

VOC 기반 연관규칙 마이닝을 이용한 통신선로설비의 장애 예측*

나 기 주** · 한 인 섭** · 조 남 욱***

Fault Prediction of a Telecommunications Network using Association Rules Mining based on Voice of the Customer

Na Gijoo · Han Insup · Cho Namwook

〈Abstract〉

Customer complaints handling helps organizations to retain existing customers and attract new customers, as well. As Voice of the Customer (VOC) is one of the main sources of customer complaints, many organizations utilize VOC to enhance customer satisfaction. Effective management of VOC has been proved as one of the best ways to maintain organization's brand image and reputation. In spite of its importance, little has been reported on the utilization of VOC to detect faults in a telecommunication industry. In this paper, association rule mining based on VOC is used to identify root fault causes of a telecommunications network. To do that, VOC of a Communication Service Provider has been collected first. Then, association rule mining has also been conducted with various support and confidence levels. As a result, root fault causes of the telecommunications network can be identified. It is expected that this study can be used as a basis for decisions about customer satisfaction management such as preventive maintenances or reduction of the customer maintenance cost.

Key Words : Voice of the Customer, Association Rules, Fault Prediction, Telecommunications

I. 서론

기업의 서비스와 상품에 대한 고객 불만이 지속적으로 발생하게 되면 충성고객의 이탈과 기업 평판(Reputation)의 하락을 불러올 수 있다. 이는 매출감소로 이어질 수 있는 위험한 신호이기 때문에 많은

기업들은 고객 불만을 해결하고 소통하기 위한 창구로써 고객의 소리(Voice of the Customer, 이하 VOC)를 운영한다. 특히 통신사가 제공하는 서비스는 최종 소비자에게 전달되기까지 매우 복잡한 가치사슬(Value Chain)을 거친다. 통신사는 일정한 기간 동안에 축적된 VOC 데이터를 이용하여 고객 불만을 조기에 파악하여 대처할 필요가 있다.

본 연구의 대상으로 선정된 A통신사에서 1년간 발생하는 VOC 현황을 살펴보면 요금에 대한 불편 신

* 이 연구는 서울과학기술대학교 교내 학술연구비 일부지원으로 수행되었습니다.

** 서울과학기술대학교 IT정책전문대학원 박사과정

*** 서울과학기술대학교 글로벌융합산업공학과 교수(교신저자)

고가 43%, 상품에 대한 불만 및 문의 사항이 34%를 차지하고 있으며, 서비스 사용의 불편신고가 17%를 차지한다. 이렇게 발생하는 VOC 중에 통신사용 불편에 대한 VOC를 분석하면 네트워크 중 가입자망에서 발생하는 VOC가 80%-90%를 차지하고 있다. 통신설비장치는 많은 비용과 시간이 투자되어야 하는 복잡한 체계이기 때문에 중단 없는 서비스 제공을 위해서는 장애상황을 빠르게 파악하기 위한 관리체계와 장애예방을 위한 사전점검이 중요하다. 잦은 서비스 중단은 통신시장의 경쟁상황에서 고객의 이탈을 가속화시키는 원인이 된다.

통신장애는 시스템 장애, 회선 장애, 가입자 시설 장애 등으로 나눌 수 있다. 시스템 장애는 회복시간이 수분 이내로 빠르기 때문에 VOC가 많이 발생되지 않는 반면, 회선이나 가입자 시설의 경우 장애 원인 파악에 상대적으로 긴 시간이 소요되기 때문에 VOC가 많이 발생하는 경향이 있다. 지하시설물의 케이블과 선로의 인입선 장애, 덕내 시설인 구내단자, 실내선 장애 등이 회선장애로 분류된다. 본 연구에서는 A 통신사가 운영하는 실제 VOC의 데이터를 수집한 다음 연관규칙 마이닝을 이용하여 통신선로설비의 회선 장애, 가입자시설 장애를 중심으로 장애원인을 진단함으로써 통신 서비스 품질 저하를 효과적으로 방지하고 고객 만족도를 개선할 수 있는 사전점검 방안을 제시하고자 한다.

논문의 구성은 다음과 같다. 2장에서는 고객민원과 연관규칙 마이닝과 관련된 선행연구와 연구이론을 살펴보고 통신선로설비에서 사용하는 구조적인 개념을 제시하였다. 3장에서는 연관규칙을 통해 도출하고자 하는 분석 데이터 변수의 선택과 분석 방법에 대해서 설명한다. 4장에서는 연관규칙을 통해 분석된 규칙의 결과에 대해서 의미를 해석하고 시사점을 도출하며, 마지막으로 5장에서는 결론과 함께 본 연구의 한계점과 추후 연구방향에 대해 논한다.

II. 관련 연구 및 연구 이론

2.1 관련 연구

VOC는 기업의 제품이나 서비스에 하자가 발생하거나 약속한 서비스 수준이 만족되지 않을 때 고객이 기업과 소통하는 창구로 활용되고 있다. IT기술의 발달로 대부분의 기업은 고객과의 소통을 위해 콜센터, 온라인게시판, 이메일, SMS, 소셜미디어 등을 운영하고 있다. 다양한 채널을 통하여 수집된 VOC 데이터는 기업이 판매하는 제품과 서비스의 품질을 분석할 수 있는 귀중한 자원으로 활용된다. 콜센터, 이메일, SMS 등에서 수집된 구조화되지 않은 정보를 비즈니스 정보와 연결하면 고객의 선호도 및 트렌드와 관련한 규칙을 찾아낼 수 있다[1].

