹들의 연관규칙에서 동시에 나타나 각각의 단어가 각 라벨 그룹에서 어떤 다른 단어들과 연관되어 있는지 살펴볼 필요가 있다. Table 4는 분류기에서 영향력이 높은 단어들 중 라벨 1과 2 그룹의 연관규칙에서 동시에 발견된 단어들에 대한 각 라벨별 연관규칙을 향상도의 순으로 정렬했을 때 상위 5개의 연관규칙들을 보여준다.

Table 4. Association rules for common words

Common word	Label 1		Label 2	
	Antecedent	Consequent	Antecedent	Consequent
입증	[입증, 시사]	[추적]	[과목]	[계정, 입증]
	[추적]	[입증, 시사]	[과목]	[절차, 입증]
	[입증, 테스트]	[추적]	[입증, 과목]	[계정]
	[추적]	[입증, 테스트]	[입증, 잔액]	[계정]
	[입증, 일반]	[추적]	[절차, 입증]	[수행]
절차	[추적]	[절차, 시사]	[점검]	[절차, 문서]
	[절차, 시사]	[추적]	[절차, 문서]	[점검]
	[절차, 일반]	[추적]	[점검]	[절차, 검증]
	[추적]	[절차, 일반]	[절차, 검증]	[점검]
	[일반]	[절차, 시사]	[점검]	[절차, 계산]
검토	[평가, 검토]	[이해]	[재무제표, 분석]	[검토]
	[중간, 검토]	[감사]	[감사, 검토]	[재무제표]
	[입증, 검토]	[감사]	[감사, 분석]	[검토]
	[이해, 검토]	[평가]	[검토, 분석]	[재무제표]
	[평가, 검토]	[감사]	[검토]	[재무제표]
수행	[시사, 수행]	[추적]	[분석, 수행]	[테스트]
	[추적]	[시사, 수행]	[테스트]	[절차, 수행]
	[설계, 수행]	[추적]	[내용, 수행]	[공시]
	[계획, 수행]	[일반]	[수행, 공시]	[내용]
	[계획, 수행]	[추적]	[테스트]	[분석, 수행]

예를 들어 ‘입증’의 경우에는 라벨 1 그룹에서는 ‘시사’, ‘추적’, ‘테스트’, ‘일반’ 등의 단어들과 연관되어 사용된 반면, 라벨 2 그룹에서는 ‘과목’, ‘계정’, ‘절차’, ‘잔액’, ‘입증’ 등의 단어들과 함께 사용된 것을 알 수 있다. 이처럼 복수의 라벨 그룹들에서 동일하게 사용된 단어라고 할 지라도 그 쓰임새가 각 라벨 그룹별로 다르다는 것을 쉽게 확인할 수 있다. 공통 핵심어휘의 경우 위험-기반 접근법의 특징을 잘 보여준다. 회사, 거래를 이해하고 내부통제의 위험을 평가한 결과에 따라 감사절차의 성격, 범위를 계획하고, 입증감사절차를 수행하는 상호 연계성이 잘 드러난다. ‘입증’, ‘절차’, ‘검토’, ‘수행’ 등의 핵심어휘가 각각 단계의 분류에 영향을 크게 준 단어들과 같이 나타난다는 것으로, 단계별 상호 연계성이 잘 드러내고 있다고 판단할 수 있다.

6. 결론 및 시사점

본 연구에서는 문서 자료의 형태로 수집된 외부감사 실시내용의 주요 감사실시내용으로부터 기계학습 기법을 이용하여 감사계획, 중간감사, 입증감사절차와 같은 그룹별로 문서를 자동 분류하고, 그룹별 차이를 결정짓는 그룹별 핵심어

휘를 추출하기 위한 텍스트 마이닝과 연관규칙 분석에 기반한 체계적 방법론을 소개하였다. 본 방법론은 텍스트 마이닝을 통한 문서의 전처리(Preprocessing), 기계학습 기법을 사용한 문서의 그룹 분류, 그리고 이를 바탕으로 분류의 성능에 지배적인 영향을 끼치는 중요 어휘들을 추출해 내고 각 그룹별 문서들에 존재하는 연관규칙의 추출 결과와 비교하여 각 그룹별 특징을 대변하는 그룹별 최종 핵심어휘들을 추출하였다.

