퍼지 및 지능시스템학회 논문지 2007, Vol. 17, No. 2, pp. 160-166

데이터 마이닝 기술을 적용한 사용자 선호 스팸 대응 온톨로지 구축

Constructing User Preferred Anti-Spam Ontology using Data Mining Technique

김종완 · 김희재 · 강신재

Jong-Wan Kim, Hee-Jae Kim and Sin-Jae Kang

대구대학교 컴퓨터 · IT공학부

요 약

사용자마다 임의의 메일에 대한 반응은 자신의 취향에 따라 다를 수 있다. 본 논문에서는 사용자 선호 온톨로지를 구축함으로서 스팸 메일을 줄이고자 한다. 사용자의 행동양식을 기술하는 온톨로지를 정의하기 위하여, 사용자들의 선호도 정보와 그들의 이메일에 대한 반응을 연구하기 위한 연관성 분류 마이닝 방법을 적용했다. 생성된 분류 규칙은 정형화된 온톨로지 언어로 표현된다. 사용자 선호 온톨로지는 어떤 메일이 스팸 또는 비스팸 인지를 의미있는 방식으로 설명할 수 있다. 또한 사용자들의 온톨로지에 대한 이해력 향상을 위해 논리합성에 기반한 새로운 규칙 최적화 절차를 제안하여 불필요한 규칙들을 제거한다.

키워드 : 데이터 마이닝, 스팸 대응 시스템, 사용자 선호 온톨로지, 논리합성, 규칙 가지치기

Abstract

When a mail was given to users, each user's response could be different according to his or her preference. This paper presents a solution for this situation by constructing a user preferred ontology for anti-spam systems. To define an ontology for describing user behaviors, we applied associative classification mining to study preference information of users and their responses to emails. Generated classification rules can be represented in a formal ontology language. A user preferred ontology can explain why mail is decided to be spam or non-spam in a meaningful way. We also suggest a new rule optimization procedure inspired from logic synthesis to improve comprehensibility and exclude redundant rules.

Key Words : data mining, anti-spam system, user preferred ontology, logic synthesis, rule pruning.

1. 서 론

스팸 메일은 발신인과 전혀 관련이 없는 수신자에게 무차별로 보내지는 원치 않는 전자우편(email)을 의미한다[1]. 대부분의 이메일 소프트웨어들은 전형적인 블랙리스트들이나 키워드 기반의 필터 등 자동적인 스팸 메일 필터링 기능을 지원한다. 이러한 필터링 방식은 초기에는 효과적이었지만 시간이 지남에 따라 그 정확성이 점차 낮아지고 있다. 왜냐하면 스팸 발송자들이 키워드기반 필터의 약점인 친숙한 문구 주제들을 사용하기 때문이다[2]. 나이브 베이지안(naive Bayesian) 분류기와 서포트 벡터 머신(Support Vector Machine)과 같은 다양한 기계 학습 알고리즘들이 다른 메타데이터상에서의 이메일 분류작업을 위해 사용되고 있다. 이들 스팸 대응 방법들은 통계적 정확성이 많이 향상되었지만, false positive(비스팸이 스팸으로 분류)와 false negative(스팸이 비스팸으로 분류)로 필터링되는 문제가 있다. 즉, 어떤 이메일은 누군가에게는 스팸이지만, 실제 상황에서 다른 사람들에게는 비스팸일 수가 있다. 이메일에 대한 사용자들의 행동양식은 사용자의 선호에 의해 달라지므로, 사용자의 선호에 기반을 둔 사용자 중심의 스팸 대응 서비스들을 제공하는 게 범용 스팸 메일 대응 시스템 보다 의미가 있다고 판단된다. 따라서 본 연구에서는, 연관성 분류 마이닝(associative classification mining) 시스템을 학습하기 위해 사용자 선호 정보와 이메일의 반응을 수집하고, 데이터 마이닝 기술에 의해 생성된 규칙들을 사용하여 사용자 선호 온톨로지(user preferred ontology)를 형식 언어(formal language)로 정의한다. 스팸 대응 도메인 온톨로지의 효과를 알아보기 위해, 사용자 선호 기반 스팸 메일 대응 시스템의 개념을 설계하고 구현한다. 이 평가는 수십 명의 사용자들로부터 수집된 사용자 정보에 기초하였으며, 실험 결과는 제안된 접근법이 타당함을 제시한다.

