

데이터마이닝 연관 기법을 이용한 전력계통 고장 해석

이준섭, 김민수, 최상열, 김철환, 김용모
 성균관대학교 전기전자및컴퓨터공학과

Analysis of Electric Power System Using Data Mining Association Rule

Joon-Sub Lee, Min-Soo Kim, Choi Sang-Yule, Kim Chul-Hwan, Ung-Mo Kim
 School of Electrical & Computer Engineering, Skungkyunkwan University

Abstract - Data Mining is a issue of Database fields. Data mining is discovered optimally interesting rules for user, which are results of specific requirements of user, through past data. Through to analyze and to statical suppose interesting rules, we can prepare future faults of system. In this paper, we present a new way which is discovered and repaired faults of Electric Power system using Data Mining techniques.

1. 서 론

전력계통의 사고에 대한 정의 및 이를 해결하기 위한 노력은 과거 많은 연구에서 진행되어 왔다. 관련연구를 유형별로 나누어 보자면 다음과 같다. 객체지향 기법을 기반으로 둔 것으로서 이는 사용자 인터페이스의 설계 [10], 데이터베이스 모델링 [11], 전력 시스템의 해석 [12], 전력, 조류 계산 [14] 등으로 구분된다. 다른 측면으로는 배전 시스템 고장 관리에 대한 것으로서 이는 고장 계산 방법 [15], 고장 진단 [13] 등으로 구분된다. 그러나 상기 연구들은 배전 시스템의 해석 및 모델링, 고장 계산 및 진단 등에 집중되어 왔다. 그러나 앞으로의 시스템이 거대화되어지고, 데이터의 양은 계속 증가 될 것이다. 그러나 기존의 시스템과 방식으로는 앞으로의 전력계통 시스템을 뒷받침하기에는 미흡할 수밖에 없다. 따라서 현재 데이터베이스 분야에서 중점 연구과제로 수행되고 있는 데이터 마이닝 기법을 이용, 수많은 데이터의 분석을 통하여 앞으로의 예상 및 같은 고장 혹은 사고가 일어날 것을 사전에 하루전이라도 준비할 수가 있다면 이는 경제적, 시간적 등 많은 부분에서 기업의 이윤을 늘릴 수가 있다. 또한 사용자에게는 신뢰성 있는 서비스가 가능할 것이다.

그러나 현재 존재하는 데이터마이닝 기법을 전력계통 시스템에 바로 적용하는데 있어서는 전력계통의 각각의 사고 유형의 성격 때문에 무리가 따른다. 그 이유는 전력계통의 개별적인 사건, 사고들에 대하여 일반적인 연관관계를 준다는 것이 힘들기 때문이다. 또한 전력계통 시스템의 고장에 대한 진단 및 복구는 일관된 사건에 대한 분석 및 복구가 아니라 단일 사건에 대한 진단과 복구가 이루어 졌다는데 기인할 수 있다. 따라서 마이닝 알고리즘을 사용하기 전, 기존의 데이터를 특정한 기준에 의해 서로에게 연관관계를 지어주어야 할 필요성이 존재한다. 본 논문에서는 전력계통에 각각의 사건들에 대해 시간이라는 인수를 두어 서로에게 연관관계를 주도록 고안되었다.

본 논문의 구성은 2.1장에서는 마이닝 알고리즘의 기본적인 지식들에 대하여 간단히 소개를 한다. 2.2장에서는 새로운 알고리즘에 대한 소개와 그 효용성에 대하여 논한다. 2.3장에서는 현재의 연구 진행 상황과 앞으로의 진행 방향에 대하여 제시한다. 3장 결론에서는 새로운 알고리즘을 통한 기대효과와 앞으로의 발전방향에 대

하여 소개한다.

2. 본 론

2.1 배경지식 & 문제점

본 논문은 데이터 마이닝 기법 중 Apriori 알고리즘 [2]에 기반을 두고 있다. Apriori 알고리즘은 Association Rule(1)을 발견해 내는 알고리즘 중 가장 최적화 되어있으며, 현재 많은 부분에서 사용이 되고 있다. 2.1 배경지식 & 문제점에서는 그러한 Apriori 알고리즘에 대한 전반적이 소개와 데이터 마이닝 기법을 직접적으로 전력계통에 적용하는데 나타난 문제점이 무엇인지에 대하여 분석해 보려한다.

