
빅데이터와 분석예측의 관계

강선경* · 이현창** · 신성윤***

*원광대학교 컴퓨터소프트웨어공학과

**원광대학교 디지털콘텐츠공학과

***군산대학교

Relationship between Big Data and Analysis Prediction

Sun-Kyoung Kang* · Hyun-Chang Lee** · Seong-Yoon Shin***

*Department of Computer(Software) Engineering, Wonkwang University, Iksan, 54538, South Korea

**Department of Digital Contents Engineering, Wonkwang University, Iksan, 54538, South Korea

***Scool of Computer Information & Communication Engoneering, Kunsan National University,
Kunsan 54150, South Korea

E-mail :{doctor10, hclglory}@wku.ac.kr, s3397220@kunsan.ac.kr

요 약

본 논문에서는 빅데이터를 이용하여 어떠한 것을 분석하고 또 분석한 것을 가지고 무엇을 예측할 것인가에 대한 중요성에 대해 이야기한다. 일상생활에서 일어나서 나도 모르게 축적되고 있는 데이터의 많은 양을 과연 어느 부분에 어떻게 적용하여 활용할 것인가의 문제는 아주 중요한 요소이다. 이러한 데이터를 가지고 무엇을 예측하고 어떻게 활용할 것인지를 구체화 시키는 작업은 여러 종류가 존재한다. 그 중에서 가장 적절한 것을 찾는 것이 예측 확률을 높이는 방법이라 하겠다. 또한 분석하여 예측한 내용이 실제 생활에 유용하게 쓰여져야 그 데이터는 의미 있는 데이터라 할 수 있다.

ABSTRACT

In this paper, we discuss the importance of what to analyze and what to predict using Big Data. The issue of how and where to apply a large amount of data that is accumulated in my daily life and which I am not aware of is a very important factor. There are many kinds of tasks that specify what to predict and how to use these data. Finding the most appropriate one is the way to increase the prediction probability. In addition, the data that are analyzed and predicted should be useful in real life to make meaningful data.

키워드

빅데이터, 정형, 비정형, 분석, 예측모델

1. 서 론

빅데이터란 기존의 데이터베이스 관리도구로 데이터를 수집-저장-관리-분석할 수 있는 역량을 넘어서 대용량 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다. 그러면서 정보 플랫폼도 진화하게 되었다. 정보시스템 통합 고도화와

Movbile, Cloud, SNS 등 제타바이트 시대에 돌입하게 된 것이다. 또한 정형/비정형 데이터의 폭 증현상으로인해 전세계 데이터는 매년 40% 증가하고 있다. 이렇게 증가되는 많은 양의 데이터를 처리하고 분석하는 기술들도 함께 중요한 이슈로 등장하게 되었다. 그리고 다양한 대용량 데이터를 분석하여 새로운 통찰력을 얻을 수 있다. 예를 들어 다중채널 고객 감성 및 경험을 분석하고 신생

아실에서 예측분석을 통해 생명을 위협하는 상황을 최대한 빨리 감지한다든지 기후/지리 데이터 분석을 통해 풍력 발전기 및 풍력 발전소 부지 플래닝과 날씨 패턴 분석등 실시간으로 다양한 데이터의 분석을 통해 사회와 경제에 도움을 주는 예측을 해 낼 수 있다. 본 논문에서는 이러한 데이터의 올바른 쓰임을 위해 분석 및 예측에 대한 방법에 대해 이야기해 보고자 한다.[1][2]