Jain[2]는 VOC에 저장된 데이터를 연관규칙 마이닝과 텍스트 마이닝을 이용하여 의사결정에 필요한 정보를 추출하였다. 통신사에서는 효율적인 망 장애 운용관리를 위해 근원장애 분석 정보를 이용하여 네트워크의 장애를 진단하는 시스템 구현에 VOC를 이용하였다[3]. VOC는 기업뿐만 아니라 공공, 의료 등의 서비스 조직에서 서비스 제공 결과의 피드백을 수렴하기 위한 목적으로 운영되고 있기 때문에 데이터 마이닝 기법을 이용하여 제품구매 만족도와 서비스 만족도를 분석하는데 유용하다[4-5].

본 연구에서 VOC 분석에 사용된 연관규칙 마이닝은 전자상거래의 상품추천 서비스나 정보 추천시스템 구현, 구매형태 분석, 통신 시스템의 장애원인 분석 등과 사회과학 연구를 위한 만족도 연구를 위해 이용되었다[6-10]. 또한 네트워크 서비스 시스템의 위험진단, 상품의 기대수익, 이탈 고객의 예측, 고객행위의 예측, 제조 공정의 불량항목 예측, 성능 최적화의 주요 분석기법으로 활용되었다[11-16].

여러 분야에서 연관규칙 마이닝을 이용한 선행연

구가 존재하나 통신사의 VOC를 분석하여 이용자의 통신선로시설에 관한 장애를 예측하는 연구는 부족한 실정이다.

2.2 연관규칙

연관규칙(association rules)은 상품추천이나 명목 척도로 측정된 연구 자료의 분석과 예측 연구 등에 광범위하게 사용되는 비지도 학습(unsupervised learning) 방식의 데이터마이닝 기법으로 친밀도 분석(affinity analysis) 혹은 장바구니 분석(market basket analysis)으로도 불린다. 연관규칙은 항목들 사이의 모든 규칙들을 조건부(antecedent)와 결론부(consequent) 형식으로 나열한 후 조건부와 결론부의 항목집합(item sets) 사이에서 가장 빈번하게 나타나는 강한 규칙들을 찾기 위한 것이다. 여기서 빈도수가 높은 항목집합들을 빈발 항목집합(frequent item sets)이라고 하며 빈발 항목집합을 찾기 위해 일반적으로 Agrawal[17]가 제안한 Apriori 알고리즘이 이용된다[18].

빈발 항목집합을 결정하기 위해서는 지지도(support), 신뢰도(confidence), 향상도(lift)의 척도를 사용하며 각 척도를 설명하기 위해 통신사의 고객센터의 발생건수로 예를 들어 설명하면 다음과 같다. 한 통신사에서 1년 동안 접수된 전체 VOC가 10,000건이고 A지역 B시설의 장애 발생건수가 1,000건일 때 이 중에서 장애원인이 C로 밝혀진 것이 500건이라고 가정해 보자. A와 B의 동시 발생으로 인한 결과가 C라고 할 때 A, B => C의 항목집합을 <식1>과 같이 표현하고 <식1>의 지지도와 신뢰도는 각각 <식2>와 <식3>으로 정의한다.

$$\begin{aligned} \text{조건부}(A \cap B) \Rightarrow \text{결론부}(C) & \quad <식1> \\ \text{지지도} = P(A \cap B \cap C) & \quad <식2> \\ \text{신뢰도} = P(A \cap B \cap C) / P(A \cap B) & \quad <식3> \end{aligned}$$

따라서 <식1>의 지지도는 0.05(5%=500/10,000)이며 신뢰도는 50%(=500/1,000)가 된다. 지지도와 신뢰도의 관계에 있어서 지지도보다는 신뢰도가 높을수록 크게 신뢰할 수 있다. <식1>에서 조건부와 결론부가 독립적이라고 가정할 때 향상도는 <식4>와 같이 정의한다.

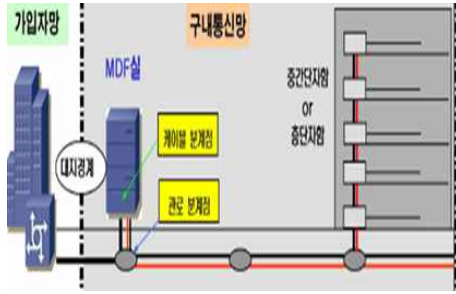
$$\text{향상도} = \frac{P(A \cap B \cap C)}{P(A \cap B) \cdot P(C)} \quad <식4>$$

전체 VOC 중 C의 발생건수가 2,500건 일 때 식1의 향상도는 2(=0.05/(0.1*0.25))가 된다. 향상도가 1보다 클 경우에는 조건부와 결론부의 동시발생 확률이 조건부와 결론부가 독립적인 때 기대할 수 있는 확률보다 더 크기 때문에 향상도가 클수록 해당 규칙이 더 유용하다[17].

2.3 통신선로설비의 구조

본 연구의 목적은 연관규칙 마이닝을 이용하여 통신선로설비에서 발생하는 장애의 원인을 진단하고 예방 대책을 수립하기 위한 것이다. 통신선로설비 구조의 이해가 필요하다.