기존의 감사 관련 연구가 문헌 조사, 전문가 평가, 델파이(Delphi) 기법과 같은 정성적(Qualitative) 연구방법에 크게 의존했던 것에 비해 본 연구는 기계학습을 기반으로 외부감사 실시 내용상에 사용된 어휘들의 특징별(그룹별) 중요도값을 계량적으로 산출하여 정량적인 연구의 토대를 마련하였다. 이는 향후 자동화된 외부감사 실시내용의 주요 감사실시 내용 분류 결과를 수작업을 통해 선별적으로 재확인하거나 감사의 단계별 핵심어휘와 관련된 후속 연구를 정량적으로 수행 할 수 있는 기틀을 마련했다는 데 그 의의가 있다. 또한 본 연구의 결과 제시된 핵심 어휘를 활용할 때 외부감사인의 업무수행 내역 공시에 대한 투자자 등 정보이용자의 이해가능성이 향상될 것이며, 따라서 정보이용자가 외부 감사 자원이 어떻게 배분되었는지 정확히 판단할 수 있을 것이라는 점에서 외부감사 업무 공시 제도에 기여하게 될 것이다. 마지막으로, 본 연구에는 업무 단계별 자원 투입 내역이 정형화된 단계별 분류 하에 공시될 수 있을 때 그 유용성이 향상될 것이라는 공시제도 개선점을 파악하였다는 데 의의가 있으며, 규제 당국의 주목이 요구되는 바이다.

하지만 본 연구의 한계점으로는 초기 훈련용 데이터의 생성을 위해 전문가 집단에 의해 라벨링 작업을 선행하는 조건이 따른다는 데 있다. 이는 974건에 대한 초기 라벨링을 수행한 본 연구결과와 달리 더 많은 문서를 훈련용으로 생성한다면 문서 분류의 성능이나 그룹별 핵심어휘의 추출 결과가 달라질 수 있음을 의미한다. 향후에 비지도학습에 기반한 군집화 기법을 통해 라벨링을 자동 생성하고 이러한 결과가 실제 전문가 집단의 라벨링과 유사한 결과를 도출하기 위하여 어떤 추가적인 노력들을 해야 하는지에 대한 연구가 필요하다고 하겠다. .

REFERENCES

- Aggarwal, C. and Zhai, C. 2012. A Survey of Text Classification Algorithms, in Charu C. Aggarwal and ChengXiang Zhai (eds) Mining Text Data. Berlin/Heidelberg: Springer, pp.163-22.
- Alloghani, M., Dhiya A., Jamila M., Abir H., and Ahmed J. Aljaaf. 2019. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science' in Michael W. Berry, Azlinah Mohamed and Bee Wah Yap (eds) Supervised and Unsupervised Learning for Data Science, Cham. Switzerland: Springer, pp.3-21.
- Analyzer for Financial Information Utilization. The Journal of Information Systems 28(4):155-174.
- Blokdijk, H., F. Driehuisen, D. A. Simunic, and M.I T. Stein. 2006. An Analysis of Cross-Sectional Differences in Big and Non-Big Public Accounting Firms' Audit Programs. Auditing: A Journal of Practice & Theory 25(1): 27-48.
- Boskou, G., Kirkos, E., and Spathis, C. 2018. Assessing internal audit with text mining. Journal of Information & Knowledge Management 17(02):1850020.
- Breiman, L. 2001. Random Forests, Machine Learning 45(1):5-32.
- Byrnes, P. E., Al-Awadhi, A., Gullvist, B., Brown-Liburud, H., Teeter, R., Warren, J.D., and Vasarhelyi, M., 2018. Evolution of Auditing: From the Traditional Approach to the Future Audit1. In Continuous auditing. Emerald Publishing Limited.
- Chan, D. Y. and M. A. Vasarhelyi. 2018. Innovation and practice of continuous auditing. In Continuous Auditing: Theory and Application Emerald Publishing Limited: 271-283.
- Cho, S., M. A. Vasarhelyi, T. Sun, C. Zhang. 2020. Learning from Machine Learning in Accounting and Assurance.

