

레그테크 기반의 자본시장 규제 해석 온톨로지 및 딥러닝 기술 개발을 위한 제언*

최승욱** · 권오병***

<목 차>

I. 서론	3.4 딥러닝 분석의 활용
II. 이론적 배경	IV. 토의 및 결론
III. 레그테크 기반의 온톨로지 및 딥러닝 기 술 개발을 위한 사례	4.1 토의
3.1 전체적인 흐름	4.2 결론
3.2 형태소 분석	참고문헌
3.3 온톨로지 요소 추출	<Abstract>

I. 서론

전세계적으로 Digital Transformation(DT)을 통해 조직과 경제 시스템을 혁신하려는 노력은 점차 증가하고 있다. DT는 조직의 비즈니스 방식, 제품과 서비스, 전략의 근본적 변화를 가져온다. 이에 기업들도 DT에 적극적으로 대응하지 않으면 다가오는 신경제에 적응하지 못할 가능성이 높다. 이러한 환경 변화에 따라 최근 인공지능과 빅데이터 기술의 발전을 기반으로 규제비용을 줄이고 효율적인 감독을 가능하게

만들기 위한 시스템의 하나로 레그테크(RegTech)가 등장하였다. 레그테크는 규제(Regulation)와 기술(Technology)의 합성어로 규제를 인공지능 기술을 통해 기계어로 전환하여 해석 및 적용하는 개념을 의미한다. 레그테크는 최근 핀테크의 중요한 분야 중 하나로 부상 중이다(우가인, 2017). 그러나 아직 핀테크 연구는 전자결제 서비스 등의 사용의도나 지속 사용의도에 머물러 있어(김은정 등, 2017; 정대현 등, 2017; 김민중, 손달호, 2020) 레그테크로의 연구분야 확장이 절실하다.

* 이 논문은 2020년도 경희대학교 [2020년 대학혁신지원사업]의 지원을 받아 수행된 연구임.
(KHU-20201407)

** 경희대학교 회계·세무학과, suchoi@khu.ac.kr (주저자)

*** 경희대학교 경영학과, obkwon@khu.ac.kr (교신저자)

레그테크 기술이 발전하게 된 데에는 두 가지의 상황적·기술적 패러다임의 변화가 영향을 미쳤다. 우선, 2008년에 발생한 금융위기는 금융규제를 더욱 강화하는 촉진제 역할을 하였다. 규제의 강화는 다양한 규제비용을 확대하고, 이로 인해 감독기관과 금융기관은 더욱 강력한 규제환경에 직면하게 되었다. 이러한 환경은 감독기관이나 금융기관 모두에게 상당한 부담으로 작용한다. 예컨대, 감독기관 관점에서는 규정의 적합성과 적용 가능성을 면밀히 살펴봐야 하고, 타 법안과의 저촉 가능성, 규정을 준수하도록 유도하는 등의 강도 높은 업무 부담에 직면한다. 이를 적용하는 자본시장의 기업들 역시 규제비용이 천문학적으로 소요되며, 규정숙지를 못하게 될 경우 막대한 손실을 입는다(Kwon et al. 2020). 예를 들어, Boston Consulting Group의 추정에 따르면 2009년에서 2016년 동안 미국과 유럽의 규제불이행 은행들에 부과된 벌금의 규모는 무려 3,210억 달러에 이르는 것으로 나타났다.

레그테크 기술의 발전을 촉진한 두 번째 변화는 과학기술과 금융의 융합을 가져온 핀테크(Fintech)의 발전이다. 금융위기 이후 저성장과 저금리 기조가 지속되어 금융기관의 영업이익이 급감하였다. 이에 기존 금융 상품과 서비스의 한계를 절감하고 새로운 성장 동력으로 핀테크를 주목하였다(UK Government Office of Science, 2015). 과거에도 과학과 금융은 아날로그 기술부터 함께 해 왔지만 이는 주로 금융회사의 이익을 위한 방향이었다. 그러나 핀테크는 간편결제, P2P, 크라우드펀딩 등과 같은 대고객 접점이 크고 생활의 편리함을 높이거나 스타트업의 자금조달을 지원하기 위한 수단 또

는 기존 금융 인프라가 성숙하지 못한 지역에서의 대안금융이 되고 있다(FCA 2019). 요약하면, 금융위기로 인한 규제비용의 증가와 핀테크의 등장은 레그테크가 발전하는 중요한 동인이 되었다.

본 연구는 국내에 적용되기 시작한 레그테크 기술의 현황과 과제를 짚어보고, 머신리더블 규제(Machine Readable Regulation)를 탐색하여 기술 개발에 대한 기초 연구자료를 제시하는 것을 목적으로 한다. 현재 영어를 기반으로 하는 온톨로지 및 딥러닝 기술은 상당히 개발되어 있으며, 독일어나 스페인어 등 문법적으로 영어와 유사성이 높은 유럽이나 중남미 계열 언어로 확장하는 것은 어렵지 않을 것이다. 그러나 문법이 상이한 한국어 등 대부분의 아시아 언어권에서는 이를 그대로 활용하는 것이 쉽지 않다. 따라서 한국어 기반의 국내 법령을 인공지능으로 하여금 인지하고 해석하도록 하는 기술의 개발은 쉽지 않다. 또한 금융당국 역시 레그테크 기반의 보고 체계에 대해 익숙해져야 할 필요가 있다. 초기에는 모든 금융기관이 이러한 기계 중심 보고 체계에 익숙하지 않을 것이므로, 감독기관과 금융기관 외에도 보고 체계를 건설당하고, 업무를 지원할 수 있는 기관이 설립되는 등의 보고 생태계 구축이 중요하다. 본 연구는 이러한 생태계 구축의 필요성을 제시한다.

본 연구는 구체적인 사례로 전자금융감독규정의 일부를 활용해 해당 규제문장을 기계어로 전환하고 형태소 분석 등을 통해 정형화된 데이터로 변환하는 과정을 제시하였다. 형태소로 분리된 각 문장은 전처리 과정을 통해 메타언어로 변환되어 문형분석을 받는데, 이때 문형

분석이 이루어질 수 있도록 구축된 문형 사전을 활용한다. 다음 단계에서는 규제문장이 있는지 판정하기 위해 형태소 분석과정과는 별개로 딥러닝 분석 과정이 수행되며, 그 진행 과정은 분석 대상 문장으로 학습데이터 셋을 구축하여 딥러닝 알고리즘으로 규제문장 여부를 판별하는 모델을 생성하는 등의 절차가 이용된다. 최종적으로 형태소 분석 결과와 딥러닝 수행 결과를 결합하여 그 의미를 산출한다. 본 연구는 이러한 일련의 과정을 통해 레그테크 기반의 온톨로지 및 딥러닝 기술의 적용가능성을 사례를 통해 제시한다.

이어지는 구성은 다음과 같다. 2장에서는 레그테크의 이론적 배경과 관련 선행연구를 제시한다. 3장에서는 자본시장 법령을 이용하여 레그테크 적용 사례를 제시한다. 4장은 토의와 결론을 통해 연구를 맺는다.

II. 이론적 배경

전술한 바와 같이 DT를 통해 조직과 경제 시스템을 혁신하려는 노력은 점차 증가하고 있다. 기업들도 DT에 적극적으로 대응하지 않으면 이미 눈앞에 현실화된 신경제와 새로운 형태의 산업혁명에 적응하지 못할 가능성이 높다. 이러한 환경 변화에 레그테크는 규제에 대응해야 하는 모든 기업들에 대한 DT의 한 형태로 해석될 수 있을 것이다. 그러나 규제를 적용하기 위한 업무에는 막대한 비용이 발생하며 그 가장 주된 이유는 비구조적이고 적시성이 떨어지거나 제출해야 하는 데이터에 대한 정의가 불명확하기 때문이다. 따라서 적시에 정확한 데이터

를 보고하도록 하는 시스템이 필요하다.