<그림 1> (a)의 구내통신선로설비의 구조에서 분계점은 설치관리 책임을 구분하기 위한 경계이다. 이는 관로시설과 케이블시설로 나뉘며 통신사의 국선과 가입자망의 경계는 <그림 1> (b)와 같이 사유지와 사유지의 경계블록으로 나뉜다. <그림 1> (a)에서 MDF(Main Distributing Frame)는 건물 내부의 층간 중간단자함 IDF(Intermediate Distributing Frame)에 연결되는 주단자함으로 가입자는 최종적으로 IDF와 통신기기를 연결하여 사용한다. 관로시설은 사설 측 부지의 경계이며, 케이블시설은 사설 측 구내 최초단자함의 국선용 단자까지 해당된다. 구내통신선로설비 중 이용자 시설은 원칙적으로 이용자(건물주)가 유지보수를 해야 하지만 이용자의 편의를 위해 통신사가



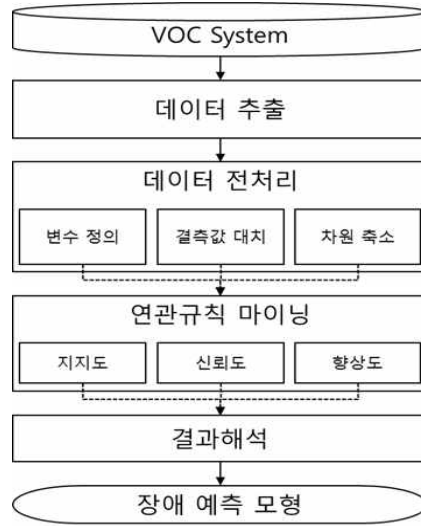
(a) 구조



(b) 경계

<그림 1> 통신선로설비의 구조 및 경계

과를 해석하고 사전예방과 장애예측을 위한 모델을 제시하였다.



<그림 2> 연구 모형

고객단말장치까지 정비한다. 통신선로설비의 구축 책임에서 사업자는 국선용 단자함의 국선접속설비(국선 수용단자반, 보호기), 국선접속설비까지의 케이블, 인입선과 부지경계까지의 관로시설 관리를 책임지며 이용자는 인입인 수공 및 접속함 설치, 구내측 관로, 배관시설 그리고 국선용단자함, 구내배관, 구내케이블, 배선, 접지시설 등을 책임진다.

III. 연구방법론

본 연구의 연구모형을 <그림 2>에 제시하였다. 먼저 통신사 VOC시스템으로부터 분석대상 데이터를 추출하여 전처리 후 분석 모델링을 위한 변수 선택, 결측치 대치, 차원 축소를 실시하였다. 데이터 전처리 후 연관규칙 마이닝으로 도출된 빈발 항목집합의 결

3.1 데이터 추출 및 전처리

본 연구를 위해 수집된 데이터는 2013년 1월 1일부터 2013년 7월 31일까지 A통신사 특정 서비스 지역의 고객센터로 접수된 VOC로 수집 건 수는 약 700건이다. 일반적으로 통신사 VOC 데이터에는 가입자 정보와 함께 VOC등록일, 관할지역(Region), 망소유(Owner), 케이블((Cable), 통신선로시설(Facility), 장애원인(Cause), 수리내용(Repair) 등이 포함되어 있다. 본 연구에서는 데이터마이닝 기법을 통해 장애가 빈번하게 발생하는 통신선로구간, 시설, 발생원인 및 조치사항과 연관된 규칙을 도출하는 것이 목적이다. 따라서 구조화된 통신선로시설 관련 정보만을 분석 데이터로 선정하여 전처리를 하였고 가입자 정보와 비구조화된 자연어 항목은 추출하지 않았다.

<표 1> 데이터 구조

변수명(종류)	속성값(변환값)	형태
Region(2)	Y지역(yj), S지역(ss)	명목
Owner(2)	통신사(sp), 사설(pv)	"
Cable(10)	7, 11, 12, 13, 14, 16, 26, 31, 51, 78	"
Facility(9)	인입선(314), MDF주단자함(315), 회선보호기(316), 시내단자함(317), 케이블지하(318) 전화기등(166), 콘센트(311), 실내선(312), 구내단자함(313)	"
Cause(5)	시설노후(121), 공법불량(122), 절연불량(123), 접속불량(124), 단자불량(304)	"
Repair(12)	교체(111), 이물질제거(112), 재접속(113), 선로재구성(115), 회선정비(116), 습기제거(117), 수리의뢰구내기사(121), 수리의뢰고객(122), 점퍼재구성국사내(302), 점퍼재구성국사외(303), EN교체(304), 멀티해체(307)	"

3.2 변수 선택 및 모델링

<표 1>과 같이 추출된 VOC 데이터의 변수들은 다양한 속성값을 가지고 있으며 속성값 모두 명목척도로 되어있다. 명목척도에서 유용한 정보를 도출하기 위해서 데이터마이닝 기법에서 연관규칙을 사용한다. 2장에서 살펴본 바와 같이 연관규칙을 통해 통신선로 시설의 장애구간 및 원인과 관련한 빈발 항목집합을 찾아내기 위해서는 <표 2>와 같이 조건부와 결론부에 해당하는 변수를 선택해야 한다.