본 논문의 2장에서는 관련 연구를 소개하고, 3장에서는 데이터 수집 및 전처리 과정을 소개하고, 4장에서는 연관성 분

접수일자 : 2006년 10월 21일
완료일자 : 2006년 12월 30일
감사의 글 : 이 논문은 2006학년도 대구대학교 학술연구비 지원에 의한 논문임

류 마이닝을 사용한 온톨로지 구축 방법론을 기술하고, 5장에서는 제안된 시스템 구조와 실험 결과를 보여주고, 6장은 결론으로 구성되어 있다.

2. 관련 연구

최근 들어 개인 이메일 관리 시스템에 대한 연구가 많이 제안되고 있다. Gray와 Haahr는 개인화된 협동(personalized collaborative) 스팸 필터를 제안하였다[3]. 개인화된 협동 필터는 네트워크상의 멤버들에 의해 알려진 모든 스팸 메시지들에 대한 정보를 수집하여 각 사용자에게 관련성이 높은 스팸 정보들을 전달해준다. 이런 개인화된 협동 필터를 개발하기 위하여 새로운 스팸이 분류되면, 유사한 메일을 받을 사용자들에게 그 메일을 스팸으로 고려하도록 그 메시지의 서명(signature)을 계산하고 전달한다. P2P (peer-to-peer) 구조가 이러한 시스템을 가능하도록 한다. 따라서 P2P 네트워크에서 서로 다른 사용자의 서명에 의존하게 되는 이러한 접근법은 독립적으로 수행되지 못한다. 반면에 Ravi 등은 네트워크 말단(network edge)에서 개인적인 이메일 관리를 제안하였다[4]. 그들의 인공신경망(artificial neural network) 기반 스팸 필터는 서버에서 스팸과 바이러스를 필터링함으로써 네트워크의 대역폭을 유지하였다. 이 시스템은 2가지 필터를 가진다. 첫 번째 필터는 이메일에서의 텍스트 패턴을 인식하고 그 패턴을 학습하고, 두 번째 필터는 이메일에 있는 이미지를 학습한다. 이 시스템은 사람이 스팸을 구별하는 것과 매우 유사하다. 그러나 이 방식은 어떤 이메일 주소의 메일이 스팸으로 확인되면 이것을 학습하고, 같은 사용자의 모든 다른 이메일 주소들로부터 비슷한 스팸 메일들을 삭제하는 단순한 중앙 관리형 스팸 필터이다.

한편 사용자 선호를 고려한 몇 가지 스팸 메일 대응 시스템들도 연구되고 있다. 이들 시스템의 대부분은 스팸 대응 시스템의 추천을 수락할 것인지 아닌지에 대한 사용자들의 선택을 요구한다. 이 시스템들에서는 명백한 스팸은 아니지만 스팸의 가능성이 높은 이메일을 특정한 영역에 저장한다. 메일의 수신자들은 이 특정 영역의 웹 링크를 받게 되고, 개별 사용자들은 특정 영역에 있는 해당 메일을 자신의 의도대로 분류하도록 요청받는다. 그러면 시스템은 각 사용자의 선호를 기억해서 특정한 소스로부터 오는 이메일들을 사용자의 받은 편지함으로 전달하거나 스팸 편지함으로 차단하는 기능을 수행한다[5]. 또한 일부 시스템은 생소한 외국어로 인해 스트레스를 받지 않도록 각 사용자들이 일반 메일이라고 가정하는 언어를 지정하는 옵션을 제공하기도 한다.

기존의 사용자 취향을 고려한 스팸 대응 시스템들과는 달리, 본 연구의 목표는 순수하게 사용자의 선호도와 사용자의 반응에 기반을 둔 사용자 선호 온톨로지 기반의 스팸 대응 이메일 관리 시스템을 개발하는 것이다.

