

자연어 처리가 금융을 바꾼다

글_배정훈 한양대학교 교수

감성 분석을 통해 효율적으로 정보를 추출하면 미래 주식 가격의 움직임을 예측할 수 있다. 이를 가능케 하는 것이 바로 인공지능의 한 분야인 자연어 처리 기술이다. 인공지능이 글의 문맥을 이해하고 미묘한 뉘앙스를 파악하여 미래의 주식 가격에 관한 가치 있는 정보를 제공하는 시대를 미리 살펴보자.

컴퓨터가 인간의 말을 알아듣다

컴퓨터가 인간의 언어를 이해할 수 있을까? 인간의 음성을 인식하고, 책의 내용을 요약하고, 뉴스에서 정보를 추출하고, 말과 글의 감성을 파악하고, 외국어를 번역할 수 있는 컴퓨터가 존재할까? 이와 같은 문제를 연구하는 인공지능의 한 분야가 자연어 처리(Natural Language Processing, NLP)이다. 자연어 처리는 컴퓨터가 인간의 말과 문자를 분석하도록 프로그래밍하는 인공지능 기술이다. 자연어 처리 기술을 통해 컴퓨터 프로그램은 인간의 두뇌와 마찬가지로 언어의 맥락을 이해하고 추론함으로써 비정형적인 문자 데이터를 이해할 수 있다. 이 인공지능 기술은 외국어 번역, 감성 분석, 주제 인식, 글의 요약, 음성 인식 등 다양한 분야에 사용된다.

페이스북, 트위터와 같이 대량의 문자적인 정보를 보유한 기업은 자연어 처리 기술을 이용하여 사용자의 관심사를 파악하고 추가적인 트래픽을 만들기 위해 사용자 맞춤형 정보를 끊임없이 제공한다. 유튜브는 자연어 처리 기술에 기반한 음성 인식과 자동 번역 기술을 이용하여 동영상에 여러 가지 언어로 된 자막을 제공한다. 애플의 시리와 삼성의 빅스비도 자연어 처리 기술의 음성 인식 기능을 활용하여 핸드폰을 조작하는 서비스를 제공한다.

감성 분석을 통해 주가 흐름을 예측

금융 시장에서도 자연어 처리 기술이 금융 투자, 리스크 관리, 신용 평가 등과 같은 여러 분야에 적용되고 있다. 일반적으로 숫자로 표현되는 정형적인 데이터가 줄 수 있는 정보는 이미 금융 시장에 반영이 되었을 가능성이 높아 정형적인 데이터를 분석하여 추가적인 가치를 만드는 일은 아주 어렵다. 하지만, 뉴스, 소셜 미디어 등에서 매일 대량으로 생산되고 있는 비정형적인 문자 데이터는 구조적으로 금융 시장에 즉각 반영되는 것은 불가능하다.

“

따라서 자연어 처리 기술을 효율적으로 활용하여 대량의 문자 데이터에서 정보를 추출한다면 금융 시장에 추가적인 가치를 더할 가능성이 높다. 자연어 처리 기술 중에 하나인 감성 분석은 뉴스 및 소셜 미디어에서 생산되는 대량의 문자 데이터에서 투자와 관련된 통찰력을 얻는 데 활용되고 있다.

주식 시장의 과거 가격, 거래 정보 그리고 회계 정보 등의 정형적인 데이터에서 얻을 수 있는 정보는 이미 많은 사람들이 주식 투자에 활용하고 있어 주식의 가격에 이미 반영이 되었을 가능성이 높다. 하지만 매일 대량으로 생산되는 경제 뉴스, 소비자들의 의견 등의 비정형 문자 데이터는 실시간으로 주식의 가격에 반영되기는 어렵다. 비정형 문자 데이터가 주식의 가격에 반영되기 전에 감성 분석을 통해 효율적으로 정보를 추출하면 미래의 주식 가격의 움직임을 예측할 수 있다. 과거의 감성 분석 방법은 긍정어 및 부정어를 정의하는 감성 사전에 기반을 두고 단순히 특정 단어의 수를 파악하는 데 그쳤지만, 최근의 딥러닝을 활용한 방법은 글의 문맥을 이해하고 미묘한 뉘앙스를 파악하여 미래의 주식 가격에 관한 가치 있는 정보를 제공하고 있다.

고차원 언어 능력 지닌 인공지능 탄생 가능

필자는 대량의 뉴스 미디어 및 증권 포털 사이트의 비정형 문자 데이터를 자연어 처리 기술로 분석하여 한국의 금융 위기를 예측하는 연구를 수행한 적이 있다. 변동성 지수, 이차율 기간 스프레드, 신용 스프레드 등의 정형적인 데이터를 사용하는 전통적인 위험 관리는 금융 위기를 선제적으로 탐지하지 못한다. 필자는 비정형 문자 데이터에서 전통적인 위험 관리 방법에서 포함하지 못하는 정보를 추출할 수 있었고, 이 정보를 활용하여 한국 시장의 금융 안정 지수를 예측하고 선제적으로 금융 위기를 관리를 할 수 있는 방법을 찾을 수 있었다. 현재 국내 몇몇 자산운용사들은 이 연구를 바탕으로 뉴스 미디어 및 증권 포털 사이트의 문자 데이터를 자연어 처리 기술을 통해서 분석하고 선제적 위험관리에 적용하고 있다. 자연어 처리 기술의 현재 수준은 아직 인간의 고차원적인 언어능력과는 격차가 크다. 구글 번역기는 완벽한 번역을 하지 못하고, 시리와 빅스비에서 음성 인식의 오류가 종종 발생한다. 금융 시장에서 사용되고 있는 자연어 처리 기술도 한계점이 많으며, 현재에는 전통적인 금융기법에 더하여 보조적인 수단으로 사용되고 있다. 하지만 수많은 천재적인 엔지니어들이 기술 발전에 매진하고 있고, 자연어 처리 기술을 탑재한 인공지능은 매일 생산되는 대량의 문자 데이터를 학습하여 진화하고 있다. 빠른 시일 내에 고차원적인 인간의 언어 능력을 지닌 인공지능이 탄생할 것은 자명하다.

자연어 처리 기술이 더 고도화되면 미래 금융업에서 파괴적인 혁신이 올 것이다. 애널리스트 리포트, 신용 평가 보고서, 회계 감사 보고서 등 인간의 언어로 이루어진 모든 금융 서비스는 자연어 처리 기술을 적용한 인공지능으로 빠르게 대체될 가능성이 높다. 인간과 기계가 어떻게 공존을 해야 할지 고민이 필요한 시점이다.

”