2.1.1 Association Rules 개요

전력계통에서 발생하는 사건, 사고들을 item이라 할 때, Transaction이란 각각의 item들의 집합으로 구성되어지는 하나의 item set이라 볼 수 있다. 간단한 예로 item X, Y가 하나의 XY라는 Transaction을 이루고 있다고 할 때, Association Rule(연관규칙)은 $X \rightarrow Y$ 로 표현된다. 직관적으로 Association Rule을 관찰하여보면, 하나의 Transaction에 X와 Y가 함께 포함되어 있다고 말할 수가 있다. 이해를 돕기 위해 Association Rule에 대한 간단한 예 [1]를 들어보면 다음과 같다. Transaction을 이루고 있는 전체 테이블 중, 30%가 맥주(beer)를 포함하고 있고, 또한 기저귀(diaper)를 포함하고 있다. 이중 2%가 양쪽 모두를 포함하고 있다. 여기서 30%를 Rule의 Confidence라고 하고, 2%를 Rule의 Support라고 한다. 여기서 Association rule의 문제점은 사용자의 min Confidence와 min Support에 맞는 모든 Rule를 찾아야 한다는데 있다. 이는 사용자에게 유용하지 않은 Rule또한 찾아진다는 문제점과 유용한 Rule을 발견하는데 있어 noise와 같은 역할을 한다. 따라서 기존의 연구 [2]들은 이러한 문제점을 해결하려는데 초점을 맞추고 있다.

위의 문제들에 대하여 최적화된 기법중의 하나가 Apriori 알고리즘이다. 우리는 여러 가지 연관규칙 알고리즘 중, 효용성과 성능면에서 이미 입증된 Apriori 알고리즘을 사용하였다. 대부분의 연관규칙의 알고리즘들은 아래에서 소개할 Apriori 알고리즘을 기반으로 작성되었다.

2.1.2 Apriori 알고리즘

- 1) $L_1 = \{ \text{large 1-itemsets} \};$
- 2) **for** (k = 2; L_k-10 ; k++) **do begin**
- 3) $C_k = \text{apriori-gen}(L_{k-1});$
//New candidates
- 4) **forall** transactions t D **do begin**
- 5) $C_t = \text{subset}(C_k, t);$
// Candidates contained
- 6) **forall** candidates $c \in C_t$ **do**

```

7)      c.count++;
8)      end
9)      Lk = { c ∈ Ck | c.count ≥ minsup }
10)     end
11) Answer = ∪k Lk

```

(Fig. 1 Apriori 알고리즘)

(Fig.1)은 Apriori 알고리즘을 보여주고 있다. Apriori 알고리즘은 frequent item 집합을 찾기 위해 multiple path를 사용하고 있다. k번째의 path에서 알고리즘은 k개의 모든 아이템 집합을 찾게 되며 이를 k-item 집합이라 한다. 각각의 pass는 두 가지의 단계로 이루어져 있다. 우선 F_k를 frequent item 집합이라 정의하고, C_k를 후보 item 집합- frequent item 집합이 될 수 있는 잠재성을 가진 집합-이라 정의를 하자. 첫 번째 단계는 (k-1) item 집합으로부터 후보 생성 단계이다. F_{k-1}을 (k-1) pass에서 찾은 frequent item 집합이라고 한다면, 이것을 이용하여 후보 집합인 C_k를 생성한다. 후보 생성 단계의 절차는, 후보 집합인 C_k에 대하여 모든 frequent item 집합의 superset임을 보장한다. 특별히 C_k는 메모리상의 hash-tree 구조를 이용하여 저장된다. 그리고 나서 생성되어진 C_k는 다음단계인 support counting 단계를 거치게 된다. 각각의 transaction마다 transaction속에 포함되어져 있는 C_k에 대한 후보들에 대하여 hash-tree 구조와 그들의 support count를 사용하여 증가하게 된다. 마지막 pass에 도달하게 되면, 각각의 C_k에 대하여 과연 그 후보 집합이 frequent item 집합 즉 F_k가 될 수 있는 지에 대하여 검사하게 된다. Apriori 알고리즘은 F_k 또는 C_{k-1}이 비워지게 되면 종료하게 된다.

