

퍼지 기반 퀴네-맥클러스키 규칙 감축 기법을 이용한 대용량 스트리밍 데이터의 고속 이벤트 탐지 기법 연구

박나영[○], 김정태^{*}, 윤희용^{*}

[○]을지대학교 의료IT마케팅학과

^{*}성균관대학교 정보통신대학

e-mail: parkny42@hanmail.net[○], {ktkim,youn}@ece.skku.ac.kr^{*}

Research on High-speed Event Detection based on Fuzzy Rule-based Quine-Maccluskey for Streaming Big Data

Na-Young Park[○], Kyung-Tae Kim^{*}, Hee-Yong Youn^{*}

[○]Dept. of Medical IT Marketing, Eulji University

^{*}College of Information and Communication Engineering, Sungkyunkwan University

● 요약 ●

최근 모바일 기기 및 무선기기의 발달로 인하여 센서 네트워크가 다양한 분야에서 응용되고 있다. 따라서 센서에서 실시간으로 발생하는 스트리밍 데이터에서 이벤트를 감지하고 분석하는 것은 중요한 연구 분야로 부각되고 있다. 단순 이벤트의 발생 조건을 빠르게 판별하기 위해 비트맵 인덱스 기반 복합 이벤트 검출 기법 등 여러 가지 방법들이 사용되고 있지만, 아직까지 이기종 센서에서 발생하는 각기 다른 형태의 데이터를 융합하여 이벤트를 검출하는 복합 이벤트 처리에 대한 연구는 미비한 실정이다. 본 논문에서는 각기 다른 형태를 가지는 스트리밍 데이터에 멤버십 함수를 적용하여 퍼지화 함으로서 이기종 센서에서 발생하는 데이터를 융합 처리가능하며, Quine-Mccluskey 감축기법을 통하여 규칙의 신뢰도 및 속도가 향상된 의사결정을 하는 고속 이벤트 탐지기법을 제안한다.

키워드: 스트리밍데이터(Streaming data), 이벤트 탐지(Event Detection), Quine Mccluskey 규칙 감축 기법(Quine Mccluskey Rule Reduction Techniques), 복합 이벤트 처리(Complex Event Processing)

I. 서론

최근 모바일 기기 및 무선기기의 발달로 인하여 센서 네트워크가 다양한 분야에서 응용되고 있다. 따라서 센서에서 실시간으로 발생하는 스트리밍 데이터에서 이벤트를 감지하고 분석하는 것은 중요한 연구 분야로 부각되고 있다.[1,2] 하지만 아직까지 이기종 센서에서 발생하는 다른 형태를 가지는 데이터를 융합 처리하는 연구는 미비한 실정이다. 본 논문에서는 다양한 형태의 데이터를 복합 처리하기 위해 퍼지 이론을 기반의 Quine-Mccluskey 감축 기법을 사용하는 이벤트 탐지기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 서술하였고, 3장에서는 이벤트 탐지 세 단계를 상세히 기술하였다. 마지막 4장에서는 결론을 도출하였다.

II. 관련 연구

1. 이벤트 기반 아키텍처

이벤트 기반 아키텍처는 서비스 기반으로 통합된 정보 시스템 구조에서 다양한 이벤트 상황을 바탕으로 이벤트 교환, 이벤트 트리거, 실시간 대응을 구현한다. 이벤트 기반 아키텍처는 현재 여러 다양한 분야에서 사용되고 있고 처리방법에 따라 단순 이벤트 처리, 스트림 이벤트 처리, 복합 이벤트 처리로 나눌 수 있다.

복합이벤트 처리는 여러 가지 센서에서 발생하는 복잡한 이벤트를 탐지하고 관리하기 위해 필요한 기술이다. 복합 이벤트 처리는 기업에서 발생하는 복잡한 이벤트를 이해하고 특정 이벤트가 무엇에 의해서 발생하는지 파악하고 그것을 바탕으로 대응하는 규칙을 만들어 실행하는데 그 목적이 있다. 즉, 복합 이벤트 처리는 어떠한 이벤트에 대한 대응을 하기 위한 기본적인 배경을 제공해주는 역할을 한다. 일반적인 복합 이벤트 처리에서 이벤트는 단순이벤트와 복합 이벤트로 구성된다. 단순 이벤트는 하나의 이벤트로 표현되며,

복합 이벤트는 단순 이벤트의 조합으로 표현할 수 있다. 예를 들어, RFID의 단순 이벤트는(T,L,ID)로 정의할 수 있다. 여기서 ID는 RFID 태그 ID를 나타내며, L은 위치, T는 발생하는 이벤트의 사건을 나타낸다. 이러한 기존의 복합 이벤트를 검출하는 기법으로는 SASE기법이 제안되었다.