온톨로지는 "실세계 혹은 특정 도메인에 존재하는 모든 개념들(concepts)과 그 개념들의 속성들(properties), 그리고 개념들이 상호간에 의미적으로 어떻게 관련되는지(semantic relations)에 대한 정보를 가지고 있는 지식베이스(knowledge base)"라 정의할 수 있고, 지능 시스템에서 정보의 의미를 정의하는데 주요한 역할을 수행한다[6]. 사용자 선호 기반 스팸 대응 시스템에서는 사용자의 취향을 기반으로 한 사용자의 행동양식을 형식적으로 정의한 도메인 온톨로지 (domain ontology)가 도움이 될 수 있다. 도메인 온톨로지 작성 방법으로는 상향식과 하향식의 두 가지가 있다. 하향식(top down) 접근법에 있어서 온톨로지 전문가들은 도메인 지식과 직관을 기반으로 개념과 이의 관계를 결정한다. 상향식(bottom-up) 접근의 온톨로지 전문가들은 관련된 데이터의 적용범위와 패턴들을 분석함으로써 중요한 개념을 선택한다. 비록 자연 언어 텍스트들로부터의 온톨로지 지식을 습득해주는 몇몇 자동화 도구들로 사람의 노력이 감소된다 하더라도 상향식과 하향식 방법은 사람의 노력을 수반한다[7]. 본 논문에서는 사용자의 선호와 사용자의 행동양식들 사이의 관계를 발견하는데 중점을 두고 있으며, 이러한 관계들은 도메인 온톨로지에서 규칙들로 나타낼 수 있다. 데이터 마이닝은 이러한 규칙 발견에 유용하지만, 데이터 마이닝 기법들로부터 생성되는 많은 규칙들 가운데 쓸모있는 규칙들을 발견하기는 쉽지 않다. 몇몇 규칙들은 너무 길거나 세분화되어 있어서 온톨로지 구성을 어렵게 하기도 한다.

규칙의 최소화와 최적화를 위해 사용될 수 있는 기계 학습에서의 다진 논리(multi-valued logic)에 대한 연구가 있다. Files와 Perkowski는 다진 논리합성(multi-valued logic synthesis) 방법을 연구하였다[8]. 기계학습의 일부 개념들이 다진 논리합성과 완전하게 일치된다는 사실을 서술하고, 잘 알려진 분류 알고리즘인 C4.5와 산업계 표준 논리 최소화 도구인 Espresso 보다 다진 논리합성의 성능이 우수하다는 것을 보여주었다. 또한 제한된 전제조건을 가진 규칙들을 반복적으로 마이닝하는 방법도 보고된 바 있다[9]. 이 방법은 연속적인 질의에 효율적으로 답하기 위해서 이전의 단계에서 얻은 마이닝 정보를 활용할 수 있는 반복적인 알고리즘이다. 본 논문에서는 부울 대수 간소화로 잘 알려진 카르노 맵(Karnaugh map)으로부터 영감을 얻어 논리합성과 데이터 마이닝에 기반한 규칙 가지치기(rule pruning) 방법을 제안하고, 이를 통해서 사람이 이해하기 쉽지 않은 다소 길고 중복된 규칙들을 배제한다.

3. 데이터 준비

샘플 도메인 데이터로부터 온톨로지를 구축하려면 먼저 데이터 준비단계가 필요하다. 일부 이메일 파일들에 대한 사용자들의 반응을 얻기 위해서, 우리는 본 연구진이 속한 학부 1학년에서 4학년 학생들에게 샘플 메일에 대한 사용자 반응을 조사하였다. 우리는 이미 내용기반 이메일 필터링에 관한 연구[10]를 수행한 경험이 있으므로, 본 연구에서는 이메일의 헤더와 본문을 포함한 이메일의 내용 대신에 특정한 사용자 그룹으로부터 각 개인의 흥미와 행동양식들을 기반으로 하는 스팸 메일 대응 시스템 개발을 목표로 하였다. 사용자의 선호 데이터를 준비하기 위해 사용자 프로파일 형식과 이메일에 대한 사용자 반응 부류를 다음과 같이 설계하였다.