이러한 의미에서 레그테크는 “각종 규제에 대하여 기존 방식보다 더 효율적, 효과적으로 대응할 수 있도록 기술의 활용에 초점을 맞춘 핀테크의 한 하위 분야(FCA 2016)”로 정의된다. 즉, 자료 수집 기술부터 복잡한 의사결정 자동화 및 보고서 주체에 대한 행동 분석에 이르기까지 매우 폭넓은 기능을 포괄한다(Fröhlich 2019). 산업이 보다 더 디지털화되고 표준화되면서 금융규제와 그 적용에 관한 데이터는 더욱 대량으로 수집될 것이고, 레그테크는 이 데이터들을 보다 더 유용하고 효율적으로 활용할 수 있는 정보기술의 역할을 하게 될 것이다. 여기서의 유용함이란 규제 보고를 보다 더 정확하고 신속하게, 또한 규제 법령의 변경이 일어났을 때 이에 대한 대응을 좀 더 원활히 할 수 있는 가치를 의미한다(Alamaki and Broby 2019). 이러한 레그테크는 크게 네 개의 키워드로 요약된다. 이는 각각 복잡하게 결합된 데이터를 분리 및 추출하는 민첩성(Agility), 규제에 따른 보고를 빠르게 수행하는 속도(Speed), 짧은 시간 내에 솔루션을 찾아내는 통합(Integration), 빅데이터를 분석하는 분석기법(Analytics)이다(Deloitte 2016).

지난 2008년의 금융위기는 자본시장의 규제 강화를 야기하였고, 이러한 규제를 효율적으로 적용하는 방법론적 논의가 지속되어 왔다. 전술한 바와 같이 이러한 환경적 변화는 핀테크의 발전과 맞물려 레그테크가 등장하는 중요한 요인이 되었다. 구체적으로, 레그테크는 핀테크의 일종으로 금융과 기술의 결합을 의미하는 핀테크를 기업중심적인 관점에서 적용하는 것을 의미한다. 예컨대, 핀테크는 개인 고객 중심 금융

기술에서의 편의를 제공하는 것이라면, 레그테크는 금융 규제를 적용받는 회사의 관점에서 인공지능을 활용하여 규제에 대한 적응력을 높이는 기술을 의미한다(조창훈 2017). 즉, 레그테크는 복잡한 규제를 적용하는데 발생하는 비용을 획기적으로 줄여 효율을 높인다는 점에서 금융, 회계감사, 위기 관리(risk management) 분야에서 적용될 것으로 기대된다(Anagnostopoulos 2018). 이러한 레그테크의 적용은 그동안 감사법인 혹은 법무법인이 해왔던 금융규제의 검토와 적용, 감사와 위험관리 등의 업무를 규제를 적용받는 기업이 내부적으로 구현할 수 있다는 점에서 비용 효율을 달성할 것으로 기대된다.

이러한 레그테크와 관련된 선행연구는 많지는 않다. 예외적으로, Colaert(2018)는 레그테크의 실행으로 인한 명암이 존재하며, 장점으로 효율적 향상, 개인에 의해 발생될 수 있는 위험 감소, 감독책임의 감소 등이 존재한다고 주장한다. 반면, 단점으로는 새로운 유형에 따른 시스템리스크의 증가, 높은 비용, 데이터 보호의 문제, 비인간화의 부작용 등을 언급하였다. Yang and Li(2018)는 데이터 모니터링에 중점을 둔 기술 중심의 규정은 기존 규정의 비효율성에 대한 해결책이 될 수 있으며, 금융 소비자의 권리 및 이익에 대한 효과적인 보호를 강화할 수 있다고 주장한 바 있다. 또한 새로운 시스템 위험이라고 볼 수 있는 핀테크 위험을 레그테크 기술로 인해 효과적으로 제어할 수 있다고 주장한 바 있다. 여기에서 핀테크 위험이란, 급속도로 발전한 핀테크를 통한 혁신이 오히려 규제기관이 민첩하게 이를 파악하기 전에 규제를 역행하는 탈법적 금융행위를 야기하

는 위험, 신기술을 잘못 적용하여 발생하는 기술적 위험, 규제기관이 규제를 받는 기업보다 정보 열위에 있어 발생하는 정보비대칭의 위험 등을 포괄한다(Yang and Li 2018). 이상과 같이 레그테크 관련 연구는 비교적 최근에 이루어지고 있으며, 주로 레그테크 적용의 효익이나 나아가야 할 방향성에 대한 문헌들이 대부분이다.

레그테크를 구성하는 주된 정보기술로는 인공지능(Huang et al. 2014; Omar et al. 2017), 기계학습(Chen 2016; Gray et al. 2014), 자연어 처리(Butler and O'Brien 2019), 빅데이터 분석 기술(Yoost 2016), 분산원장 기술(Kavassalis et al. 2018), 그리고 블록체인 기술(Colgren 2018; Peters 2018) 등이 있다. 특히 레그테크에서의 인공지능은 주로 자연어의 형태로 존재하는 규제 문장에서 규칙을 탐지 및 추출하여 활용하는데 관심이 있다(Fröhlich, 2019). 특히 규제 문장을 기계학습이 가능한 문장으로 인지하는 기술이 중요하다. 이러한 머신러닝 규제를 인지하는 것은 규제 개념 요소에 대한 명료화(unambiguity)에 그 핵심이 있다. 감독기관이 컴퓨터 등 기계가 그 규제 내용을 스스로 인식하고 규제를 자동적으로 이행할 수 있도록 하는 머신러닝 규제를 제공하게 되면, 금융기관 등 시장참여자들은 규제기관이 제공해 준 것을 토대로 컴퓨터 등에 입력함으로써 손쉽게 규제를 이행하고 자동적으로 그 보고서가 규제당국에 보고됨으로써 규제준수(regulatory compliance)를 달성하게 된다(유제민 2019).

이상과 같이 레그테크의 필요성에 대해서는 실무와 학계에서 논의를 진행하고 있다. 다만, 한국은 경제규모나 기술발전 수준을 고려할 때

이에 미온적이었으며, 특히 자본시장의 규제에 대한 레그테크 적용은 아직 미흡한 상황이다 (Kwon et al. 2020).

Ⅲ. 레그테크 기반의 온톨로지 및 딥러닝 기술 개발을 위한 사례

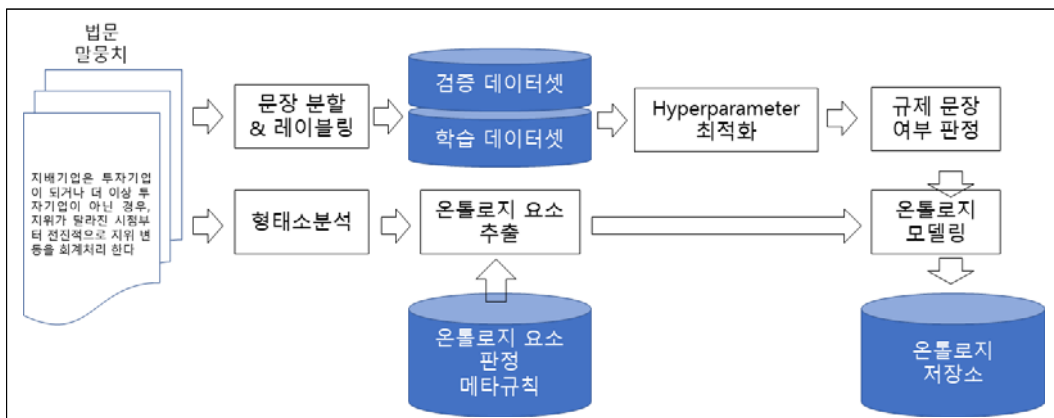
3.1 전체적인 흐름

본 연구는 레그테크 기반의 온톨로지 및 딥러닝 기술 개발 방법을 제안하는 것을 목적으로 한다. 한국어 자연어처리는 최근 BERT를 이용한 한국어 의미역 추출하기(배장성 등 2020)나 자동 띄어쓰기(황태욱, 2019), 딥러닝을 활용한 품사 태깅하기(김선호 2018) 등 기계 학습 방법을 활용한 연구가 활발해지고 있다. 그러나 아직 법문에 대해서 이러한 한국어 자연어처리 기법을 적용한 사례는 많지 않다.

