빅데이터 기반의 스마트 팩토리와 CEP를 결합한 실시간 분석시스템 개발

이원창*, 이이섭**

Ⅰ. 서 론

1. 빅데이터

빅데이터는 최근 많은 기술 학문 분야에서 각광받고 있는 분석 기술의 한 분야 이다. 이 기술이 소개됨에 따라 분석하기 어려웠던 기존에는 여러 가지 데이터들에 대한 새로운 분석 및 해석이 가능하게 되었다. 이에 따라 다양한 분야에서 빅데이터 수집 및 분석기법을 사용하여 유의미한 분석 결과를 내놓고 있어 앞으로 활용가능성이 매우 높을 것으로 전망 된다. 그러나 현재의 빅데이터 솔루션은 고비용, 고용량의 분석시스템으로 인하여 중, 소기업에서는 그 활용성이 저조하다.

2. 실시간 분석 시스템

오늘날 제조업 환경은 3V(Volume, Velocity, Variety)속성의 빅데이터 환경으로 변화하고 있다. 따라서 이를 처리하기 위한 배치분석기반의 하둡 보다는 실시간 이벤트 분석 기반의 CEP 시스템 개발이 필요하다. 본 논문에서는 실시간 이벤트 분석을 위한 중, 소규모의 생산설비에 적합한 개선된 CEP 시스템을 제안한다. 제안한 CEP 시스템은 생산설비들과 결합하여 실시간 이벤트 데이터를 분석함으로서 체계적인 대처와 계획을 할 수 있는 시스템을 구축하고 그 실효성을 제시한다.

Ⅱ. 관련 연구

* 이원창, 연구원, 010-8805-0144, lwczzang61@gmail.com
** 이이섭, 금오공과대학교 컴퓨터공학과 교수, 054-478-7532, eesub@gmail.com

2.1 빅데이터(Big Data)

빅 데이터란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석하는 대량의 정형 또는 비정형 데이터 집합이다. 빅데이터 기술의 발전은 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동하게 한다. 또한 개인화된 현대 사회 구성원 마다 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 실현시키기도 한다. 이와 같이 빅데이터는 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게 가치 있는 정보를 제공할 수 있는 가능성을 제시하며 그 중요성이 부각되고 있다. 세계 경제 포럼은 2012년 떠오르는 10대 기술 중 그 첫 번째를 빅 데이터 기술로 선정 했으며 대한민국 지식경제부 R&D 전략기획단은 IT 10대 핵심기술 가운데 하나로 빅 데이터를 선정하는 등 최근 세계는 빅 데이터를 주목하고 있다.

2.2 하둡(Hadoop)

대량의 자료를 처리하는 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프리웨어 자바 소프트웨어 프레임워크인 하둡은, 분산처리를 지원하기 위해 개발되어 하둡 분산 파일 시스템(HDFS: Hadoop Distributed File System)과 맵리듀스를 구현, 아파치 루씬의 하부 프로젝트이다. 구글, 야후 그리고 페이스북등 글로벌 IT기업에서 핵심 기술로 사용되고 있으며, 그 외 많은 기업들이 자사 솔루션에 응용 및 활용하고 있다.

2.3 복합 이벤트 처리 기술(CEP, Complex Event Processing)

CEP는 여러 이벤트 소스로 부터 발생한 이벤트를 대상으로 실시간의 의미 있는 데이터를 추출하여 대응되는 액션을 수행하는 것을 말한다. 이때 이벤트 데이터는 스트림 데이터로써 대량의 지속으로 입력되는 데이터, 시간 순서가 필요한 데이터, 끝이 없는 데이터이다. 이러한 스트림 데이터는 전통적인 관계형 데이터베이스에서는 실시간 처리 분석이 불가능하다. CEP는 바로 이런 스트림 데이터를 실시간으로 분석하는 이벤트 데이터 처리 솔루션이다. 즉, 데이터베이스나 파일, 하둡에 저장하지 않고, 다양한 고속의 이벤트 스트림을 In-Memory 기반으로 초당 수백/수십만건의 실시간 처리가 가능하다.

Ⅲ. 본론

기존의 데이터베이스 기반 솔루션은 클라이언트에서의 요청, 서버에서 저장 후 분석된 결과를 응답하는 형식으로 선 저장, 후 분석의 기술이다. 그러나 CEP 기반 솔루션은 요청을 분석 후 클라이언트에게 응답, 서버에 저장하는 형식으로 선 분석, 후 저장의 기술이다.

그림1 과 같이 바코드 리더 등과 같은 센서에서 발생된 데이터를 데이터 수집 미들웨어 단계에서 처리 및 가공한다. 빅데이터 플랫폼 단계에서 처리 및 가공한 데이터를 활용할 다양한 방법으로 응용, 활용하게 된다.