Yahoo와 Comcast 같은 많은 웹메일 시스템들은 사용자의 개인 정보를 요구한다. 이때 요구되는 사용자의 등록 정보처럼, 본 논문은 사용자 프로파일(user profile)에 포함할 특성들로 14가지 {Age, Gender, RequiredHits, News, Finance, Sports, Adults, TvMovieMusic, Kids, Games, Travel, Shopping, Jobs, RealEstates}를 선택하였다. 나이(Age) 특성은 출생년도에 따라 5단계로 했다가, 실험에 참여한 사용자들이 대학생이므로 다시 두 그룹 FS(1-2학년)과 JS(3-4학년)으로 나누었다.

RequiredHits(지금부터 RHit로 표기)는 Spam Assassin[2]에서 제안된 용어로서, 어떤 메일이 스팸으로 간주되기 전에 스팸과 관련된 용어들이 본문에 얼마나 많이 나

타나는 지를 의미한다. 숫자를 사용하는 Spam Assassin과는 다르게, 본 연구에서는 이메일의 내용을 고려하지 않고 언어 표현이 사용자에게 보다 편하기 때문에 매우 약함(VW), 약함(W), 보통(N), 강함(S), 매우 강함(VS)처럼 언어항으로 구별했다. 만약 사용자들이 강한 스팸 필터를 원한다면 RHit값으로 약함(W)을 선택할 것이고, 약한 필터를 원할 때는 강한(S) RHit를 선택하면 된다.

특성 추출(feature selection)은 예측에 적합한 특성들의 부분집합을 발견하기 위해 후보 특성 집합 안에 있는 특성들의 모든 가능한 조합을 조사한다. 최적의 특성수를 찾기 위해 고안된 방법들로는 정보획득량(information gain: IG), 상호 정보(mutual information: MI), 카이제곱 통계량(chi squared test) 등이 있다. 이전의 데이터 마이닝 연구[11]에 의하면, IG는 이러한 문제에서 좋은 해결책이 된다. 우리는 모두 14개의 속성들에 대한 IG를 계산하고, 그들로부터 상위 몇 개의 속성들을 선택했다. 이 과정은 5.2절에서 설명될 것이다.

이메일을 받는 사람은 받은 이메일에 대해 보통 4가지, 즉 Reply, Delete, Store, Spam으로 응답한다. 물론 전달이라는 Forward도 있지만, 본 연구에서는 고려하지 않는다. 사람들은 어떤 메일에 흥미가 없으면 Delete하고, 만약 중요하거나 반응할 가치가 있다고 생각되는 메일이면 보내는 사람에게 Reply한다. Reply 여부에 상관없이 때때로 그 이메일이 언젠가는 필요할 수 있기 때문에, 편지함에 메일들을 보관(Store)하기도 한다. 스팸 메일이 그들에게 주어졌을 때 대부분의 사용자는 다시 스팸 메일을 받는 것을 싫어하기 때문에 이와 같은 종류의 메일들을 스팸 편지함으로 이동(Spam)한다. 물론 일부 사용자들은 스팸 메일을 바로 삭제하고 스팸 편지함으로 이동시키지만, 또 다른 부류의 사용자들은 스팸 메일들을 삭제만 하거나 이동만 시키는 단일한 행동을 수행한다. 아무튼 중요한 사실은 대부분의 사용자들은 자신의 편지함에 비스팸만 보관하기를 원한다는 점이다.

따라서 우리는 샘플 이메일들, 조사에 참여한 사용자들의 개인 정보와 취향, 메일에 대한 반응들을 수집하였다. 개인 정보에 해당하는 속성들인 Age는 {FS, JS}로, Gender는 {male(M), female(F)}인 이진 형태로 주어지며, 대부분의 다른 취향 속성들도 이진 형식으로 되어 있다. 예를 들어, 만약 사용자가 News에 관심이 있다면 그 속성에 대해 true(T)를 선택하며 관심을 가지고 있지 않으면 false(F)를 선택한다. Finance, Sports, Adults, TvMovieMusic, Kids, Games, Travel, Shopping, Jobs, RealEstates와 같은 속성들도 T와 F 값들을 가진다. 반면에, RHit에서는 다진 속성값 (VW, W, N, S, VS)을 가진다. 분류의 목표 변수인 Response는 위의 4가지 범주형(category) 값을 가진다.