본 장에서는 이를 위한 사례를 다음과 같이 제시한다. 우선, 첫 번째 단계는 법문으로 기계어로 전환이 가능한 요소를 추출하는 것이다.

이를 위해 법문을 기반으로 온톨로지 요소를 추출하는 방법을 다음의 <그림 1>과 같이 요약한다. 먼저 법문 말뭉치를 수집한 후에 이 중에서 일부를 선택하여 문장 단위로 분할하고, 여기에 전문가에 의하여 규제 규칙(rule)이 포함되어 있는지의 여부를 YES/NO로 하여 레이블링을 한다. 그리고 이를 학습용 및 검증용 데이터셋으로 임의의 분배한다. 그런 후에 여러 개의 대안이 되는 딥러닝 모델(DNN, CNN, LSTM 등)에 대해서 추론 성능이 극대화되는 방향으로 하이퍼파라미터(hyperparameter) 값을 최적화한다.

한편, 법문말뭉치의 일부에 대해서 형태소분석을 실시한 후에 사전에 제안하는 온톨로지 요소 판정을 위한 메타규칙을 적용하여 형태로 분석된 각 요소에 대해 온톨로지 관점에서의 유형을 정한다. 그러면 이 유형화된 요소 및 형태소분석의 대상이 된 문장에 대해 규제문장 여부 딥러닝을 수행하여 규제문장 존재 유무를 추론한 후에 그 결과(유형화된 요소, 해당 문장에 규제문장 존재 유무 값)를 온톨로지 저장소에 저장한다. 추후에 저장된 값은 온톨로지 작



<그림 1> 연구범위

성도구(예: OntoEdit, Protege 등)를 활용하여 온톨로지 형태의 머신리더블 규제(온톨로지화된 규제법령)로 변형 생성되며, 이를 활용한 인공지능 시스템이 법률 서비스에 활용하게 된다.

3.2 형태소 분석

상기한 방법에서 가장 먼저 이루어지는 형태소 분석 과정은 분석 대상 입력, 분석 대상 법령의 형태소 분석, Class와 Property의 인식으로 나뉠 것이다. 다음으로, 적용하고자 하는 규제의 문장을 입력한다. 우선, 분석하고자 하는 법인세법 도메인의 규제 문장은 <그림 2>와 같다.

우선, 입력된 규제 문장에 대하여 형태소 분석을 수행한다. <그림 3>에서와 같이 입력 텍

스트는 먼저 띄어쓰기 단위의 어절 단위로 분리되고, 어절 단위의 내용 각각에 대하여 형태소 분석이 이루어진다. 이때 형태소는 항상 어절 단위와 같거나 그보다 작으므로 먼저 어절 단위로 분리해 형태소 분석을 효율적으로 수행할 수 있다. 다음으로 언어 단위에서 가장 작은 단위의 형태소 단위로 나누어 반복적인 언어 단위가 가능한 한 많이 발견되게 하다. 이때 반복적인 단위가 많이 발견될수록 두 문장이 실질적으로 같은 내용의 문장인지 여부가 쉽게 파악될 것이다. 최소 의미 단위로 분리된 형태소는 반복된 단위로서 이후 정형데이터로 바뀔 수 있으며, 태깅된 품사는 메타언어로 변환할 때와 문형 분석 시 대상 요소 여부를 알게 해주는 정보를 제공한다.



<그림 2> 규제문장 사례

법령	어절	형태소
제13조(과세표준) 1항 내국법인의 각 사업연도의 소득에 대한 법인세의 과세표준은 각 사업연도의 소득의 범위에서 다음 각 호의 금액과 소득을 차례로 공제한 금액으로 한다. 다만, 제1호의 금액에 대한 공제는 각 사업연도 소득의 100분의 60〔조세특례제한법〕 제6조 제1항에 따른 중소기업(이하 “중소기업”이라 한다)과 회생계획을 이행 중인 기업 등 대통령령으로 정하는 법인의 경우는 100분의 100]을 한도로 한다.	제13조 (과세표준) 1항 내국법인의 각 사업연도의 소득에 대한 법인세의 과세표준은	제/XPN + 13/SN + 조/NNB (/SS 과세/NNG + 표준/NNG)SS 1/SN + 항/NNB ▶ 내국/NNG + 법인/NNB + 의/VCP 각/NNG 사업/NNG + 연도/NNG + 의/VCP 소득/NNG + 에/VCP 대하/VV + ㄴ/ETM 법인세/NNG + 의/VCP 과세표준/NNG + 은/VCP

<그림 3> 분석 대상 규제 문장의 형태소 분석

3.3 온톨로지 요소 추출

다음으로, 형태소단 위로 분리된 데이터에서 온톨로지의 Class와 Property를 인식한다. 인식의 기본 원칙은 다음과 같다.

(1) Class와 ObjectProperty는 각각 주어와 목적어, 그리고 주어와 목적어를 잇는 동사인 경우가 대부분임. 또한 능동태의 경우에는 앞의 Class가 domain, 뒤의 Class가 range 가 됨.

(2) DataProperty는 ‘(Class)의 (DataProperty)’와 같이 Class의 주변에 많이 나타남. 따라서 Class와 ObjectProperty를 먼저 파악하고 난 후에 Class list를 참고해 가면서 ‘인근(adjacency)’의 Class가 아닌 명사들을 찾음.

특히 한국어 문법적으로 Class와 Property가 등장할 수 있는 패턴들이 있는데 이들을 모아놓은 것을 메타 규칙 저장소라고 한다. 머신러닝을 규제 요소 추출 위한 메타규칙의 예는 <그림4>와 같다. 메타규칙을 활용하는 궁극적인 목적은 문형 사전에 기록된 어휘의 위치정보를 이용하여 온톨로지서 활용된 유효한 어휘를 추출하고자 하는 데 있다. 유효 어휘란 온

톨로지를 구성하는 요소로서, Class와 Property 요소이다. 메타규칙은 각 문형 조건을 제시하고, 각 조건에서 Class Domain, Class Range, Object Property, Data Property가 어느 위치에 있는지를 명시한 사전을 말한다.

위의 <그림 4>에서 첫 번째 문형 조건의 의미는 다음과 같다. 먼저, 문형 조건은 ‘ANY NOUN1이(가/은/는), ANY NOUN2를(을) VERB1’으로서 ANY 조건으로 시작하고 있다. ANY 조건은 어떠한 표현이 와도 뒤에 기록된 내용이 오기까지는 PASS해야 한다는 것을 말한다. 따라서 이 문형 조건은 앞에 어떤 단어를 올 수 있으며, 명사로 시작하여(NOUN1) 조건문이 시작됨을 말하고 있다.

그 다음으로는 조사가 오는데 조사로는 -이, -가, -은, -는 중 어느 것이 와도 동일하게 처리하는 것이다. 그 다음에 또다시 ANY 규칙이 있어 어떠한 어휘가 와도 넘기며(PASS), 다시 두 번째 명사(NOUN2)가 와야 한다. 그리고 목적격 조사 ‘-를’ 또는 ‘-을’이 오고, 마지막으로 동사 ‘하다’가 와야 함을 말하고 있다. 한편, 우측에서는 온톨로지 구성요소가 문형 조건의 어

편 위치에 속하는지를 명시하고 있다. 예를 들어, Class Domain은 NOUN2 자리에 있는 것임을, Class Range는 NOUN1 자리에 있는 것임을, Object Property는 VERB1 자리에 있는 것임을 의미한다.