<표 2> 변수 선택

조건부	결론부
Region(지역)	Cause(장애원인) Repair(수리내용)
Owner(소유구분)	
Cable(설치케이블)	
Facility(선로시설)	

변수값에 이상치(outlier)나 결측치(missing value)가 포함된 경우 제거하거나 대체해야 한다. 상관관계

가 높은 변수들이 분석대상에 포함될 경우 과적합(overfitting)이나 다중공선성(multicollinearity) 문제가 생길 수 있다.

<표 3> 조건부 빈도표

Region	빈도	퍼센트	Owner	빈도	퍼센트
ss	321	48.0	pv	395	59.0
yj	348	52.0	sp	274	41.0
합계	669	100	합계	669	100
Cable	빈도	퍼센트	Facility	빈도	퍼센트
7	41	6.1	166	63	9.4
11	147	22.0	311	22	3.3
12	84	12.6	312	273	40.8
13	90	13.5	313	37	5.5
14	80	12.0	314	153	22.9
16	69	10.3	315	17	2.5
26	92	13.8	316	5	0.7
31	25	3.7	317	39	5.8
51	19	2.8	318	60	9.0
68	22	3.3			
합계	669	100	합계	669	100

<표 3>과 <표 4>는 조건부와 결론부 변수들의 기술통계량이다. 분석모형의 정확도 및 신뢰도 증대를 위해 실시하는 차원축소(dimensionality reduction) 이전에 명목척도 변수들의 상관관계를 파악하기 위하여 조건부와 결론부 변수들의 빈도를 확인하였다.

<표 4> 결론부 빈도표

Cause	빈도	퍼센트	Repair	빈도	퍼센트
121	137	20.5	111	112	16.7
122	9	1.3	112	3	0.4
123	215	32.1	113	219	32.7
124	305	45.6	115	67	10.0
304	3	0.4	116	206	30.8
			117	10	1.5
			121	3	0.4
			122	36	5.4
			302	3	0.4
			303	7	1.0
			304	2	0.3
			307	1	0.1
합계	669	100	합계	669	100

조건부와 결론부의 명목척도 변수들 간 독립성 여부를 검정하여 연관성이 높은 변수를 제거하기 위해 교차분석(χ^2 , Chi-Square)을 이용할 수 있으나, 표본이 독립표본이고 기대빈도 값이 5미만인 경우가 20% 미만이어야 한다[19]. 본 연구의 표본은 기대빈도 조건을 만족하지 못하여 변수 간 교차표와 통신선로 도메인 지식(domain knowledge)을 기준으로 차원축소를 실시하였다.

<표 5> 선로시설 * 소유구분 교차표

	Facility	Owner		전체
		pv	sp	
	166	63	0	63
	311	22	0	22
	312	273	0	273
	313	37	0	37
	314	0	153	153
	315	0	17	17
	316	0	5	5
	317	0	39	39
	318	0	60	60
전체		395	274	669

<표 5>는 조건부 선로시설과 소유구분의 교차표이다. 선로시설의 166, 311, 312, 313는 사설(pv)에만 나타나고 나머지 314, 315, 316, 317, 318은 통신사(sp)에만 해당되므로 선로시설과 소유구분은 상관관계가 매우 높다. 따라서 본 연구에서는 선로시설을 선택하고 소유구분 변수는 제거하였다.

<표 6>은 케이블과 지역 변수의 교차표이다. <표 6>을 통해 두 변수의 독립성 여부를 판단할 수는 없지만 통신케이블은 설치지역에 따라 설치년도를 <표 7>과 같이 특정할 수 있다.

케이블 설치년도와 지역을 조합하여 <표 7>과 같이 4개의 값을 가진 지역-설치년도(RnC) 변수를 새롭게 생성하였다.

RnC 변수에서 y85는 yj지역의 케이블 11, 12, 13, 14, 16, 26, 31, 51가 포함되었으며 케이블 설치년도가

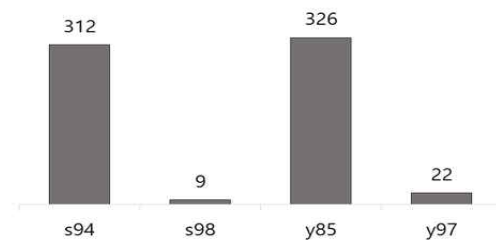
<표 6> 케이블 * 지역 교차표

	Cable	Region		전체
		ss	yj	
	7	41	0	41
	11	64	83	147
	12	47	37	84
	13	43	47	90
	14	31	49	80
	16	32	37	69
	26	54	38	92
	31	9	16	25
	51	0	19	19
	68	0	22	22
전체		395	274	669

<표 7> 케이블 설치년도

	Cable	Region		RnC
		ss	yj	
	7	1994	N/A	y85
	11	1994	1985	s94
	12	1994	1985	y97
	13	1994	1985	s98
	14	1994	1985	
	16	1994	1985	
	26	1995	1985	
	31	1998	1985	
	51	N/A	1987	
	68	N/A	1997	

1985 - 1987에 해당된다. y97은 yj지역의 68케이블로 설치년도가 1997년이다. s94는 ss지역 케이블 7, 11, 12, 13, 14, 16, 26이며 설치년도가 1994-1995에 해당된다. s98은 ss지역의 31케이블로 설치년도가 1998년 도이다. <그림 3>은 RnC변수의 빈도표이다.