4. 온톨로지 구축

스팸 대응 온톨로지 구축은 3단계로 수행한다. 첫 번째 단계는 사용자들의 선호와 그들의 이메일 반응 사이의 연관성(규칙)들을 알기 위해 연관성 분류 마이닝을 사용한다. 다음 단계에서는, 불필요한 규칙들을 제거하고 이해하기 쉬운 규칙들만 남기기 위해 새로운 규칙 가지치기 과정을 적용한다. 최적화 규칙들을 도메인 온톨로지 안에 있는 공리(axiom)들로 변환하는 작업은 마지막 단계에서 수행된다. 이제부터 각 단계를 자세히 설명하겠다.

먼저, 사용자들의 다양한 그룹과 샘플 이메일 데이터에 대한 반응들 사이의 연관 규칙을 발견하고자 노력하였다. 예를 들어, 여자들은 보통 쇼핑을 좋아하고, 학생들은 취업에 강한 흥미를 갖고 있을 것이라고 기대한다. 이러한 직감은 사용자 선호 데이터 집합에 연관성 마이닝(associative mining)을 적용한 후에 확인할 수 있었다. 같은 방법으로 사용자 프로파일과 샘플 이메일에 대한 사용자의 반응을 포함한 사용자 로그 파일(user log file)들 사이의 알려지지 않은 상관관계도 밝히기를 원했다. 따라서 우리는 샘플 이메일에 대한 선호 데이터를 학습하기 위해 대표적인 의사 결정 트리(decision tree) 알고리즘인 ID3[11]을 선택하였다. 3장에서 설명했듯이 샘플 데이터들은 대부분 이진 속성을 가지고, 일부 범주형(nominal) 특성들로 구성되어 있으므로, ID3가 데이터 집합으로부터 대표 규칙들을 발견하기에 적합하다. ID3 마이닝을 수행하여 의사 결정 트리를 생성시킨 후, 트리의 각 경로를 하나의 규칙으로 기술함으로써 의사 결정 트리를 규칙들로 변환한다. 이때 각 경로안의 루트 노드에서 내부 노드까지는 각 규칙의 전제조건(antecedent condition)들이 되고, 단말 노드는 각 규칙의 결론(conclusion)으로 간주한다. 어떤 규칙이 좋은 지를 평가하기 위해서 그 규칙과 일치하는 테스트 패턴들의 비율을 계산함으로써 정확도(accuracy)가 계산된다.

둘째 단계에서는 중복된 규칙들을 제거하고 이해하기 쉬운 규칙들을 선택하기 위하여 새로운 규칙 최소화 과정을 적용한다. 따라서 논리합성(logic synthesis)으로부터 영감을 얻은 일종의 규칙 가지치기 접근법을 제안한다. 카르노 맵은 불리언 논리 단순화를 이해하기 위한 간단하고 쉬운 방법으로 잘 알려져 있다[12]. 카르노 맵에서는 동일한 함수에 대하여 2개 이상의 단순화된 논리 표현들을 발견하는 것이 가능하다. 예를 들면, 그림 1에 있는 함수 $F(A,B,C)=\sum(1,3,4,5,6,7)$는 세 개의 입력 변수 A, B, C로 구성된다. 이 함수 F는 6개의 최소항(min term)들인 {001, 011, 100, 101, 110, 111}을 가진다. 그림 1(a)와 (b)에서 보이는 것처럼, 두 종류의 논리 최소화가 가능하다.

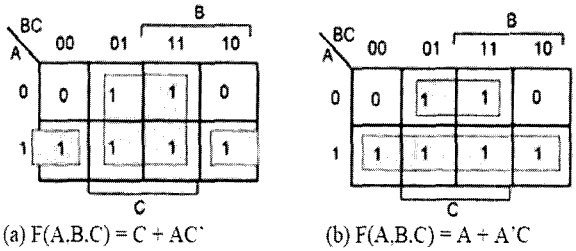

그림 1. 카르노 맵 방법의 예.
Fig. 1. An example of Karnaugh map method.