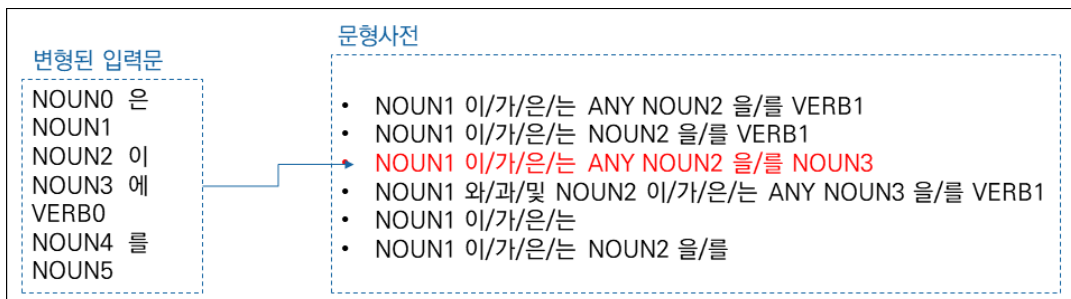
메타언어로 변환된 문장에 대하여 문형조건 사전을 탐색하여 문형 조건에 해당하는지, 해당한다면 어떠한 문형 조건에 해당하는지를 검색한다. 입력문은 메타언어 변형 결과 ‘NOUN0 은 NOUN1 NOUN2 이 NOUN3 에 VERB0 NOUN4를 NOUN5’가 되었으며, 이 메타언어

에 맞는 문형을 탐색하면 ‘NOUN1 이/가/은/는 ANY NOUN2 을/를 NOUN3’가 발견되는 것이다. <그림 5>는 이러한 온톨로지 요소를 추출하는 과정을 보이고 있다.

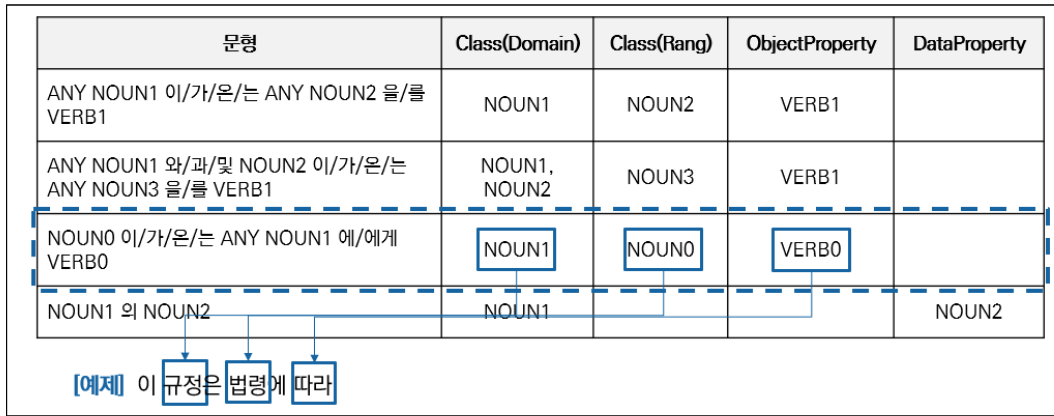
단, 입력문의 메타언어에서 조건문의 ANY에 해당하는 부분이 있는데 이는 온톨로지 구성 요소 찾기에는 도움이 되지 않으므로 이 부분은 삭제한다. 예를 들어, ‘내국법인의 각 사업 연도의 소득’이라는 같은 문장이 입력되었을 때, ‘내국법인’부터 ‘소득’까지는 ANY에 해당하는 부분이므로 삭제한다.

문형	Class(Domain)	Class(Range)	ObjectProperty	DataProperty
ANY NOUN1 이(가/은/는) ANY NOUN2 를(을) VERB1	NOUN1	NOUN2	VERB1	
ANY NOUN1 와(과/및) NOUN2 이(가/은/는) ANY NOUN3 를(을) VERB1	NOUN1, NOUN2	NOUN3	VERB1	
NOUN1 은(는/이/가) ANY NOUN2 를(을) 하 NOUN1 의 NOUN2	NOUN1 NOUN1		NOUN2	NOUN2
NOUN1 은 NOUN2 이 ANY NOUN3 에 VERB0 지	NOUN1	NOUN2	NOUN3	
NOUN1 와 VERB1 L NOUN2	NOUN1			NOUN2
ANY NOUN1 이(가/은/는) ANY NOUN2 예(예게) VERB1	NOUN2	NOUN1	VERB1	
ANY NOUN1 와(과/및) NOUN2 이(가/은/는) ANY NOUN3 예(예게) VERB1	NOUN2	NOUN1	VERB1	
NOUN1 은(는) ANY NOUN2 를(을) ANY NOUN3 예(예게) VERB1	NOUN3	NOUN1	VERB1	
NOUN1 이(가) NOUN2 예 하는 NOUN3 NOUN1 와(과/및) NOUN2 의 NOUN3	NOUN1 NOUN1, NOUN2	NOUN2	NOUN3	NOUN3
NOUN1 은(는) NOUN2 로 하여금 ANY NOUN3 를(을)	NOUN1	NOUN2	NOUN3	
NOUN1 은(는) NOUN2 예게 ANY NOUN3 를(을)	NOUN1	NOUN2	NOUN3	
NOUN1 은(는) NOUN2 로 하여금	NOUN1	NOUN2		
NOUN1 은(는) NOUN2 예게	NOUN1	NOUN2		
NOUN1 의 NOUN2 및 NOUN3	NOUN1			NOUN2, NOUN3
NOUN1 는(은/이/가) ANY NOUN2 예 VERB1 아	NOUN1	NOUN2	VERB1	
NOUN1 는(은/이/가) ANY NOUN2 예 VERB1 ㄹ	NOUN1	NOUN2	VERB1	
NOUN1 는(은/이/가) ANY NOUN2 예 VERB1 L	NOUN1	NOUN2	VERB1	
----	----	----	----	----
NOUN1 은(는/이/가) NOUN2 를(을) NOUN3 예(예게)	NOUN1	NOUN3	NOUN2	

<그림 4> 머신리더블 규제 요소 추출 위한 메타규칙



<그림 5> 온톨로지 요소 추출 과정



<그림 6> 온톨로지 요소 추출 과정 예

이러한 과정을 통해 최종적으로 남겨진 문장에서 Class와 Property를 추출한다. 이때 현재 남겨진 문장은 규칙에 해당하는 부분만 남겨진 것이므로, 최초 문형 조건에서 Class, Property가 있는 것으로 지정된 부분을 찾아 해당 어휘를 가져오는 작업이다. 아래의 예는 최종 남은 문장이 ‘이 규정은 법령에 따라’이고 해당 문형은 ‘ANY NOUN1 이/가/은/는 ANY NOUN1 예/에게 VERB0’이다. Class Domain에 속하는 것은 NOUN1, Class Range에 속하는 것은 NOUN0, Object Property에 속하는 것은 VERB0이므로, 이에 해당하는 어휘 ‘규정’, ‘법령’, ‘따르다’를 각각 추출하게 된다. 이상의 온톨로지 요소 추출 과정의 예는 <그림 6>과 같다.

3.4 딥러닝 분석의 활용

법령을 온톨로지로 변경함에 있어서 다음의 단계는 법문 중에서 규칙으로 인식할 문장이 있는지를 판정하는 것이다. 이를 위해 형태소

분석 및 온톨로지 요소 추출 과정과는 별개로 지도기반 학습을 위한 딥러닝 분석 과정을 수행한다.

먼저 대상 법령은 추론에 대한 평가를 위해 사용되어야하므로, 지도기반 학습을 위한 학습용 데이터셋은 금융감독에 관련한 법령을 대상으로 하였다. 학습데이터 SET으로 이용한 법령은 아래와 같다.