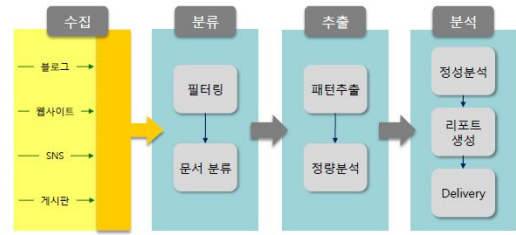


그림 2. 데이터 분석 과정

II. 데이터의 종류

일상생활에 존재하고 있는 데이터의 종류는 한 눈에 데이터의 정보를 알아 볼 수 있는 정형데이터와 그 의미를 한번에 알아볼 수 없는 비정형 데이터로 구성되어 있다. 아래 보이는 그림1은 데이터의 종류를 보여주고 있다. 표에서 보이는 것과 같이 정형데이터는 전체 15%이고, 비정형 데이터는 전체의 85%를 차지하고 있다. 이렇게 한 눈에 데이터를 확인할 수 있는 양보다 의미를 분석해 내지 않으면 그 데이터가 어떠한 내용인지 알 수 없는 데이터들이 대부분이라는 것을 알 수 있다. 따라서 데이터 분석을 통해 의미 있는 데이터를 뽑아내는 기술이 매우 중요하다.

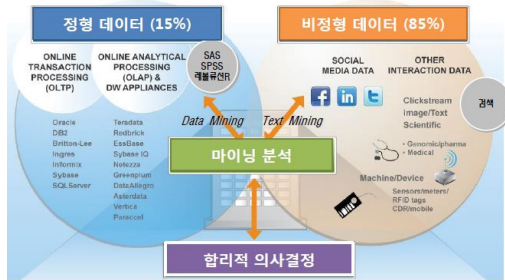


그림 1. 정형 비정형 데이터 종류

III. 데이터 분석

많은 양의 데이터에서 의미 있는 데이터를 뽑아내기 위해서는 데이터의 분석 과정이 필요하다. 이러한 데이터의 분석과정이 바로 데이터의 예측률을 높이는 중요요소를 결정 짓게 되는 것이다. 그림2에서 보는 것과 같이 수집단계에서는 블로그, 웹사이트, SNS 그리고 게시판에서 비정형 데이터를 모은다음 분류단계에서 필터링을 하고 정의해 놓은 틀에 맞추어 분류를 하게 된다 그렇게 분류해 놓고나면 일정한 패턴을 추출하여 그를 정량적으로 분석하는 작업을 수행하게 된다. 정성적 분석 후 리포트를 생성하고 그것을 필요에 맞는 곳에 공급하게 되는 과정으로 이루어져 있다.[3]

IV. 예측

예측분석은 보다 나은 의사결정을 하기 위해 경험(데이터)으로부터 학습하여 개별 사안(사람)의 미래 행위를 예측하는 예측모델을 생성하고 활용하는 테크놀로지이다. 그리고 예측모델이란 클릭, 구매, 거짓말, 사망 또는 부동산 담보대출 조기상환과 같은 한 개인에게 일어날 일을 예측하는 메커니즘으로 그것은 해당 개인의 특장들(변수들)을 입력값으로 하여 예측 점수를 결과값으로 산출한다. 그 점수가 높을수록 해당 개인이 예측된 행위를 할 가능성이 높음을 의미한다. 즉 우리는 무엇을 예측할 것인가와 어떻게 활용할 것인가가 빅데이터 분석에서는 매우 중요하다는 것을 다시 한번 이야기해 주고 있는 것이라 할 수 있다. 예측결과에 기반하여 의사결정을 하고 각각의 예측결과에 따라 취하는 행동이 전혀 다를 수 있기 때문이다.

V. 결론

데이터의 수집 및 분석 그리고 예측모델 설정의 관계는 아주 밀접하다. 비정형 데이터에서 사용자가 원하는 데이터를 추출해 내기 위해서는 어떠한 내용의 결과값을 도출하고 싶은지부터 결정되어 주어야 한다. 그것이 바로 빅데이터와 분석예측의 관계라 할 수 있다.

참고문헌

- [1] 김근태, "빅데이터 분석을 위한 기업의 Big Analytics 환경 변화," 한국정보처리학회 19, 정보처리학회지, 2012
- [2] 최성, 우성구, "빅데이터 정의, 활용 및 동향," 한국정보처리학회 19, 정보처리학회지, 2012
- [3] 김진국, 최성수, 지수영, 류관희, "데이터베이스 연동을 통한 빅데이터 분석결과 가시화," 한국빅데이터서비스학회, 한국빅데이터서비스학회논문지 21, 2015, pp 1-10