그림 1에서 보는 것처럼 함수 F가 고정되어도 두 개의 다른 논리 표현이 가능하며, 사실 두 표현은 논리 관점에서 동등하다. 우리는 이러한 카르노 맵의 예로부터 규칙 최소화에 대한 아이디어를 얻었다. 제안된 규칙 최소화 방법은 불리언 변수와 범주형 변수들을 함께 다룰 수 있기 때문에 혼합형 논리합성 규칙 가지치기(hybrid logic synthesis rule pruning: HLSRP)라 부르겠다. 데이터 마이닝으로부터 유도된 규칙 집합에는 몇 가지 변수들이 있다. 대부분의 변수들은 불리언 또는 이진 변수이지만, RHit와 같은 다진값 {VW, N, S, VS}을 가지는 변수도 있다. 약함(W)값은 이 연구에 참여한 사용자들에 의해 전혀 선택되지 않았기 때문에 RHit속성

으로부터 W는 제거되고 4가지 값이 남았다. 혼합 논리합성에서는, 만약 어떤 변수에 대한 두 개의 대응되는 논리가 다르다면, 그 변수를 가진 두 개의 규칙들은 하나로 병합되고 해당 변수에 있던 전제조건은 병합 규칙에서 생략되므로 간단한 규칙들이 유도될 수 있다. 4개의 인위적인 규칙들을 가진 아래의 예는 이러한 아이디어를 잘 설명해준다.

R1: if Age=JS and RHit=S and News=F and Adults=F and Games=F then Response=Delete

R2: if Age=JS and RHit=S and News=F and Adults=F and Games=T then Response=Delete

R3: if RHit=N and News=T and Adults=T then Response=Store

R4: if RHit=VS and News=T and Adults=T then Response=Store

Games 속성의 전제조건만 다른 두 개의 비슷한 규칙 R1과 R2가 있으므로, 논리합성 연산에 의해 Games=T는 Games=F와 결합하기 때문에 변수 Games의 전제조건인 F와 T가 Null로 병합되고, 결국 두 개의 규칙 R1과 R2는 Null 조건을 배제한 새로운 규칙 R5로 대치된다.

R5: if Age=JS and RHit=S and News=F and Adults=F then Response=Delete

범주형 변수들에게도 유사한 기능들이 수행되어진다. 앞에서 설명한 것처럼, 변수 RHit는 4개의 카테고리들을 가진다. 만약 비교되는 두 개의 규칙들에 대한 RHit 값들이 다르면 두 개의 규칙들은 하나의 규칙으로 병합된다. 위의 예에서 R3은 RHit에 N의 값을 가지고 R4에서는 RHit가 VS를 갖는다. 규칙 R3과 R4의 논리합성으로부터 RHit에 대한 전제조건은 Null을 가지게 되고, 새로운 규칙 R6을 구축하는데 그 조건은 제외된다.

R6: if News=T and Adults=T then Response=Store

두 개 이상의 규칙들이 때때로 단지 하나의 다른 전제조건을 가진 다른 규칙들을 병합하는 기회를 얻기 위해 경쟁하는 경우도 있다. 이런 상황을 해결하기 위해서, 규칙 집합에 있는 각 속성에 대한 IG를 고려한다. 병합될 후보가 두 개 이상 나타나면, 우리는 속성이 가장 낮은 IG를 가지는 변수를 선택해서 그 변수에서 구별되는 속성값을 가진 두 개의 규칙들을 병합한다. 직관적으로 높은 IG를 가진 속성들이 규칙 집합에 끝까지 남는 것이 바람직하다. 우리가 기대한 것처럼 5.2절의 실험 결과는 기존 ID3 마이닝과는 약간 다른 결과를 보여준다.

셋째로, 분류 규칙들을 온톨로지와 온톨로지 간의 매핑을 표현하는데 적합한 강한 형식의 일차 논리 언어(strongly-typed first-order logic language)인 Web-PDDL[13]을 사용하여 온톨로지로 표현한다. 이전의 결과들을 기초로 Web-PDDL로 사용자 선호 온톨로지를 정의한다. 우리는 클래스(class)와 프로퍼티(property)로서 아래의 개념들(concepts)을 선택하였다.