- 금융규제 운영규정
- 전자금융감독규정
- 전자금융감독규정 시행세칙
- 전자금융거래법 시행령
- 전자금융거래법

위 법령의 각 문장에 대해 전문가에 의하여 규칙이 존재한다고 판단되면 YES, 그렇지 않으면 NO로 레이블링을 하였다. 본 연구에서 이용한 규제문장 중 학습데이터 SET은 전체 6,429건 중 75%인 4,821 문장이고 나머지는 25%의 문장은 훈련된 모델의 정확도를 확인하는데 사용되었다. 학습데이터 SET의 예시는 다음 <그림 7>과 같다.

A	B	C
1 id	분문	is_rule
2 10	제1장 총칙	0
3 11	제1조(목적) 이 혼령은 금융위원회 및 금융유관기관이 금융회사등에 대한 금융규제	0
4 12	제2조(정의) 이 혼령에서 사용하는 용어의 뜻은 다음 각 호와 같다	0
5 14	금융회사등이란 다음 각 호의 자를 말한다	0
6 16	「은행법」에 따른 인가를 받아 설립된 은행(같은 법 제59조에 따라 은행으로 보는 자	0
7 18	「자본시장과 금융투자업에 관한 법률」에 따른 금융투자업자	0
8 20	「보험업법」에 따른 보험회사(같은 법 제4조제6항에 따라 보험회사로 보는 자를 포	0
9 22	「상호저축은행법」에 따른 상호저축은행	0
10 24	「신용협동조합법」에 따른 신용협동조합	0
11 25	바	0
12 26	「여신전문금융업법」에 따른 여신전문금융회사	0
13 27	사	0
14 28	「금융지주회사법」에 따른 금융지주회사	0
15 29	아	0
16 30	「농업협동조합법」에 따른 조합, 농업은행, 농업생명보험 및 농업신협보험	0
17 167	제9조(의견청취) 금융위원회 및 금융감독원은 새로 금융회생지도를 하려는 경우 또	1
18 301	② 제1항에 따른 유효기간은 금융회생지도의 목적을 달성하기 위하여 필요한 최소	1
19 315	④ 금융회생지도 유효기간의 연장은 1회에 한정한다	1
20 316	다만, 다음 각 호의 어느 하나에 해당하는 경우에는 금융회생지도 유효기간을 2회	0
21 325	④ 금융위원회 또는 금융감독원이 제3항 단서에 따라 금융회생지도의 유효기간을 2	1
22 326	이 경우 금융위원회 또는 금융감독원은 해당 금융회생지도 유효기간 연장의 필요성	1
23 344	이 경우 금융위원회 규제개혁법무담당관은 소관부서의 검토의견을 들어야 하며, 소	1
24 438	제21조(교육) ① 금융위원회 및 금융감독원은 소속 공무원 또는 임직원이 이 혼령을	1
25 535	제5조(전자금융사고 책임이행)을 위한 보험 등의 가입에 관한 기준) ① 금융회사 또는	1
26 578	④ 금융회사 또는 전자금융업자가 전자금융사고 책임이행을 위한 준비금을 적립하	1
27 586	④ 금융회사 또는 전자금융업자가 보험 또는 공제 가입과 준비금 적립을 병행하는	1
28 699	② 제2항 각 호의 사항을 이행하지 못하는 금융회사 또는 전자금융업자는 그 사유	1
29 700	다만, 허가, 등록 또는 인가를 마친 후 1년이 지나지 않은 금융회사 또는 전자금융업	1

<그림 7> 학습데이터 SET 예시

다음으로, 학습 데이터의 문장에 대하여 형태소 분석을 수행하였다. 이는 하나의 문장을 변형 없이 입력값으로 하여 학습할 경우, 문장 전체가 하나의 종류가 되기 때문에 패턴 학습이 올바르게 이루어지지 않기 때문이다. 따라서 하나의 문장에 사용된 어휘가 무엇인지 분석하여 사용된 어휘의 유사성을 확인하고 이를 통해 패턴 학습이 이루어지도록 할 필요가 있었다. 또한, 형태소 분석 대상으로는 문형 분석 및 문형 인지에 도움을 주는 형태소만 추출하기 위해 실질형태소만 취하였다. 형식형태소는 의미가 크고, 사용빈도가 높은 것만 채택하였다. 이때, 기호류는 문장구분에 도움 되는 것 외에는 모두 제외한다. 사용할 실질형태소는 명사(NNG), 고유명사(NNP), 대명사(NP), 동사(VV), 형용사(VA), 어근(XR), 관형사(MM), 부사(MAG), 접속사(MAJ), 감탄사(IC)이다. 또한 사용할 형식형태소는 부정지정사(VCN), 연결어미(EC), 종결어미(EF), 명사형전성어미(ETN), 관형사형전성어미(ETM), 주격조사

(JKS), 보격조사(JKC), 관형격조사(JKG), 부사격조사(JKB), 보조사(JX), 마침표, 물음표, 느낌표(SF)이다.

다음 단계는 딥러닝 분석을 위하여 데이터 전처리를 수행하는 것이다. 딥러닝 알고리즘을 위해 문자열 형태를 숫자로 변형하고자 한다. 예를 들어 문장에 ‘금융, 법령, 조항’이라는 3개의 어휘가 있다면 각각을 ‘101, 2133, 35’와 같이 식별할 수 있는 숫자로 변형한다. 딥러닝을 구동하기 위하여 숫자로 변형된 문자열은 다시 데이터 패딩(Data Padding) 과정을 거친다. 데이터 패딩 과정은 모든 각 문장을 동일한 크기로 만들어주는 것으로 크기가 작은 문장은 앞에 ‘0’을 붙여 지정된 크기로 동일하게 만들었다.

다음으로, 복수의 딥러닝 알고리즘의 성능을 비교하여 최적 알고리즘을 선택한다. 전처리된 데이터를 복수의 딥러닝 알고리즘을 적용해 최적 알고리즘을 선정하는 단계이다. 본 연구에서 검토한 알고리즘은 ANN, RNN, LSTM 3개이다. ANN은 기본모형으로 채택한 것이며, RNN과 LSTM은 과거에 학습한 결과를 기억하는 모델로서 비정형 문장 단위의 학습 및 추론에 적합한 딥러닝 알고리즘으로 알려져 있기 때문이다. 이들의 일반적 특징은 <표 1>과 같다.

이어서 최적 딥러닝 모델을 결정하여 생성한다. 선정된 딥러닝 알고리즘에 대하여 최적의 옵션인 하이퍼파라미터를 찾아 가장 좋은 성능을 내는 모델을 생성하는 단계이다. 탐색할 하이퍼파라미터는 유사 어휘 사용을 위한 Embedding 층을 위한 사용 어휘의 수를 설정하는 Embedding Dimension, 은닉층의 깊이를 설정하는 Hidden Layer, 각 은닉층에서의 노드의 수를 설정하는 Hidden Layer Node로 총 세

<표 1> 사용된 알고리즘의 특징

알고리즘	설명
ANN 인공신경망(기본)	ANN은 기본 신경망 알고리즘으로서 단순하나 속도가 빠르고 딥러닝의 기본 성능을 가늠하려 할 때 주로 사용된다.
RNN 순환신경망	RNN은 LSTM과 같이 순환 계열의 기본 알고리즘으로 순환 계열 알고리즘은 문장과 같이 긴 단위에 좋은 성능을 보인다.
LSTM 장단기메모리신경망	LSTM은 처리 단위가 긴 경우, 앞의 정보를 쉽게 소실하지 않아 긴 문장의 텍스트를 처리할 때 좋은 성능을 보인다.