<그림 3> RnC 빈도표

결론부의 장애원인과 수리내용 변수의 표본들은 <표 4>를 참고할 때 상관관계가 높다고 볼 수 없다. 두 변수는 예방 점검과 장애 예측에 필요한 변수이므로 본 연구의 분석 모델을 <표 8>과 같이 결정하였다.

<표 8> 분석 모델

조건부	결론부
RnC(지역-설치년도) Facility(선로시설)	Cause(장애원인) Repair(수리내용)

3.3 분석방법

본 연구에서는 변수의 기술통계량과 상관분석을 위해 SPSS를 이용하여 교차분석 및 Chi Square Tests를 수행하였다. 분석 모델의 연관규칙 마이닝은 오픈소스 통계패키지인 R 3.2를 이용하였다. 분석 모델의 빈발 항목집합을 도출하기 위하여 초기 지지도와 신뢰도 값은 R 패키지 “arules”의 “Apriori” 함수의 기본설정(지지도: 10%, 신뢰도: 80%, 조합수: 최대10개)을 활용하였다. 분석 모델의 표본수가 많지 않기 때문에 연관규칙 마이닝을 실행한 후 통신선로시설의 운용경험을 바탕으로 장애를 특정할 수 있는 유용한 규칙이 나올 때 까지 지지도와 신뢰도를 하향 조정하면서 연관규칙 마이닝을 실시하였다.

IV. 연구결과 분석

4.1 장애원인 분석

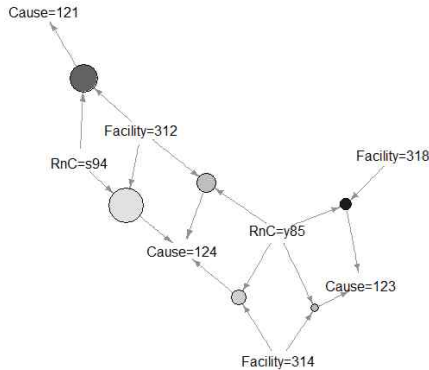
연관분석 결과를 살펴보면 다음과 같다. 통신선로 시설의 특성상 조건부와 결론부 표본의 속성값이 다양하게 분포하고 있고 출현빈도가 적기 때문에 지지도를 0.05로, 신뢰도를 50%/30%로 설정하여 연관규칙

분석을 실시하였다. 지지도 0.05, 신뢰도 50%이상 조건 일 때 3개의 항목집합이 도출되었고 지지도 0.05, 신뢰도 30%이상 조건 일 때 <표 10>과 같이 6개의 항목집합이 도출되어 변수들의 연결망을 <그림 4>에 제시하였다. 원의 크기는 조건부 변수의 상대적인 지지도의 크기를 나타낸다.

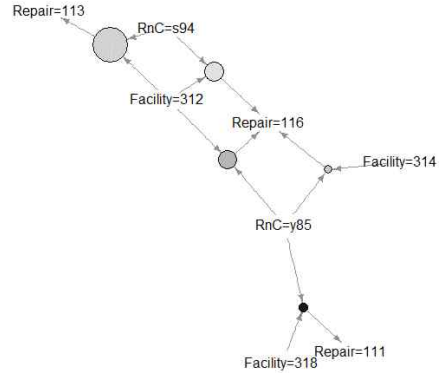
<표 10> 장애원인 분석 결과

no	antecedent	consequent	sup.	conf.	lift
1	RnC=y85, Facility=318	Cause=123	0.064	0.843	2.623
2	RnC=s94, Facility=312	Cause=121	0.103	0.408	1.994
3	RnC=y85, Facility=314	Cause=123	0.057	0.417	1.299
4	RnC=y85, Facility=314	Cause=124	0.071	0.527	1.157
5	RnC=s94, Facility=312	Cause=124	0.118	0.467	1.025
6	RnC=y85, Facility=312	Cause=124	0.082	0.579	1.270

신뢰도 50%이상 조건 일 때의 규칙들을 살펴보면 1번 규칙은 Y지역 1985년 설치 케이블, 지하케이블 시설에서 장애원인이 절연불량(123), 4번 규칙은 같은 케이블 구간 인입선 시설에서 접속불량(124), 6번 규칙은 Y지역 1985년 설치 케이블, 실내선 시설인 경우 장애원인이 접속불량(124)으로 나타났다. 1번, 4번, 6번 항목집합의 향상도는 각각 2.623, 1.157, 1.270이다. 이 중에서 향상도가 가장 높은 1번 항목집합이 빈발 항목집합으로 통신선로시설의 장애원인을 예측하기 위해 가장 유용한 규칙이다. 하지만 1번, 4번 규칙은 장애발생 구간이 통신사 국선구간이고 6번 규칙의 장애발생 구간은 구내통신망 구간이기 때문에 국선구간과 구내통신망 구간에서 향상도가 가장 높은 1번과 6번이 선로경계에 따른 장애원인을 예측하기에 유용한 규칙이다.