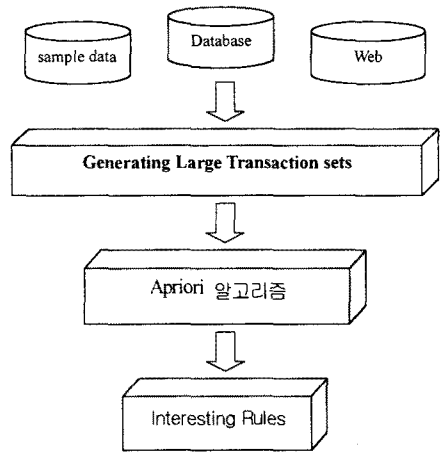
2.1.3 입력 데이터 형식

Transaction 테이블 T는 두 개의 컬럼을 가지고 있다. 하나는 transaction 정의부(tid)이고 다른 하나는 item 정의부(item)으로 정의 할 수 있다. tid에 대한 item의 수는 변수이며, 테이블이 생성되는 동안에는 알 수가 없다. 하나의 tid에 모든 item들은 하나의 단일 튜플의 컬럼에 나뉘며, 다른 튜플에서도 나타날 수 있다. transaction에 대한 item들의 개수는 DBMS가 지원하는 최대 컬럼의 개수보다 많이 나타는 것을 종종 볼 수 있다. 간단한 실례로 우리의 실세계의 데이터베이스의 최대 각 transaction당의 item들의 개수는 872이거나 다른 경우에는 700정도 밖에 되지 않음을 볼 수가 있다. 정리하자면 transaction당 item들의 평균 개수는 단지 9.6에서 4.4 정도 밖에 기대 될 수밖에 없다. 정리해보면 하나의 일련의 사건에 대하여 약 5개에서 10정도의 사건정도가 파악 가능한 것이다. 그러나 사용자의 min conf.나 min sup.의 설정에 의해 이는 더 많은 수의 item들도 하나의 transaction에 포함될 수 있기에 앞으로의 연구에는 크게 지장을 주지 안 수 있다.

2.1.4 문제점

위에서 간략하게 Association Rule과 그 중 최적화된 하나의 기법인 Apriori 알고리즘에 대하여 알아보았다. 위에서 볼 수 있듯이 Apriori 알고리즘은 Transaction이 생성되어 있는 상태에서 적용이 가능하다. 그러나 전력계통의 사고를 면밀히 분석해 보면, 각각의 사고에서 특별한 연관관계를 찾을 수 없다는 것을 발견할 수 있다. 기본적으로 전력계통에 문제 혹은 사고가 발생하면 각각의 사고에 대한 처리와 복구가 이루어지기 때문에, A라는 사고로 인하여 B라는 사고가 일어났다고 규정짓기가 상당히 힘든 측면이 있다. 물론 그동안 전력계통 분야의 연구에서, 그러한 사고들에 대해 일반적인 정의와 각각의 사고에 대한 원인 분석 및 예방책이 없었던 것은 아니다. 그러나 앞으로 전력계통

의 시스템이 거대화되고 복잡해져간다면 각각의 사건, 사고들에 대하여 통합 관리한다는 것은 어려워지게 될 것이다. 또한 모든 관리 시스템을 자동화해 가는 데 있어서도 많은 제약 사항이 따르게 되며, 현재와 비슷한 방식인 고장 발생에 대해 그 자체로서 복구하는 방식으로는 미래의 전력계통제어에 심각한 문제를 초래 할 가능성도 있다. 따라서 데이터 마이닝을 이용하여 고장 진단 및 복구에 사용한다면, 지난 기간의 데이터를 이용하여 앞으로의 사고에 대하여 미연에 방지하고, 그 동안 간과되거나 무시되었던 새로운 연관관계를 발견가능 할 것으로 보이기에, 데이터 마이닝 기법의 도입의 필요성이 대두된다. 그러나 처음에도 언급했듯이 Apriori 알고리즘은 기본적으로 transaction이 존재하고 있는 상태에서 이용할 수가 있기에, 전력계통의 사고처럼 일정한 관계를 규정 지을 수 없는 사고들, 즉 아무런 연관관계도 없는 데이터를 이용하기 위해 새롭게 transaction들을 생성해야 할 필요성이 있다.