Classes (Types): Preference, Event, Email, Action, Client, Gender, RHit, Response

Properties (Predicates): name, gender, age, prefer, accuracy

Web-PDDL에서는 위의 개념들을 다음과 같이 표현할 수 있다.

```
(define (domain spam_email)
    (:extends(uri http://orlando.drc.com/daml/ontology/
  Person/G3/Person-ont-g3r1" :prefix pdt)
              (uri "http://www.w3.org/2000/10/
  XMLSchema" :prefix xsd))
    (types: Preference Event Email - Object Action -
Event Client - @pdt:Person Gender - @xsd:string
  RHit - @xsd:string Response - Action)
    (:Objects Store Delete Spam Reply - Response
  News Finance Sports Adults TvMovieMusic Kids
Games Travel Shopping Jobs - Preference)
    (:predicates (name c - Client n - @xsd:string)
            (sex c - Client s - Gender)
            (age c - Client a - Number)
            (prefer c - Client p - Preference)
            (respond c - Client r - Response)))
```

여기에서 사용자의 반응들(예를 들어 Store, Delete, Spam, Reply)과 선호들(예를 들어 News, Adults, Games 등)은 Response 클래스와 Preference 클래스의 개체들(instances)로 정의할 수 있으며, 데이터 마이닝을 통해 얻은 규칙들은 공리로서 사용자 선호 온톨로지에 표현될 수 있다.

5. 시스템 구조와 실험

5.1 제안된 시스템의 구조

사용자들은 동일한 메일의 헤더와 내용에도 다르게 반응할 수 있는데, 이러한 상황은 개개인의 취향과 잠재적인 행동양식에 의해 주로 일어난다. 사용자들의 예측할 수 없는 행동양식들은 본 논문의 연구 범위 밖의 내용이기 때문에 고려하지 않는다. 우리는 사용자 마다 동일한 이메일에 대한 반응도 다를 수 있다는 가정에서 출발하여, 실제 상황에서 이 가정이 타당함을 보여주려고 하였다.

먼저, 컴퓨터공학부의 1학년에서 4학년까지의 사용자들의 개인적 선호들을 수집하였다. 샘플 이메일에 대한 사용자들의 잠재적인 반응들을 분석하기 위해, 사용자 그룹에게 샘플 이메일을 발송하였고, 그들에게 {Reply, Delete, Store, Spam} 반응 가운데 하나를 요구했다. 이 조사에서, "Reply", "Delete", "Store" 반응들은 비스팸 이메일들로 간주되고, 단지 "Spam" 반응만이 스팸으로 간주된다. 따라서 우리의 연구는 사용자들의 4가지로 세분화된 반응들을 스팸과 비스팸으로 분류하기 때문에 기존의 스팸 메일 필터링 연구들과는 차이가 있다.

제안된 구조는 그림 2에 나타난다. 그림 2에서 보이는 대로, 여러 사용자들로부터 사용자 프로파일을 수집하였고, 사용자 로그 파일은 샘플 이메일들에 대한 반응들로 작성하였다. 선호와 반응사이의 연관성 분류 규칙들을 발견하기 위해 Witten과 Frank에 의해 개발되어 잘 알려진 마이닝 도구인 WEKA[11]를 사용하였으며, 사용자 선호 온톨로지는 데이터 마이닝과 규칙 최소화 작업 후에 구성되었다. 온톨로지를 사용하는 추론 엔진 OntoEngine[14]은 사용자 선호와 개인 정보를 기반으로 전향 추론을 통해서 이메일들을 4가지 카테고리들로 분류할 수 있다.

일반적인 스팸 대응 소프트웨어는 내용 중심의 필터링 서비스를 제공한다. 그러나 제안된 시스템은 스팸에 대한 정보를 주는 것뿐만 아니라 받은 메일에 대한 사용자의 반응도 추정할 수 있기 때문에 사용자 중심의 스팸 메일 대응 서비스를 지원할 수 있다. 이러한 접근법은 새로운 방법이고, 무수한 스팸 메일로 고생하는 이메일 사용자를 위한 핵심 서비스가 될 수 있다.

