

Term Project 계획서			
학 과	방송통신대학교 대학원 정보과학과	학 번	201584-010070
이 름	(한글) 김 호 현		
	(영문) Hohyun Kim		
Term Project Subject	Moving Median 및 B-spline 회귀 시계열 요소 분해를 이용한 효율적인 판매 급증 이상 탐지 시스템		
	Efficient Sales Abnormal Growth Anomaly Detection System Using Time Series Decomposition with Moving Median and B-spline Regression		
Abstract	<p>Anomaly Detection은 기대되는 패턴을 따르지 않는 이상 값을 탐지하는 것으로 Intrusion Detection, Fraud Detection, Fault Detection 등의 다양한 분야에 이용되고 있다. 항공 회사들은 안전한 대금관리를 위하여 대리점의 판매 이상 급증을 조기에 탐지하는 것이 필요하여 이를 위한 시스템을 운용하고 있다. 그러나 시계열 Data인 일별 판매액이 갖는 Seasonality 및 Trend 속성으로 인해 이상 값의 정확한 탐지가 어렵다.</p> <p>Trend 요소 분해를 위해 Moving Average 사용 시는 이상 값의 영향을 받아 False Alarm Rate가 높아지는 문제가 있다. Twitter에서는 매우 효율적인 Long-Term Time-Series Data Anomaly Detection 알고리즘을 개발하여 Open Source로 공개하였다. 그러나 Daily 로 한 건의 Data 만 발생하는 Data 량이 많지 않은 경우는 이를 이용하기 어렵다.</p> <p>본 논문에서는 Moving Median + B-spline Regression 방법을 통하여 시계열 Trend 요소를 보다 정확히 분해해 냄으로써 탐지 신뢰성을 높일 수 있는 방안을 제시한다. 또한 탐지 알고리즘을 병렬로 처리하여 효율적으로 동작하는 시스템을 만들어 보임으로써 현업에서 이를 참고하여 활용 가능토록 한다.</p>		
Motivation	<p>현재 회사에서는 대리점의 판매 이상 급증을 탐지하여 지점에 경고하는 시스템을 운용하고 있다. 그러나 단순히 과거 실적 평균 대비 일정수준 이상 증가 시 이상 건으로 판별하는 현재의 방법은 높은 False Alarm Rate를 보이고 있어 개선이 필요하다.</p> <p>이번 학기의 팀 프로젝트를 통해 신뢰성 높은 판매 급증 이상 탐지 알고리즘을 연구해 보고자 한다.</p>		
Contribution	<p>시계열 (Time-series) Data에 대한 Anomaly 탐지를 위해서 Data로부터 Trend 요소를 정확히 분리해 내는 것이 매우 중요한데, 기존의 Moving Average, Moving Median의 방법보다 개선된 알고리즘을 제시하고자 한다. (Moving Median + B-spline Regression 방법을 사용하여 Trend 요소를 추출하면 정확도를 높일 수 있을 것으로 기대하고 있다.)</p> <p>또한 수 만개에 이르는 대리점의 데이터를 병렬로 처리하여 효율적으로 동작하는 실제 시스템을 만들어 보일 예정이며, 이상 탐지 시스템 개발 시 좋은 참고 사례가 될 것으로 생각한다.</p>		

<p>Schedule</p>	<ul style="list-style-type: none"> - 3.14 : Term Project 계획서 제출 - 3.15 ~ 4.23 : 관련 논문 분석 및 개선방안 고찰 <ul style="list-style-type: none"> • Anomaly Detection Survey 논문을 통하여 Anomaly Detection의 개념, 적용분야, 접근 방법론에 대한 이해도 증진 • Time-series Data에 대한 Anomaly Detection 알고리즘 Study • 참고 논문으로 선정한 논문을 분석 (국내 논문 6편, 국외 논문 6편) 및 기존의 제안된 방법들의 부족한 고찰 • 특히 국외 참고 논문으로 선정한 “A Novel Technique for Long-Term Anomaly Detection in the Cloud”에 제시된 알고리즘을 심도 있게 분석하고, 이를 현업에 적용 시 문제점 파악 및 이의 개선방안 고찰 • 구현 시스템 Architecture 구상 • 실험에 필요한 자료 획득 - 4.23 : Term Project 중간보고서 제출 - 4.24 ~ 5.10 : 제안 알고리즘을 시스템으로 구현 및 기존 방법들과 비교 <ul style="list-style-type: none"> • 제안 알고리즘을 프로그래밍 하여 실제 System 으로 구현 • 기존 논문에서 제안된 방법들과 비교 평가 • 비교 평가 결과 자료 정리 - 5.11 ~ 5.31 : 논문 및 PPT 작성 <ul style="list-style-type: none"> . 학회 양식에 맞추어 논문 작성 . 발표 용 PPT 작성 - 6. 1 ~ 6. 5 : 작성 논문 리뷰 및 수정 보완 - 6. 6 ~ 6. 9 : 논문 발표 Rehearsal - 6.10 : 논문 및 발표자료 제출 - 6.11 ~ 6.12 : Term Project 발표
-----------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------