(Fig.2 알고리즘 흐름도)

(Fig.2)는 Transaction 집합 생성 단계의 위치를 나타내고 있다.

2.2 Transaction 집합 생성

기본적으로 Transaction 집합 생성은 전력계통의 사고가 서로 연관성을 크게 찾아 볼 수 없다는데 기인한다. 단일적인 고장이 주로 이루어지며, 혹 연관된다 하더라도 그 관계는 극히 일부분으로 나타내지고 있다. 그러나 시스템이 거대화 되어간다면, 서로 다른 지역에 속한 고장들간의 연관관계를 갖는다는 것이 너무나 어렵게 된다. 간단한 예를 들어 설명해 보자. A라는 지역에서 고장이 발생했다고 가정하자. 그렇다면 A라는 지역을 관리하는 전력관리시스템에서 이를 복구 등의 조치를 취하게 될 것이다. 그러나 그와 비슷한 시간에 그동안 전력계통의 이론상으로는 발견이 되지를 못했던 연관관계에 의해 B라는 지역에서 문제가 발생할 수도 있다. 따라서 각각의 사건, 사고들을 연관지어 새로운 Transaction 집합을 생성해야 할 필요가 있다. 아래의 (Fig.4)은 알고리즘에 사용되는 상수들을 정의하고 있고, (Fig.3)는 Transaction 생성 알고리즘을 보여주고 있다.

```

1) LE = {events};
   //각각의 event들은 서로 독립적인 사건
2) Let LT = 0; // Large Transaction sets
3) for(m = 0 ; m=m-1; m++)
4) for(f=0; f=(발생된 Event의 갯수)in TH; f++)
5) Temp = make transaction set;

```

```

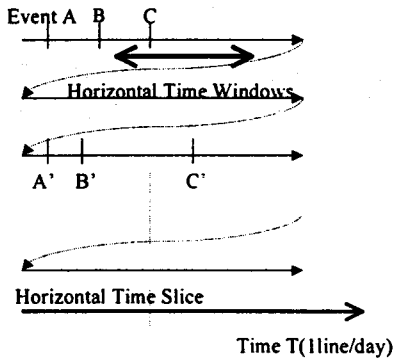
//as like LT+LT+1* (ex. AB, ABC, ABCD*)
7)if (Temp ∈ LT)
8)LT.count++ where Temp = LT.name;
9)else
10)Add Temp to LT.tranname & set LT.count=1;
11)End
12)Using The Apriori Algorithm
(Fig. 3 Transaction 데이터 생성 알고리즘)

```

상수명	정의
L_E	생성전의 데이터 소스
L_T	새롭게 생성된 Transaction 데이터
$L_{T.name}$	L_T 테이블의 컬럼(transaction 이름)
$L_{T.count}$	L_T 테이블의 컬럼(transaction 발생횟수)
m	L_E 데이터 소스에 속한 모든 item들의 갯수
T_H	사용자 정의 시간범위

(Fig.4 상수 정의)

위의 알고리즘의 역할은 존재하지 않는 연관관계를 시간이라는 사용자 정의 상수를 두어 연관규칙을 만들어 내는데 있다. 즉 일정한 시간 간격내에 위치하는 사건들을 서로 연관이 있다는 가정하에 데이터를 분석하게 되는 것이다. 가정되어진 사건들의 패턴이 축적되어있는 데이터 속에서 다수 발생을 하게 된다면, 이는 무시할 수 없는 각 사건의 연관관계라 볼 수 있게 된다. (Fig.5)가 간단한 예를 나타낸 그림이다.



(Fig.5 알고리즘 개요)

위 그림에서 보듯이 각각의 독립적인 사건 ABC가 하나의 시간 선 안에 존재를 한다. A라는 사건에 대하여 사용자가 정의 한 시간의 범위가 시작되게 된다. 그 시간 범위 안에 B혹은 C라는 사건이 포함이 된다고 가정하자. 만약 다른 시간 선 안에 A와 같은 사건인 A'가 발생을 하게 되고 같은 시간 범위 안에서 B',C'가 발생을 하게 된다. 이러한 Transaction이 다수 발생을 하게 된다면 A라는 사건이 일어나게 되면, B라는 사건이 일어나게 되고, 이어서 C라는 사건도 발생하게 된다는 규칙이 발생하게 된다. 즉 A→B→C라고 정의 될 수 있다.