<그림 4> 장애원인 규칙 연결망



<그림 5> 수리내용 규칙 연결망

4.2 수리내용 분석

장애원인 분석 결과와 동일하게 수리내용 분석은 지지도를 0.05, 신뢰도를 50%/30%로 설정하여 연관규칙 분석을 실시하였다. 지지도 0.05, 신뢰도 50%이상 조건 일 때 2개의 항목집합이 도출되었고 지지도 0.05, 신뢰도 30%이상 조건 일 때 <표 11>과 같이 5개의 항목집합이 나타났다. 5개 항목집합 변수들의 연결망을 <그림 5>에 제시하였다. 신뢰도 50%이상 조건 일 때의 규칙들을 살펴보면 1번 규칙은 Y지역 1985년 설치 케이블, 지하케이블 시설에서 수리내용이 교체(111), 4번 규칙은 Y지역 1985년 설치 케이블, 실내선 시설인 경우 수리내용이 회선정비(116)로 나타났다.

<표 11> 수리내용 분석 결과

no	lhs	rhs	sup.	conf.	life
1	RnC=y85, Facility=318	Repair=111	0.061	0.804	4.802
2	RnC=y85, Facility=314	Repair=116	0.059	0.439	1.427
3	RnC=s94, Facility=312	Repair=116	0.079	0.314	1.018
4	RnC=y85, Facility=312	Repair=116	0.078	0.547	1.777
5	RnC=s94, Facility=312	Repair=113	0.104	0.414	1.265

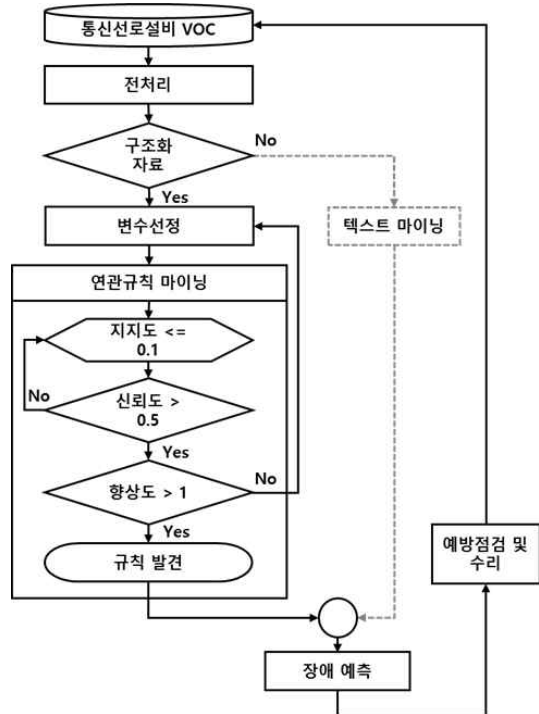
1번, 4번 항목집합의 향상도는 각각 4.802, 1.777이다. 이 중에서 향상도가 4.802인 1번 규칙이 빈발 항목집합으로 통신선로시설의 예방점검 및 장애방지를 위해 가장 유용한 규칙이다. 1번 규칙은 통신사의 국선구간에서 발생하는 장애원인을 예측할 수 있다면 4번 규칙은 선로경계에 따른 가입자의 구내통신망 장애발생 원인을 예측하기에 유용한 규칙이므로 통신사 국선과 가입자 구내에서 발생하는 장애 원인을 예측하기 위해서는 1번과 4번 규칙을 이용해야 한다.

4.3 결과 해석

장애원인 분석에서 신뢰도 50%이상 규칙들을 살펴보면 통신사 국선과 가입자 구내에서 동일하게 Y지역 1985년 설치 케이블에서 절연불량과 접속불량이 발생하고 있다. 이러한 현상은 1985년에 설치된 케이블(11, 12, 13, 14, 16, 26, 31, 51)의 노후화가 원인으로 추정된다. 또한 가입자의 구내통신망은 통신사가 변경될 때 마다 회선을 재구성하기 때문에 접속불량 장애가 많이 발생하는 경향이 있다.

수리내용 분석에서 신뢰도 50%이상 조건 일 때 1번 규칙은 Y지역의 1985년 설치 케이블(11, 12, 13, 14, 16, 26, 31, 51)인 구내통신망 지하케이블(318) 시

설에서 통신선로설비의 교체가 자주 이뤄졌다. 같은 신뢰도에서 4번 규칙도 역시 Y지역의 1985년 설치 케이블의 가입자망 구간에서 실내선(312) 시설에서 회선정비 문제가 빈번하게 발생하고 있다. 4번 규칙과 같이 가입자의 구내통신망에서 자주 발생하는 장애 현상은 초고속 인터넷이 나오면서 증가하기 시작했다. 구내통신망의 IDC(Insulation Displacement Contacts)형 단자는 통신사가 회선을 개통하기 위해서 자주 사용하는 시설로 정확한 공기구를 사용하지 않으면 단자에 설치되어 있는 칼날이 벌어짐으로써 연결된 회선이 자주 빠지게 된다. 따라서 3~4년 주기로 단자를 교체하여야 하지만 현재는 이를 책임지는 주체가 없는 실정이다. 과거에는 공기업 통신사에서 교체작업을 하였지만 현재는 국내 통신 3사가 통신망을 공동으로 사용하기 때문에 책임소재가 불분명하다. 따라서 가입자가 관리해야하는 사설구간을 주기적으로 정비할 수 있는 주체를 정책적으로 선정하여야 한다.