가지를 고려한다. 그리고 각각에 대해서 최적화를 위해 여러 값을 비교한 결과 <표 2>와 같은 결과를 얻었다. 단, <표 2>는 4 x 3 x 4 = 총 48가지의 경우의 수를 모두 탐색하였다는 의미이다. 비교 결과, Embedding Dimension은 100개, Hidden Layer는 1개, Hidden Layer Node는 16개일 때가 최적으로 판정되었으므로 이에 따라 모델을 생성한다. Keras는 Tensorflow를 사용하기 쉽게 만들어주는 Wrapper 패키지이다. Input 층은 학습데이터에서 사용된 2,330개의 단어 종류가 노드로 구성된다. Embedding 층은 유사한 단어는 같은 숫자를 갖도록 하는 것으로서 학습 데이터에서 사용된 문장이 그대로 실제 데이터에서 사용되지 않아도 비슷한 문장임을 판별하게 한다. 출력층의 노드는 하나로서 Rule이 존재할 것인지를 확률로 계산하여 Y/N 방식으로 판정한다.

한편 본 연구에서의 신경망은 어떤 문장 속에 규제규칙(regulation rule)이 들어있는지의 여부를 추론하는 것이므로 출력층의 값은 0/1 값을 가지는 노드가 하나인 것으로 설정하였다. 이상과 같이 수립한 네트워크의 형태는 <그림 8>과 같다. 한편, 모델 생성에 사용된 딥러닝 플랫폼은 Tensorflow 1.16, Keras 2.2.4 버전이었다. 수행 결과 규칙여부 판정 정확도는 0.971

로서 매우 높은 정확도를 보였다. 참고로 경쟁 모델인 ANN과 RNN과의 비교는 <표 3>과 같다. 거의 유사한 성능이 나왔으나 가장 성능이 좋은 대안은 LSTM이었다.

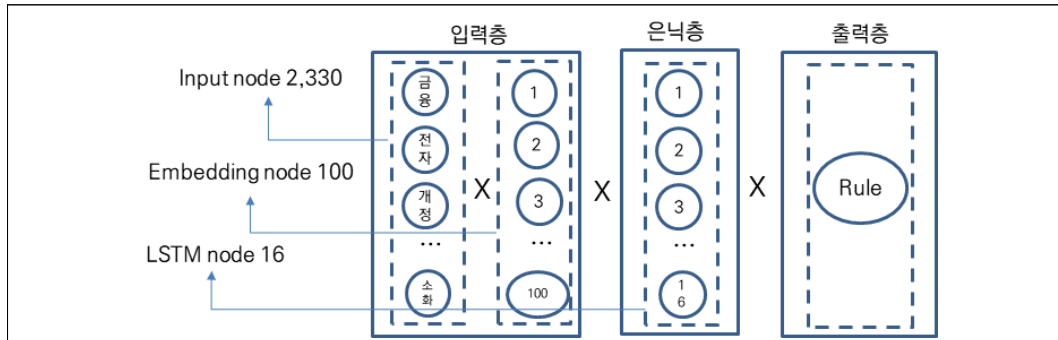
<표 2> 최적 하이퍼파라미터 탐색

Hyperparameter	대안
Embedding Dimension	50, 100, 150, 200
Hidden Layer	3, 2, 1
Hidden Layer Node	64, 32, 16, 8

<표 3> 후보 모델 성능 비교

모델	정확도 평균
ANN	0.966
RNN	0.967
LSTM	0.971

최종적으로 형태소 분석 결과와 딥러닝 수행 결과를 결합한다. 형태소 분석 결과와 딥러닝 수행 결과를 결합하여 온톨로지 작업을 수행하는 전문가가 더 빠르고 편리하게 작업을 할 수 있다. 형태소 분석의 처리 결과는 <그림 9>와 같이 산출되며, 각 칼럼의 의미는 <표 4>와 같다.



<그림 8> 출력 노드의 룰 존재 여부 판단

A	B	C	D	E	F	G	H	I	J	K	L
id	문장	문장이 문형으로 변형된 결과	해당하는 조건 규칙	조건 규칙에 해당하는 부분	ANY를 제외하고 조건에 일치하는 부분	class_domain class_range	object_property	data_property	nouns	Rule 존재 시 1	
1	①전자회계의 발행 및 관	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3		전자회계	발행	전자회계, 발행	0	
3	13 ②다음 라 호의 양무를 행	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2		호	양무	다음, 호, 양무	0	
5	28 직불전자지급수단의 발행	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3	NOUN1 NOUN2 NOUN3		직불전자지급수단	발행	직불전자지급	0	
5	47 이용자가 미리 직접 대가	VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1		가치 이용자	대하다	이용자, 선불	0	
6	49 자금이동에 직접 관여하	NOUN1 VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1	VERB8 VERB8 NOUN1		전자지급결제 대통행형	관하다	자금이동, 전	0	
7	76 ⑤제28조제1항의 규정	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1		항	규정	항, 규정, 허가	1	
8	78 ② 제28조제2항제1호	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1		업무	종류	항, 규정, 다음	1	
9	87 ③ 제28조제2항제4호	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1	NOUN1 NOUN1		구분 기본재산	마르다	항, 규정, 상1	1	
10	93 분기별 전자금융거래 중	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2		이력	범위	분기, 전자금	1	
11	96 ③ 제3항제1호에 해당	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2	NOUN1 NOUN2		금융위원회	신고하다	항, 자가, 등4	1	

<그림 9> 형태소 분석결과와 딥러닝 수행

<표 4> 분석 결과의 칼럼 의미

칼럼명1	칼럼명2	의미
A	id	원 텍스트의 기록 순서
B	문장	원 텍스트의 문장
C	문장이 문형으로 변형된 결과	원 텍스트의 명사류와 동사류를 NOUN과 VERB로 치환한 결과
D	해당하는 조건 규칙	원 텍스트에 적용된 문형 규칙
E	조건 규칙에 해당하는 부분	원 텍스트에서 문형 조건 규칙에 해당하는 부분
F	ANY를 제외하고 조건에 일치하는 부분	중간 보충 부분을 제거하고 남은 부분
G	Class Domain ~ Data Property	문형 조건 규칙에 따라 추출된 온톨로지 구성요소
H	Nouns	원 텍스트에서 사용된 모든 명사류
I	Rule 존재 시 1	원 텍스트가 규제 문장을 가지고 있는지를 딥러닝으로 판정한 결과

IV. 토의 및 결론

4.1 토의

자본시장 관련 규제 법령에 대해 기계가 해

독가능하고 추후 법률서비스가 가능한 인공지능 시스템 구축을 위해 자동 처리가 온톨로지의 형태로 변형하는 연구를 수행하였다. 최근 들어 영국 등을 중심으로 영문 규제 법령에 대해 머신리더블 규제로 표현하는 방법에 대한 연구는 추진된 바 있고(Ponkin, 2020), 자본시

장 규제 법령은 아니나 규칙을 추출하려는 시도도 있었으나(Zhong et al., 2020), 한국어로 표현된 규제 법령을 기반으로 온톨로지를 구축하는 연구가 성공한 것은 거의 찾아보기 어렵다.

이에 본 연구에서는 형태소분석과 메타 규칙 기반 온톨로지 요소를 추출, 그리고 LSTM을 기반으로 하는 문장의 규제 규칙 포함 여부를 추론하는 방법론을 제안하였다. 법령 텍스트로부터 Machine Readable Regulation을 자동 변환 또는 효율적 변환하기 위하여 메타규칙을 적용한 것, 특히 한국어에서 시도한 것은 본 연구가 최초의 것이다. 본 연구는 다음과 같은 학술적·실무적 의의가 있을 것이다.

첫째, 한국어 법령도 온톨로지 변환을 위한 방법이 가능함으로 보였다. 한국어 법령은 영문 법령보다 더 의미가 모호하고 비록 법문 자체의 본질적인 모호성에도 불구하고(김혁기, 2009; 김경하 등, 2019), 법령 용어의 난해성과 모호성(이계수, 2011), 문법적 측면도 이해도가 떨어지는데 영향을 준다는 지적이 있는데, 이러한 상황에서도 유의한 효율성을 보였다는 것이 의미가 있다.