SASE기법은 연속적으로 발생하는 대용량의 RFID 스트림 데이터에서 이벤트를 처리하기 위한 방법이다. SASE기법은 5가지의 처리 과정을 거치면서 이벤트의 적합성을 검사한다. 하지만 대용량 데이터 스트림에서 이벤트를 처리하는데 연산이 크게 증가하고 이로 인해 주어진 시간 내에 이벤트를 검출하지 못하여 시간 지연이 발생하는 문제점이 제기되고 있다.[8]

III. 본 론

이벤트 탐지 단계는 이기종 센서 데이터에 멤버십 함수를 적용한 데이터 퍼지화 과정, 규칙 감축 기법을 이용한 의사결정 과정, 데이터 패턴 융합 및 비-퍼지화 과정으로 나누어진다. 퍼지화 과정은 각 센서에서 여러 가지 다른 형태로 유입되는 데이터를 융합 및 복합 처리하기 위해 멤버십 함수를 통해 정형화된 속속도로 변환한다. 의사결정 과정에서는 속속도로 변환된 수치들을 각 항목별로 분류한 후 if-then 형식의 규칙을 이용하여 패턴들을 찾아낸다.[3] 이 과정에서 Quine-Mccluskey 규칙 감축 기법을 사용함으로써 규칙의 신뢰도가 향상된 고속 의사결정이 가능하다. 비-퍼지화 과정에서는 찾아낸 패턴들을 Centroid of gravity 알고리즘을 통해 융합하여 이벤트 발생 여부를 판단한다.

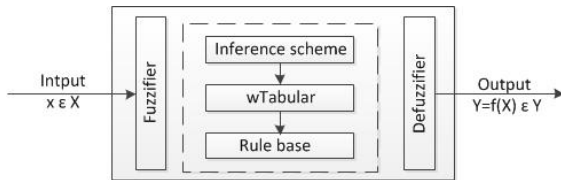


그림 1. 퍼지기반의 이벤트 탐지기 구조도
Fig. 1. The Architecture of Fuzzy Rule-based Event Detection

1. 이기종 센서 데이터에 멤버십 함수를 적용한 데이터 퍼지화 단계

퍼지화 단계는 서로 다른 이기종의 센서 데이터를 멤버십 함수를 통해 정형화된 속속도로 변환 하는 과정이다. 퍼지화의 첫 번째 단계는 입력변수를 주어진 퍼지시스템에 전달하는 것이다. 입력 변수는 사전에 정의한 입력 멤버십 함수를 통해 퍼지화된다. 멤버십 함수의 차수는 수평축에서 선택한 입력 변수를 삽입함으로써 결정되며, 수직축은 입력변수의 속속도를 나타낸다. 멤버십 함수를 충족하는 유일한 조건은 0과 1사이에서 변화하는 것이다. 0은 입력변수가 퍼지 집합의 소속이 아니라는 것을 의미하고, 1은 입력변수가 퍼지집합에 완전히 소속된다는 것을 뜻한다. 각 입력 변수는 대응하는 유일한 멤버십 함수를 가지고 있으며 이는 유효한 규칙을 이용해 입력변수

의 가중치를 구할 수 있다. 이 가중치를 통해 각 속속도(DOM)와 active rule의 영향도를 결정해 퍼지 출력반응 정도에 따른 퍼지 집합을 생성함으로써 이기종의 센서 데이터를 정형화된 속속도로 변환할 수 있다.

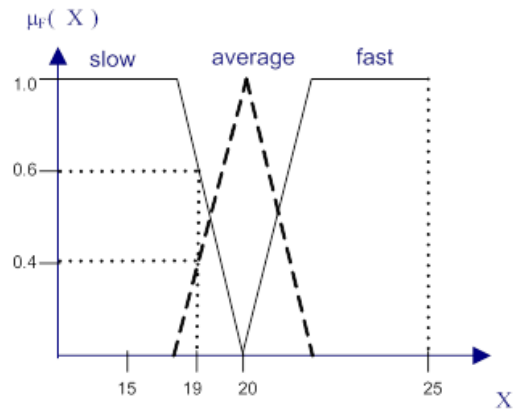


그림 2. 퍼지로지 멤버십함수의 예
Fig 2. The Example of Fuzzy Logic Membership Function

위 그림은 퍼지 변수속도에 대한 3개의 퍼지집합("slow", "average", "fast")의 멤버십 함수를 나타낸다. 전체 집합은 모든 가능한 속도 값을 생성한다. 예를 들어, X=19, 속도 값 19Km/h에 대한 퍼지집합은 'slow'의 속속도 0.6을 갖는다. 따라서 $\mu_{slow}(19) = 0.6$ 이며 $\mu_{average}(19) = 0.4$, $\mu_{fast}(19) = 0$ 으로 표현할 수 있다.