<그림 6> 통신선로설비 장애 예측 모형

4.4 장애 예측 모형

본 연구에서는 통신사 VOC시스템의 구조화된 데이터(structured data)만을 추출하여 연관규칙 마이닝을 실시하였다. VOC시스템은 다양한 경로로 저장된 비구조화된 데이터(unstructured data)가 포함되어 있지만 본 연구에서는 데이터 수집의 한계로 제외하였다. 통신사 VOC의 연관규칙 마이닝 결과를 바탕으로 통신선로설비의 장애 예측 모형을 <그림 6>과 같이 제시하였다.

장애 예측 모형에서 조건부와 결론부의 변수를 올바르게 선정하는 것이 매우 중요하다. 선택된 조건부 변수는 장애 원인을 찾기 위한 중요한 단서가 되기 때문에 변수 선정은 관련 분야의 업무 전문가가 참여해야 예측 모형의 정확도를 높일 수 있다. 지지도 0.

1, 신뢰도 0.5를 기준으로 신뢰도가 0.5이상인 규칙이 나오지 않을 경우 신뢰도가 0.5이상인 규칙이 나올 때까지 지지도를 0.1이하로 점진적으로 낮춰가면서 마이닝을 반복적으로 실시한다. 지지도와 신뢰도의 조건이 만족하더라도 향상도가 1이상 나오지 않는 경우에는 유용한 규칙이 없거나 변수 선정이 잘못된 결과로 볼 수 있기 때문에 변수 선정을 다시 고려해야 한다.

연관규칙 마이닝으로 유용한 결과가 도출된 경우에는 예방점검과 사전조치 작업을 통해 미래에 발생할 수 있는 서비스 이용단절과 이로 인한 고객의 불편을 감소시킬 수 있다.

V. 결론

본 연구에서는 A 통신사의 실제 VOC 데이터에 기반한 연관규칙 마이닝을 이용하여 통신선로설비의 회선 장애, 가입자시설 장애를 중심으로 장애원인을 진단함으로써 통신 서비스 품질 저하를 효과적으로 방지하고 고객 만족도를 개선할 수 있는 방안을 제시하였다. 본 연구를 통해 A 통신사의 통신선로시설 장애의 원인이 특정지역에 설치된 케이블, 가입자가 관리하는 사설 선로구간과 많은 연관이 있음을 알 수 있었다. 가입자가 관리하는 시설의 정비는 국내통신 3사가 협의하여 정비를 실시할 수 있도록 관련 부처에서는 관리감독을 하여야 한다. 시설관리는 원칙적으로 건물주의 책임이나 건물주가 통신선로에 관한 지식이 부족하기 때문에 정비 지침이 필요하다. 또한 통신선로시설에 대한 VOC가 다수 발생하는 경우 통신사용에 불편함이 발생하는 신호로 감지하여 관련 부처에서는 강력하게 통신선로시설의 장애 절감 대책을 수립하여 시행할 수 있도록 지도감독을 하여야 한다.

다만, 본 연구에 사용한 데이터가 특정 통신사 및 지역에 국한되어 표본수가 충분하지 못하였다는 한계점이 있다. 따라서 통신사에서 관리하는 데이터를 더 활용하여서 지속적으로 분석하여 정비 대책을 세워야하며 정책적으로는 건물에 사용하는 통신시설 관리 기준을 만들어서 분쟁의 요소가 발생되지 않도록 조치하여야 한다. 연구 결과에서 나타났듯이 고객이 관리하는 시설 불량에 대하여 무조건 무료수리를 함으로써 통신사의 비용이 증가하게 할 수는 없는 사항이다. 가입자의 민원 처리에 소요되는 비용은 건당 접수에 2500원이 소요되며, 현장 출동에 따른 비용은 19,000원 정도가 소요되는 것으로 추정된다. 민원처리 비용과 소비자 불편을 줄이기 위해서는 사전적 예방대책을 포함한 통신선로시설의 품질관리 방안을

수립하여서 시행하여야한다.

본 연구는 실제 현업에서 수집된 VOC에 연관규칙 마이닝 결과를 이용하여 실제 통신선로시설의 품질을 개선할 수 있도록 판단 근거를 제공할 뿐만 아니라 사전 예방정비를 통해 가입자에게 높은 서비스 품질과 만족도를 제공해 줄 수 있는 실증연구로서 의의를 갖는다.

추후 연구 방향은 다음과 같다. 첫째, 본 연구에서 인위적으로 설정한 빈발 항목집합 결정을 위한 지지도와 신뢰도 결정에 있어서 그 유효성을 평가하기 추가적인 연구가 필요하다. 이를 위해서는 2013년 8월 이후에 접수된 고객민원 데이터를 검증 데이터로 활용하여 본 연구를 통해 도출된 규칙을 지속적으로 검증하고 보정할 수 있어야 한다. 둘째, 연관규칙 분석을 통해 도출된 규칙을 기반으로 지속적으로 품질을 관리하고 개선할 수 있도록 경영관리기법과 연계한 연구가 필요하다. 셋째, 도출 결과의 신뢰성을 높일 수 있도록 충분한 표본수를 확보해야 하고 비구조화된 VOC 데이터의 추가적인 연구가 필요하다.