2.3 연구 진행

현재 본 논문에서 제시된 알고리즘과 전력계통에서의 적용도에 대해서는 연구 진행 중에 있다. 가장 큰 초점은 기존의 전력계통 사고들에 대한 정의의 일수도 있지만, 큰 초점은 그 동안 발견하지 못했던 각 사고의 연관관계를 추정해 내는 데는 것이 더 중요한 목표로 삼고 있다. 또한 현재의 과정에는 Transaction을 생성하는 단계와 Apriori 알고리즘을 적용하는 단계로 이루어져 있다. 그러나 두 단계를 사용함으로써 결과 도출 시간등에

서 효율이 떨어진다고 본다. 따라서 현재 두 알고리즘을 하나의 알고리즘 속에서 구현하려는 시도가 계속되고 있다.

3. 결론

본 논문에서는 데이터마이닝 기법을 전력계통 시스템의 고장 해석에 이용하려는 연구에 대한 고찰 및 소개이다. 마이닝 기법을 통한 데이터 분석결과는 그 이용범위가 무한하다고 본다. 특히 전력계통 시스템 상의 고장 진단 및 복구에 새로운 지평을 열 수 있는 방법이라 생각된다. 가장 중요한 것은 기존의 알지 못했던 각 사고들간의 연관관계를 정의 할 수 있으며, 이러한 각 사고들을 분류하여 앞으로 발생할 사고에 대한 미연의 방지와 신속한 대처를 가능하게 한다는데 있다.

(참고 문헌)

- [1] R. Agrawal, T. Imielinski, and A. Swami., "Mining association rules between sets of items in large database", ACM-SIGMOD, pp 207-216, In Proc. 1993
- [2] R. Agrawal, R. Srikant., "Fast algorithms for mining association rules", VLDB, pp 487-499, In Proc. 1994
- [3] Robert Cooley, Bamshad Mobacher, Jaideep Srivastava., "Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, In Proc. 1999
- [4] Brian Lent, Arun Swami, Jennifer Widom., "Clustering association rules", ICDE, In Proc. 1997
- [5] M. Klemettinen, H. Mannila, P. Ronkainen, and H. Toivonen, "Finding interesting rules form large sets of discovered association rules", In 3rd International Conference on Information and Knowledge, November 1994
- [6] Heikki Mannila, Hannu Toivonen, and A. Inkeri Verkamo, "Efficient algorithms for discovering association rules", KDD-94, AAAI Workshop on Knowledge Discovery in Database, July 1994
- [7] G. Piastestsky-Shapiro, "Discovery, analysis, and presentation of strong rules", In G.Piastestky-Shapiro, editor, Knowledge Discovery in Database. AAAI/MIT Press, 1991
- [8] C. Schaffer, "Domain-Independent Function Finding", PhD thesis, Rutgers University, 1990
- [9] 김용모, "객체지향모델을 이용한 배전자동화시스템의 능동적 관리에 관한연구", 기초전력공학연구소, 1998.9
- [10] M.Foley, et. al, "An Object Based Graphical User Interface for Power System", IEEE Transaction on Power Systems, Vol.8, No.1, pp.97-194, February 1993
- [11] S. R. Choi, M. C. Shin, U. M. Kim, H. and M. Kim, "Object-Oriented Applications to EMS Database Design", International Conference on Electrical Engineering (ICEE), Vol.2, pp.352-256, 1998
- [12] 김건중, "객체지향개념의 ODBC 시스템을 이용한 전력계통해석 데이터베이스 구축", 충남대학교 석사학위논문, 1995
- [13] 이재용, 백영식, "객체지향 프로그래밍 기법을 이용한 전력계통의 고장진단 시스템에 관한 연구", 대한전기학회 논문지, Vol.44, No.4, 1995
- [14] 김정년, "객체지향 기법을 이용한 전력 조류계산", 경북대학교 전기공학과, 석사학위 논문
- [15] 이재용, 김주용, 백영식, "4단자망을 이용한 고장 계산에 관한 연구", 대한전기학회 논문지, Vol.43, No.4, 1994