2. Quine-Mccluskey 규칙 감축 기법을 이용한 고속 의사결정 단계

대규모 데이터를 효율적으로 감축하고 규칙을 찾아내기 위해서 Apriori 알고리즘과 FP-growth 알고리즘이 주로 적용되고 있다. 스트리밍 데이터에서 이벤트를 감지하기 위해서는 빠른 시간 내에 변화에 강인한 데이터 처리가 중요하지만 Apriori 알고리즘은 여러 번 데이터베이스를 스캔하므로 많은 시간이 지연된다. FP-growth 알고리즘은 다양한 형태의 데이터 추가가 어렵다. 본 논문에서는 Quine-Mccluskey 방법을 규칙 감축기법에 적용하여 빠른 시간 내에 신뢰성 높은 데이터 처리가 가능한 wTabular 알고리즘을 연구했으며, 이를 통해 고속 이벤트 탐지가 가능하다. 가장 먼저 수집된 각각의 데이터를 각 데이터 속성별 기준을 통해 이진수로 표현한다. 그 후 데이터 집합을 전진부의 규칙의 이상치(1로 표현) 개수에 따라 오름차순으로 나열하며 이 중 이벤트 발생(후진부를 1로 표현)의 규칙을 추출하여 정렬한다. 정렬된 데이터 집합에 사전 확률을 계산하여 그에 따른 가중치(사후 확률)를 도출한다. 이때 최소 가중치에 부합 하지 못하는 규칙은 데이터 집합에서 제거한다. 다음 이렇게 정렬된 데이터 집합을 이상치의 개수가 n개인 규칙과 n+1인 규칙을 비교하고 don't care bit가 일치하는 규칙들을 결합한다. 이때 결합된 값은 don't care bit(-)와 care bit(이진)로 표시하며 결합을 통한 규칙 감축을 진행한다. 이는 규칙 감축이 더 이상 진행되지

않을 때 까지 진행되며 감축이 끝난 규칙들의 집합에서 중복 규칙을 제외한다. 예를 들어, 제안하는 방법을 사용하여 표1과 같은 임의의 데이터 집합에 대한 규칙 감소를 실행하면 표2와 같은 최종 규칙을 도출할 수 있다.

표 1. 임의의 데이터 집합
Table 1. A Set of Random Data

0	0	0	0	0
0	0	0	1	1
0	0	1	0	0
0	1	0	0	0
0	1	0	1	1
1	1	0	0	1
0	0	1	1	1
0	1	1	0	1
1	0	1	0	0
1	1	1	0	0
1	0	1	1	1
1	1	0	1	0
0	1	1	1	1

표 2. 최종 연관 규칙
Table 2. Final Association Rules

S_1	S_2	S_3	S_4	f
0	-	-	1	1

이렇게 생성된 규칙은 신뢰도와 지지도를 각각 식 (1)과 (2)를 통하여 계산하여 규칙의 연관성을 측정할 수 있다.

$$S(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

$$C(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

위에서 정의한 식에 따르면 표2에서 규칙 $S_4 \rightarrow f$ 의 경우 지지도 100% 신뢰도 100%로 나타낼 수 있으며 규칙 감소율은 92%로 계산할 수 있다.

3. 데이터 패턴 융합 및 비-퍼지화 단계

비-퍼지화 작업은 퍼지 집합을 요약하는 최종 결과 값을 찾는 것이다. 이를 위해 다양한 수학적 기법(중심, 이등분, 평균, 최대값 및 가중치 평균)이 사용된다. 그림3은 각 방법에 대한 값을 결정하는 그래프이다.

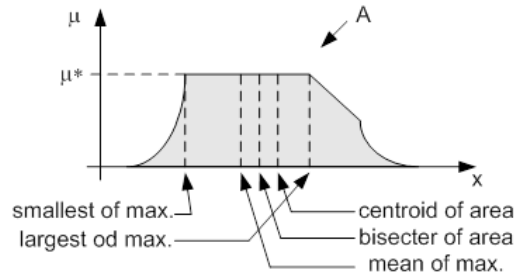


그림 3. 디퍼지화 방법의 그래프
Fig. 3. The Graph of Defuzzification Method

본 연구에서는 가장 정확도가 높은 Centroid of gravity 알고리즘을 사용하여 이벤트의 발생 여부를 판단한다. Centroid of gravity 알고리즘은 멤버십 함수의 곡선 아래 영역으로 중심을 결정함으로써 정확한 판단이 가능하다. 복잡한 멤버십 함수의 경우, 높은 수준의 계산이 요구되며 아래와 같은 수식으로 처리할 수 있다. 여기서 z_0 은 비-퍼지화된 출력값, u_i 는 멤버십함수, x 는 출력변수이다.