둘째, 자본시장 법령에 대한 인공지능 기반 지능형 법률 서비스의 가능성을 열었다는데 의의가 있다. 특히, 최근 디지털 무역의 확장세가 두드러지고 디지털세 논의가 이루어지는 등 급변하는 거시경제 환경 속에서 기업들은 개정 및 제정되는 각종 법률에 대해서 신속하고 정확한 이해 및 자신의 기업에 미칠 파급효과에 대한 분석을 해야 하는바 적지 않은 법률서비스 비용이 발생할 것인데 본 연구의 방법 등을 발전시켜 이러한 서비스 비용을 혁신적으로 절

감할 수 있을 것이다. 또한 국가적으로 각종 소송 등으로 인한 사회적 비용을 저감하고 경제 주체들이 위기를 관리할 수 있도록 하는 것이므로, 본 연구는 공공성을 가지기도 한다.

셋째, 본 연구는 전자금융감독규정의 법령을 판독하여 기계어로 변환하는 일련의 과정을 제시한다. 즉, 법령의 형태소를 분석하여 문장을 메타언어로 변환하고, 규제 문장이 있는지를 판정하기 위해 딥러닝 분석과정을 수행한다. 이러한 본 연구의 결과는 레그테크 기반의 기술 개발을 위한 기초적 사례를 보여준다. 향후 연구는 이를 보완하여 머신러닝을 규제 모델을 직관적으로 보여주는 시각화 프로그램 개발할 수도 있을 것이다.

넷째, 본 연구는 레그테크가 적용 가능한 여러 산업에 실무적 시사점을 제시한다. 예를 들어, 금융회사, 회계감사법인, 법무법인 등이 제공하던 회계감사업무, 법률자문서비스 등의 규제 검토를 해당 규제를 적용받는 기업이 레그테크를 적용하여 내부적으로 금융규제를 해석하고 적용함으로써 비용을 절감하는 효과가 기대된다. 본 연구는 이러한 레그테크를 적용하기 위한 규제의 분석과 그 결과에 대한 기초자료를 제공하고 있다. 또한 금융위기 이후 촉발된 금융규제의 강화는 핀테크의 발전을 불러왔으나, 오히려 규제기관이 규제기업보다 핀테크의 적용에 있어 민첩하게 시대적 변화에 대응하지 못하고 있다는 우려가 있다. 레그테크는 기계학습을 통해 규제와 법령을 신속하게 해석한다는 점에서 규제기관 측면에서도 효과적인 관리와 감독을 위한 기술적 토대를 제공할 가능성이 크다. 또한 본 연구는 이러한 규제기관-규제-규제기업으로 이어지는 보고체계에 있어 레그테크

크를 적용하기 위한 생태계 구축 노력이 필요함을 제시한다.

다만, 본 연구는 전자금융감독규정의 간단한 사례를 통해 비교적 단순한 법령의 판정 및 알고리즘을 제시하고 있다. 이는 본 연구의 한계점으로 향후 연구는 보다 복잡한 법령이나 금융규제의 적용을 통해 본 연구에서 제시한 규제문장의 알고리즘을 확장할 수 있을 것이다. 예를 들어, 새로운 규제와 기존 규제의 문장간 차이를 기계어를 통해 판독하고 알고리즘을 개발하여 기업들의 규제적응력을 높이는 방법론적 탐색도 가능할 것이다.

4.2 결론

금융위기가 촉발한 규제강화의 환경적 요인과 핀테크를 비롯한 인공지능을 활용한 기술발전은 레그테크의 등장을 야기하였다. 적절한 규제환경을 조성하고, 규제관점에서 이해관계들의 필요를 충족하기 위해서는 규제준수비용을 줄이고, 감독관점에서 효율적인 감독이 가능한 체계가 필요하게 되었으며, 이를 기술적으로 해결할 수 있는 대안체계로서 레그테크가 제시된 것이다. 즉, 레그테크는 시장참여자의 효율적인 규제준수를 촉진할 뿐 아니라 입법자와 규제기관 관점에서도 복잡한 규제가 등장하고 수많은 신산업과 스타트업 등 신규 시장참여자가 등장하는 시대에서 규제 목적이 제대로 달성되는지 효과적으로 감시할 기제를 제공하는데 그 의의가 있다.

참고문헌

- 김민중, 손달호, “핀테크 결제 서비스에서 프라이버시 및 신뢰요인의 영향,” 인터넷전자상거래연구, 제20권, 제1호, 2020, 87-98.
- 김은정, 김주현, 김종원, “핀테크 사용의도에 영향을 미치는 요인에 관한 연구,” 정보시스템연구, 제26권, 제1호, 2017, pp. 75-91.
- 우가인, 조재형, “핀테크 기반의 간편결제 서비스 지속적 사용의도에 대한 요인 연구,” 정보시스템연구, 제26권 제3호, 2017, pp. 25-46.
- 유제민, “레그테크(RegTech)의 도입과 규제법학의 과제,” 경제규제와 법, 제12권, 제1호, 2019, pp. 7-25.
- 정대현, 장활식, 박광오, “핀테크 활성화를 위한 사용환경특성과 지속사용의도,” 정보시스템연구, 제26권, 제2호, 2017, pp. 123-142.
- 조창훈, “국내 레그테크의 시장성 검토 및 도입시 고려사항,” 전자금융과 금융보안, 2017, pp. 65-79.
- Alamaki, P., and Broby, D., “The Effectiveness of Regulatory Reporting by Banking Institutions,” 2019, *Working Paper*.
- Anagnostopoulos, I., “Fintech and Regtech: Impact on Regulators and Banks,” *Journal of Economics and Business*, Vol. 100, No. Nov-Dec, 2018, pp. 7-25.
- Butler T., and O'Brien L., “Understanding RegTech for Digital Regulatory

- Compliance,” *Disrupting Finance*, 2019, pp. 85-102.
- Chen, S., “Detection of Fraudulent Financial Statements Using the Hybrid Data Mining Approach,” *Springer Plus*, Vol. 5, No. 1, 2016, pp. 1-16.
- Colaert, V., “RegTech as a Response to Regulatory Expansion in the Financial Sector,” 2018, *Working Paper*, Available at SSRN 2677116.
- Colgren, T. D., “XBRL, Blockchain and New Technologies: A Focus on Innovative Solutions and Core Competencies in Technology and Analytics is a Must for Management Accountants,” *Strategic Finance*, Vol. 99, No. 7, 2018, pp. 62-64.
- Deloitte, “RegTech is the New FinTech: How Agile Regulatory Technology is Helping Firms Better Understand and Manage Their Risks,” 2016, <https://assets.ctfassets.net/sdIntm3tthp6/resource-asset-r162/630efe5964256dd06e2731f2e75846a2/09172435-9d76-429e-8658-5477a3dca62c.pdf>
- FCA, “Feedback Statement, Financial Conduct Authority, Call for Input on Supporting the Development and Adopters of RegTech, 2016, p. 3 at <https://www.fca.org.uk/publication/feedback/fs-16-04.pdf>.
- FCA, “RegTech Applies to New Technologies Developed to Help Overcome Regulatory Challenges in Financial Services,” 2019, Available at: <https://www.fca.org.uk/firms/regtech>
- Fröhlich, M., “Enabling RegTech Up Front: Unambiguous Machine readable Regulation,” *The RegTech Book*, 2019.
- Gray, G., and Debreceny, R., “Ataxonomy to Guide Research on the Application of Data Mining to Fraud Detection in Financial Statement Audits,” *International Journal of Accounting Information Systems*, Vol. 15, No. 4, 2014, pp. 357-380.
- Huang, S., Tsaih, R. and Yu, F., “Topological Pattern Discovery and Feature Extraction for Fraudulent Financial Reporting,” *Expert Systems with Applications*, Vol. 41, No. 9, 2014, pp. 4360-4372.
- Kavassalis, P., Stieber, H., Breymann, W., Saxton, K., and Gross, F. J., “An Innovative RegTech Approach to Financial Risk Monitoring and Supervisory Reporting,” *The Journal of Risk Finance*, Vol. 19, No. 1, 2018, pp. 39-55.
- Kwon, O., Kim, Y. S., and Choi, S. U., “Current Status and Challenges of RegTech: Focused on Machine Readable Regulation,” *Logos Management Review*, Vol. 18, No. 3, 2020, pp. 45-60.
- Omar, N., Johari, Z. and Smith, M., “Predicting