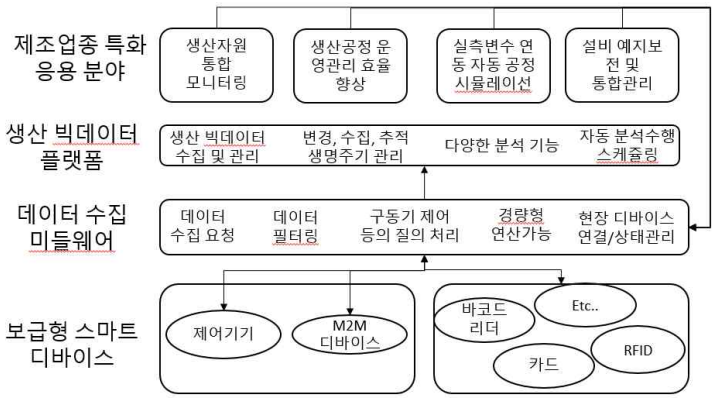

그림 1 시스템 아키텍처

센싱되는 데이터들을 효율적으로 관리하기 위하여 다음의 알고리즘을 사용하였다. 수신되는 데이터는 CEP엔진에 등록, 미등록 된 데이터거나 사용 불가능한 데이터가 포함된다. 사용 불가능한 데이터를 걸러내고, 등록되지 않은 데이터는 새롭게 생성, 추가하여 새로운 패턴의 분석을 하게 된다.

Algorithm1 Event Detection Algorithm

Input: Raw RFID events e, Event pattern p
Output: Events that satisfy the event query

```
check the state for classification
1: for all instance a in raw RFID events e do
2:       if a triggers the state of Event pattern then
3:          builder.setBolt(a)
4:       else if a triggers the state of trash then
5:                discard event a
6:                add the discard rate
7:       else create a new run r and
8:            add event a to the run r
9:            check if r already exist, add to data
10: end
```

그림 2 Event Detection Algorithm

아파치 스톰(Apache Storm)의 데이터 처리는 at least once원칙을 따르기 때문에 데이터의 중복이 발생할 수 있다. 발생하는 데이터의 빈도[1]에 따라서 분석 및 활용이 달라지기 때문에 중복성 또한 중요한 문제이다. 아파치 스파크(Apache Spark)의 경우 짧은 배치를 사용하여 ack가 유실될시 그 배치를 새로 전송하기 때문에 중복의 문제가 없다. 따라서 다음의 알고리즘과 같이 ack의 넘버링을 사용하여 데이터의 중복을 막는다.

Algorithm2 Duplication Prevent Algorithm

Input: Raw RFID events e, Event pattern p
Output: Events that satisfy the event query

Add the ackNum to InsertItem in their tree
1: **for all** instance a in raw RFID events e **do**
2: **if** a ackNum equals in tree **then**
3: discard event
4: **else** run Algorithm1
5: **end**

그림 3 Duplication Prevent Algorithm

Ⅳ. 결론

다양한 분야에서 빅데이터 수집 분석기법을 이용한 관련 산업에 대한 관심이 급증하고 있다. 하지만 고비용으로 인하여 중, 소기업에서는 구축이 힘든 것이 현실이다. 본 연구를 통해 빅데이터가 저비용으로 CEP 기반의 실시간 분석이 가능해지고 사용자의 요구에 더 빠르고 정확한 기능을 구현한다. 즉 효율인 생산설비 관리 환경을 제공하는 것이다. 향후 스마트 팩토리 생산설비 뿐만 아니라 선박, 물류 등의 많은 기기에서 사용할 수 있도록 사용자의 요구에 부합할 수 있도록 추가인 시스템 보완 등 지속인 연구가 필요하다.

참고 문헌

N.a Mao and J.ie Tan (2015), "Complex Event Processing on uncertain data streams in product manaufacturing process", Advanced Mechatronic Systems, Beijing, China, pp. 22-24.

이원창

금오공과대학에서 컴퓨터공학 학사학위를 취득하고 현재 금오대학교 컴퓨터공학과 석사과정 중이다. 관심분야는 소프트웨어 시스템, 빅데이터 등이다.

이이섭

서강대학에서 수학과 이학사를 취득하였으며, 전자계산학과 공학석사학위를 받았다. 고려대학교 컴퓨터과에서 이학박사를 취득하였으며 현재 금오공과대학 컴퓨터공학부 교수로 근무 중이다. 주요 연구 분야는 소프트웨어공학, 모바일컴퓨팅 등이다.