$$z_0 = \frac{\int \mu_i(x) x dx}{\int \mu_i(x) dx} \quad (3)$$

IV. 성능 평가

본 실험은 기존의 이벤트 탐지 기법인 BIED(Bitmap Indexing for Event Detection)과 CMMED(Contour Map Matching for Event Detection)를 비교 대상으로 하였다. 그림 4는 기존 스트리밍 데이터에서의 이벤트 발생 여부에 따른 탐지 시간을 평가한 결과를 보여 준다. 성능평가 결과 탐지되는 이벤트의 수가 적을 때는 큰 차이를 보이지 않지만 이벤트의 양이 증가함에 따라 연구된 기법의 탐지 시간이 가장 짧다는 것을 알 수 있다.

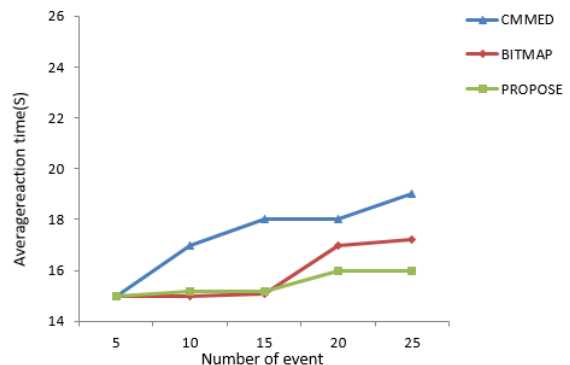


그림 4. 이벤트 탐지 시간 비교
Fig. 4. The Comparison of Event Detection Time

V. 결론

본 논문에서는 이기종 센서에서 발생하는 각기 다른 형태의 데이터를 융합하여 이벤트를 검출하기 위해서 퍼지 이론을 적용하였으며, 고속 융합처리를 위해 Quine-Mccluskey 감축 기법을 사용하는 이벤트 탐지기법을 제안하였다. 실험을 통한 성능평가 결과 이벤트의 양이 증가함에 따라 제안하는 이벤트 탐지 기법이 기존의 기법에 비해 탐지 시간이 가장 짧은 것을 확인할 수 있었으며, 이는 빠른 시간 내에 신뢰성 높은 데이터 처리가 필요한 대용량 스트리밍 데이터 처리 분야에 적합하다는 것을 의미한다.

ACKNOWLEDGMENT

본 연구는 한국산학연합회(C0017380), BK21+사업, 한국연구재단 기초연구사업 (2013R1A1A2040257), (2013R1A1A2060398), 미래부가 지원 한 2013년 정보통신·방송 (ICT) 연구개발 사업 (1391105003)의 지원을 받아 수행되었음.

참고문헌

- [1] W. Xue et al. "Contour Map Matching for Event Detection in Sensor Networks," Proceedings of the 2006 ACM SIGMOD international conference on Management of data, ACM, 2006.
- [2] C. C. Chang, W. E. Snyder, and C. Wang, "Robust Localization of Multiple Events in Sensor Networks," Proceeding of IEEE International Conference on SUTC, vol. 1, pp. 168-177, 2006.
- [3] K. Krasimira, S. H. Son, and K. D. Kang, "Using Fuzzy Logic for Robust Event Detection in Wireless Sensor Networks," Ad Hoc Networks, vol. 4, 2012.
- [4] W. S. Seol, H. W. Jeong, B. J. Lee, and H. Y. Youn, "Reduction of Association Rules for Big Data Sets in Socially-Aware Computing," IEEE BDSE 2013, 2013.
- [5] H. Malazi, K. Zamanifari, and S. Dulman, "Fed: Fuzzy Event Detection Model for Wireless Sensor Networks," International Journal, vol. 3, 2012.
- [6] D. J. Abadi, S. Madden, and W. Lindner, "REED: Robust, Efficient Filtering and Event Detection in Sensor Networks," Proceedings of the 31st International Conference on Very Large Data Bases, VLDB Endowment, 2005.
- [7] S. Li, S. H. Son, and J. A. Stankovic, "Event Detection Services using Data Service Middleware in Distributed Sensor Networks," Information Processing in Sensor Networks, Springer Berlin Heidelberg, 2009.
- [8] M. Mukesh et al, "Event Composition and Detection in Data Stream Management Systems", Proceedings of International Conference on Database and Expert Systems Applications, pp. 756-765, 2005.