참고문헌

- [1] L. V. Subramaniam, T. A. Faruquie, S. Ikbal, S. Godbole and M. K. Mohania, "Business Intelligence from Voice of Customer," 2009. ICDE '09. IEEE 25th International Conference, 2009, pp. 1391-1402.
- [2] S. S. Jain, B. B. Meshram and M. Singh, "Voice of customer analysis using parallel association rule mining," 2012 IEEE Students' Conference on Electronics and Computer Science, 2012, pp. 1-5.
- [3] 박숙지, 김진식, 양영문, 이성우, 정병덕, "VOC 와 근원장애 분석 정보를 활용한 네트워크 장애티켓

- 관리시스템 개발,” KNOM Review, 제12권, 제1호, 2009, pp. 1-10.
- [4] 김유신, 정승렬, “오피니언 마이닝을 이용한 지능형 VOC 분석시스템,” 지능정보연구, 제19권, 제3호, 2013, pp. 113-125.
- [5] C. H. Choi, J. E. Lee, G. S. Park, J. H. Na and W. S. Cho, “Voice of Customer Analysis for Internet Shopping Malls,” International Journal of Smart Home, Vol. 7, No. 5, 2013, pp. 291-304.
- [6] 김용, “Apriori 알고리즘 기반의 개인화 정보 추천 시스템 설계 및 구현에 관한 연구,” 한국비블리아학회지, 제23권, 제4호, 2012, pp. 283-308.
- [7] 이연정, 김경재, “다중모형조합기법을 이용한 상품 추천시스템,” 지능정보연구, 제19권, 제2호, 2013, pp. 39-54.
- [8] 임영희, 이종욱, 박대회, 장진경, “연관규칙 마이닝을 이용한 한국 신노년층의 생활 만족도에 관한 연구,” 한국가정관리학회 기타간행물, 2008, pp. 164-173.
- [9] W. X. Xie, H. N. Qi and M. L. Huang, “Market Basket Analysis Based on Text Segmentation and Association Rule Mining,” 2010 First International Conference on Networking and Distributed Computing, 2010, pp. 309-313.
- [10] M. Klemettinen, H. Mannila and H. Toivonen, “Rule Discovery in Telecommunication Alarm Data,” Journal of Network and Systems Management, Vol. 7, No. 4, 1999, pp. 395-423.
- [11] 김동성, 김종우, “인터넷 상점의 웹 페이지 방문 로그를 활용한 고객 행위 예측,” 한국지능정보시스템학회 학술대회논문집, 2013, pp. 125-128.
- [12] 이지영, 김종우, “실시간 CRM을 위한 분류 기법과 연관성 규칙의 통합적 활용: 신용카드 고객 이탈 예측에 활용,” 한국경영정보학회 학술대회논문집, 2007, pp. 135-140.
- [13] 임숙중, 김정자, 원용관, “연관규칙탐사 기법을 이용한 네트워크 서비스 시스템의 위험분석 및 평가,” 전자통신기술논문지, 제6권, 제1호, 2003, pp. 73-78.
- [14] 정영수, 강창욱, “연관규칙을 이용한 제조 공정의 불량항목 예측,” 한국산업경영시스템학회 학술대회, 2004, pp. 310-313.
- [15] 황인수, “연관규칙을 이용한 상품선택과 기대수의 예측,” 경영정보학연구, 제14권, 제4호, 2004, pp. 87-97.
- [16] 신대철, 김홍윤, “침입탐지 알고리즘 성능 최적화 및 평가 방법론 개발,” 디지털산업정보학회 논문지, 제8권, 제1호, 2012, pp. 125-137.
- [17] R. Agrawal, T. Imielinski and A. Swami, “Mining associations between sets of items in massive databases,” Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data, ACM Press, New York, 1993, pp. 207-216.
- [18] G. Shmueli, N. R. Patel and P. C. Bruce, Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner SE, John Wiley & Sons, New Jersey, 2012.
- [19] 최현철, 사회과학 통계분석, 나남, 서울, 2012, p. 125.

■ 저자소개 ■



나 기 주
Na Gijoo

2014년 4월~현재
가트너 사업개발(BD) 디렉터
2006년 10월~2014년 4월
한국오라클 부장
2003년 4월~2006년 10월
티맥스소프트 과장
2013년 3월~현재
서울과학기술대학교
산업정보시스템 박사과정
2008년 8월 서울시립대학교 경영학과(석사)
2000년 2월 전북대학교 항공우주공학과(학사)

관심분야 : BPM, 사회연결망분석,
데이터마이닝
E-mail : dolbaram74@naver.com



한 인 섭
Han Insup

2014년 9월~현재
장광(주) 미래연구소장
1977년 5월~2014년 4월
KT 수도권 강남본부
2011년 9월~현재
서울과학기술대학교
산업정보시스템 박사과정
2011년 8월 서울과학기술대학교
산업정보시스템(석사)
2000년 2월 한국방송통신대학교
컴퓨터공학과(학사)

관심분야 : 미래 네트워크 구조 및 설계
E-mail : his3325@daum.net



조 남 옥
Cho Namwook

2003년 3월~현재
서울과학기술대학교
글로벌융합산업공학과 교수
2003년 3월~2004년 3월
삼성SDS 책임컨설턴트
2001년 3월~2002년 3월
Lucent Technologies, MTS
2001년 2월 Purdue University 산업공학(박사)
1996년 2월 서울대학교 산업공학과(석사)
1994년 2월 서울대학교 산업공학과(학사)

관심분야 : BPM, 사회연결망분석, 이상치
탐지
E-mail : nwcho@seoultech.ac.kr

논문접수일: 2015년 11월 13일
수 정 일: 2015년 11월 24일
계재확정일: 2015년 11월 27일