- Fraudulent Financial Reporting Using Artificial Neural Network,” *Journal of Financial Crime*, Vol. 24, No. 2, 2017, pp. 362-387.
- Peters, G., and Vishnia, G., “Blockchain Architectures for Electronic Exchange Reporting Requirements,” *Handbook of Blockchain, Digital Finance, and Inclusion 2*, 2018, pp. 271-329.
- UK Government Office for Science, “Fintech Futures - The UK as a World Leader In Financial Technologies,” 2015, GS/15/3 available at www.gov.uk/go-science.
- Yang, D., and Li, M., “Evolutionary Approaches and the Construction of Technology-Driven Regulations,” *Emerging Markets Finance and Trade* Vol. 54, No. 14, 2018, pp. 3256-3271.
- Yoost, D. A., and Mathaisel, B. F., “Board Oversight of the Risks in Using Big Data and Advanced Analytics,” *The RMA Journal*, Vol. 98, No. 5, 2016, pp. 38-42.

Appendix. 데이터 처리를 위한 코드

A. 데이터 토큰나이징 및 데이터 패딩 처리를 위한 코드

```
print("\n##### Data Tokenizing #####")
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
import numpy as np
import math

tokenizer = Tokenizer(num_words=max_words)
tokenizer.fit_on_texts(texts)
word_index = tokenizer.word_index

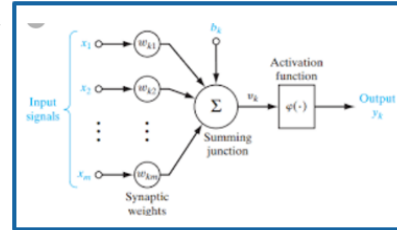
print("\n##### Data Padding #####")
data = tokenizer.texts_to_sequences(texts)
data = pad_sequences(data, maxlen=maxlen)

if use_val:
    data_val = tokenizer.texts_to_sequences(texts_val)
    data_val = pad_sequences(data_val, maxlen=maxlen)
```


B. ANN 알고리즘을 사용하기 위한 주요 코드

```
max_words = 10000
maxlen = 100
class_number = 2
embedding_dim = 100
model_type = 'ANN'
need_one_hot_encoding = True
epochs = 10
batch_size = 32
```

가장 빈도 높은 10,000 개의 단어사용
길이를 고정
분류할 클래스의 수
임베딩 층 차원의 크기
기본적으로는 원-핫-인코딩을 수행
수행할 에포크의 수
한 번에 훈련할 배치 사이즈

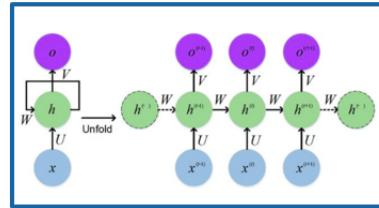


```
model_type == 'ANN':
model = Sequential()
model.add(Embedding(input_dim=max_words, output_dim=embedding_dim, input_length=maxlen))
model.add(Flatten())
model.add(Dense(units=32, activation='relu'))
model.add(Dense(units=final_unit, activation=final_activation))
model.summary()
```

C. RNN 알고리즘을 사용하기 위한 주요 코드

```
max_words = 10000
maxlen = 100
class_number = 2
embedding_dim = 100
model_type = 'RNN'
need_one_hot_encoding = True
epochs = 10
batch_size = 32
```

가장 빈도 높은 10,000 개의 단어사용
길이를 고정
분류할 클래스의 수
임베딩 층 차원의 크기
기본적으로는 원-핫-인코딩을 수행
수행할 에포크의 수
한 번에 훈련할 배치 사이즈

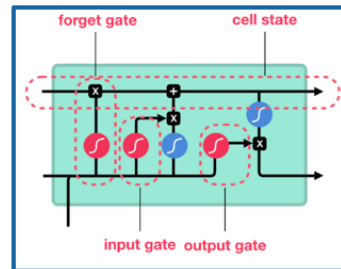


```
model_type == 'RNN':
model = Sequential()
model.add(Embedding(input_dim=max_words, output_dim=embedding_dim))
model.add(SimpleRNN(units=32))
model.add(Dense(units=final_unit, activation=final_activation))
model.summary()
```

D. LSTM 알고리즘을 사용하기 위한 주요 코드

```
max_words = 10000
maxlen = 100
class_number = 2
embedding_dim = 100
model_type = 'LSTM'
need_one_hot_encoding = True
epochs = 10
batch_size = 32
```

가장 빈도 높은 10,000 개의 단어만 사용
길이를 고정
분류할 클래스의 수
임베딩 층 차원의 크기
기본적으로는 원-핫-인코딩을 수행한다
수행할 에포크의 수
한 번에 훈련할 배치 사이즈



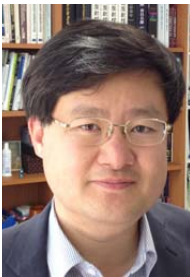
```
model_type == 'LSTM':
model = Sequential()
model.add(Embedding(input_dim=max_words, output_dim=embedding_dim))
model.add(LSTM(units=32))
model.add(Dense(units=final_unit, activation=final_activation))
model.summary()
```

최 승 욱 (Choi, Seung Uk)



고려대학교 학사와 석사 및 박사학위(회계학전공)를 취득하였다. 현재 경희대학교 회계·세무학과 교수로 재직하고 있으며, 주요 관심분야는 회계 품질, 감사품질, 빅데이터 등이다.

권 오 병 (Kwon, Oh Byung)



서울대학교 경영학사와 KAIST 석사와 박사학위(MIS 전공)를 취득하였다. 현재 경희대학교 경영학과 교수로 재직하고 있으며, 주요 관심분야는 빅데이터분석, 레그테크 등이다.

<Abstract>

Suggestions for the Development of RegTech Based Ontology and Deep Learning Technology to Interpret Capital Market Regulations

Choi, Seung Uk · Kwon, Oh Byung

Purpose

Based on the development of artificial intelligence and big data technologies, the RegTech has been emerged to reduce regulatory costs and to enable efficient supervision by regulatory bodies. The word RegTech is a combination of regulation and technology, which means using the technological methods to facilitate the implementation of regulations and to make efficient surveillance and supervision of regulations. The purpose of this study is to describe the recent adoption of RegTech and to provide basic examples of applying RegTech to capital market regulations.

Design/methodology/approach

English-based ontology and deep learning technologies are quite developed in practice, and it will not be difficult to expand it to European or Latin American languages that are grammatically similar to English. However, it is not easy to use it in most Asian languages such as Korean, which have different grammatical rules. In addition, in the early stages of adoption, companies, financial institutions and regulators will not be familiar with this machine-based reporting system. There is a need to establish an ecosystem which facilitates the adoption of RegTech by consulting and supporting the stakeholders. In this paper, we provide a simple example that shows a procedure of applying RegTech to recognize and interpret Korean language-based capital market regulations. Specifically, we present the process of converting sentences in regulations into a meta-language through the morpheme analyses. We next conduct deep learning analyses to determine whether a regulatory sentence exists in each regulatory paragraph.

Findings

This study illustrates the applicability of RegTech-based ontology and deep learning technologies in Korean-based capital market regulations.

Keyword: RegTech, Ontology, Deep Learning, Capital Market Regulation

* 이 논문은 2021년 1월 13일 접수, 2021년 1월 29일 1차 심사, 2021년 3월 3일 2차 심사, 2021년 3월 25일 게재 확정되었습니다.