Journal of Emerging Technologies in Accounting 17(1):1-10.

- Goltz, N. and Mayo, M., 2018. Enhancing regulatory compliance by using artificial intelligence text mining to identify penalty clauses in legislation. *RAIL*, 1, p.175.
- Gray, G. L. and R. S., Debreceny. 2014. A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems* 15(4): 357-380.
- Hastie T, Tibshirani R, and Friedman J. 2001. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. Spring.
- Heo, B. K. and Jung, Y. K. 2013. The Effects of Data Mining Ensemble Techniques on Audit Risk Reduction. *Korean Management Review* 42(5):1523-1559.
- Jung, G.Y., Yoon, S.S., and Kang, J.Y., 2019. Development of Text Mining-Based Accounting Terminology.
- Kamaruddin, S.S., Bakar, A.A., Hamdan, A.R., Nor, F.M., Nazri, M.Z.A., Othman, Z.A., and Hussein, G.S., 2015. A text mining system for deviation detection in financial documents. *Intelligent Data Analysis* 19(s1):S19-S44.
- Kim, D. Y. 2019. A Study on the Effects of Human Resource in KEJI Internal Accounting Control System on Audit Times. *Korean Journal of Business Administration* 32(6):1087-1107.
- Kim, K.S. and Cho, N.W. 2021. A Study on Networks of Defense Science and Technology using Patent Mining. *Journal of the Korean Society for Quality Management* 49(1):97-112.
- Kim, S., Park, E., Cho, H., Hong, S., Sohn, B., and Hong, J. 2021. Pattern Analysis of Nonconforming Farmers in Residual Pesticides using Exploratory Data Analysis and Association Rule Analysis. *Journal of the Korean Society for Quality Management* 49(1):81-95.
- Kwon, S.Y., Jung, K.C., and Yun, Y.S. 2016. The Effects of Audit Planning and Interim Audit on Audit Hours and Audit Fees. *Study on Accounting, Taxation & Auditing* 58(3):137-172.
- Na, H. J., Lee, K. C., Choi, S. U., and Kim, S. T. 2019. 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* 44(4):175-214.
- Siroky, D. S. 2009, Navigating and Random Forest and Related Advances in Algorithmic Modeling, *Statistics Survey*, 3, 147-163.
- Strobl, C. and James, M. and Gerhard, T. 2009. An Introduction to the Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging and Random Forests, *Psychological Methods*. 14(4), Dec.:323-3.

저자소개

- 성윤석** 현재 동국대학교-서울캠퍼스 경영대학 회계학과 석박사통합과정 재학 중이며, 선진회계법인 품질관리업무 담당이사로 재직 중이다. 주요 연구관심분야는 재무회계와 회계감사이다.
- 이동희** 현재 동국대학교-서울캠퍼스 경영대학 경영학과 학부과정에 재학 중이다. 주요 연구관심분야는 품질경영, 기계학습, 운영관리이다.
- 정 욱** 현재 동국대학교-서울캠퍼스 경영대학 정교수로 재직 중이다. Georgia Institute of Technology에서 산업시스템공학 박사 학위를 취득하였다. 이후 삼성SDS SCM사업단에서 SCM관련 컨설턴트로 재직하였다. 주요 연구관심분야는 품질경영, 데이터마이닝, 공급사슬관리, 기술경영 등이다.