

B-spline 회귀 시계열 요소 분해를 이용한 효율적인 판매 급증 이상 탐지

김 호 현

한국방송통신대학교 대학원

hohkim@koreanair.com

Efficient Sales Growth Anomaly Detection Using Time Series Decomposition with B-spline Regression

Hohyun Kim

Korea National Open University Graduate School

요 약

*Anomaly Detection*은 데이터 가운데서 기대되는 정상적인 행동을 따르지 않는 패턴을 찾아내는 것으로 *Fraud Detection*, *Intrusion Detection*, *Fault Detection* 등의 다양한 분야에 이용되고 있다. 항공 회사들은 대리점의 급작스런 부도로 인한 매표대금의 손실을 방지하기 위하여 대리점의 비정상적인 판매 급증을 조기에 탐지하여 경보하는 시스템을 운용할 필요가 있다. 그러나 입력 데이터로 주어지는 일별 판매액은 *Time Series Data*의 일반적 특성인 *Seasonality*와 *Trend* 속성을 가짐으로 인해 단순히 전년 동기와 비교하는 현 시스템은 높은 *False Alarm Rate*를 보이는 경향이 있다.

본 논문에서는 *Anomaly Detection*을 위해 *Time Series Decomposition*과 *Generalized ESD(Extreme Studentized Deviation) Test*를 결합하여 사용하되, *Time Series Decomposition* 시 *Robust Statistic*와 *B-spline Regression*을 사용하여 *Trend* 요소를 추출해냄으로써 정확성을 높이는 독특한 방법을 제시한다.

1. 서론

일부 항공 회사들은 대리점의 급작스런 부도로 인한 판매대금 손실을 예방하기 위하여 대리점의 판매 이상 급증을 탐지하여 지점에 자동 통보하는 조기 경보 시스템을 운용하고 있다. 그러나 현 시스템은 단순히 전년 동 분기의 평균 실적과 비교하여 일정수준 이상 증가 시 *Anomaly*로 판별하는 방법을 사용하여 높은 *False Alarm Rate*를 보이고 있어 개선이 필요하다.

현 시스템이 정확한 탐지를 해 내지 못하는 주 이유는 입력 데이터로 사용하는 일별 판매 실적이 *Time Series Data*의 일반적 특성인 *Seasonality*와 *Trend* 속성을 가지고 있기 때문이다.

*Twitter*에서는 *Long-Term Time-Series Data Anomaly Detection* 알고리즘을 개발하여 *Open Source*로 공개하였는데, 이는 탐지능력이 상당히 우수하지만 일별 판매 실적과 같이 데이터량이 많지 않은 경우는 적용하기 어려운 문제가 있었다.

*Time Series Data*에 대한 *Anomaly Detection*은 데이터가 내

포하고 있는 *Trend* 요소를 어떻게 처리하느냐에 큰 영향을 받는다. 본 논문에서는 *Time Series Data*로부터 *moving median* + *B-spline regression*을 이용하여 *Trend* 요소를 추출하는 새로운 방법을 제안한다.

2절은 동 논문의 기여점을 밝히고, 3절은 *Time Series Data*의 일반적인 특성과, 기존에 개발된 *Anomaly Detection* 방법론들을 기술한다. 4절에서는 *Time Series Data*와 *Generalized ESD(Extreme Studentized Deviation) Test*를 결합한 제안 알고리즘을 설명하고, 5절에서는 실험 결과를 통해 제안 알고리즘의 우수함을 보인 후 6절에서 결론으로 맺는다.

2. 논문의 기여

- 기존 *Time Series Decomposition* 알고리즘의 개선

기존 *R-Package*의 *STL Time Series Decomposition*에서 *Trend* 속성 추출 시 단순히 이동평균법을 사용하고 있

으나, 이동평균법과 B-spline regression을 결합하여 Smoothing 하여 탐지 능력을 개선하는 방안을 제시하였다.

- 현업에서 탐지 시스템 개발 시 참조 모델로 활용 가능
- 수만 개에 이르는 대리점에 대해 탐지 알고리즘을 병렬로 처리하여 빠르게 동작하는 실 시스템을 구현하였다.

3. 관련연구

3.1 Anomaly Detection

- *Anomaly Detection*의 정의
- *Anomaly Detection*의 활용 분야
- *Anomaly Detection Techniques*

3.2 Time Series Data

- *Time Series Data*의 특성
- *Time Series Decomposition*

3.3 Statistical Learning 기반 *Anomaly Detection* 기법

- Parametric Techniques
- Non-parametric Techniques
- Statistical Learning 기반 Technique의 장단점

4. 제안 알고리즘

4.1 *Time Series Decomposition*

...

4.2 Extracting Trend Using B-spline

...

4.3 Detecting with *Generalized ESD(Extreme Studentized Deviation) Test*

...

「알고리즘 Pseudo-code 표」

5. 실험 결과

- 현 시스템
- *STL Time Series Decomposition*
- 제안 알고리즘

6. 결론

...

참고문헌