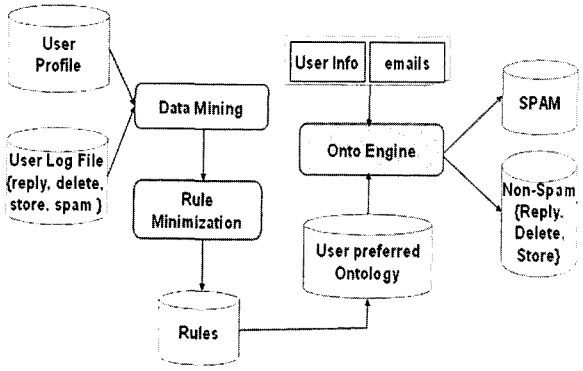

그림 2. 제안된 온톨로지 기반 스팸 메일 대응 시스템의 구조.

Fig. 2. Architecture of the proposed ontology-based anti-spam mail system.

5.2 실험

성능 척도들은 사용자 선호 온톨로지의 공리 규칙들을 이용하여 OntoEngine[14]의 추론 결과들을 평가하는데 요구된다. 스팸 메일 대응 시스템 분야에서는 오분류(misclassification), 정확률(precision), 재현율(recall) 등의 척도들이 있다[1]. 이러한 척도들은 이메일 내용으로부터 계산되므로 이메일 지향 척도라 부를 수 있다. 그러나 우리는 사용자 지향 서비스가 목적이므로 제안된 사용자 선호 온톨로지가 스팸 대응 시스템에 의미가 있다는 것을 보여주기 위해 다른 척도가 필요하다. 본 연구에서는 이러한 목적을 달성하기 위해 3가지 척도들을 제안한다.

첫째로, 공리 정확도(axiom accuracy) 또는 공리의 신뢰도(axiom confidence)는 사용자 선호 온톨로지에 있는 각 공리의 정확성을 계산하는데 유용하다. 각 공리의 전제조건과 결론을 고려해보자. 어떤 테스트 패턴의 입력 속성들이 i-번째 공리의 전제조건들과 정확하게 일치될 때, 공리에 대한 조건 매칭 점수, 즉 axiom[i].match는 1 증가한다. 동시에 테스트 패턴의 Response 값과 공리의 결론이 같다면 공리의 결론 매치, 즉 axiom[i].correct도 1 증가한다. 그러면 공리의 신뢰도는 axiom[i].correct를 axiom[i].match로 나눔으로서 계산되어진다. 당연히 신뢰도가 높은 공리들을 가진 온톨로지가 선호된다.

둘째로, 현재의 분류 정확성(classification accuracy) 척도는 이러한 스팸 대응 응용에는 적합하지 않다. 다음의 상황을 가정해보자. 갑과 을이라는 두 명의 사용자는 거의 비슷한 선호도를 가지며 학습 개체들(training instances)에 대한 반응들도 매우 유사하여, 그들은 하나의 작은 그룹으로 그룹화되며 이메일에 대한 그들의 반응들도 온톨로지에 의해 같은 것으로 결정된다. 이 경우에 갑과 을이 속해있는 사용자 그룹에 대한 반응이 어떤 공리의 특정한 전제조건으로 인해 Spam으로 추론되었다고 가정해보자. 그 공리의 특정한 전제조건에 대해 갑의 Spam 반응과 을의 Delete 반응을 포함한 모든 가능한 예제들이 데이터 마이닝 과정동안 이미 반영되었다. 그런 다음, 공리안의 특정한 전제조건과 매치된 테스트 패턴에 대하여 시스템이 갑의 반응이 옳고, 을의 반응은 틀리다고 판정한다면, 누구의 반응이 확실히 올바르다고 말하기는 어렵다. 이것은 같은 그룹에 속하는 사용자들의 개별 반응이 모든 종류의 이메일에 동일하다는 것을 보증하지는 않기 때문이다. 그래서 우리는 얼마나 많은 개체들이 각 공리에 의해 수용되는지에 대한 척도로서 공리 수용성(axiom capacity)을 소개한다. 만일 각 공리의 조건 매칭 점수의 합 (Σi axiom[i].match)이 테스트 패턴들의 총 개수와 같다면 공리 집합은 모든 개체들을 수용할 수 있고 수용성 문제는 없다. 하지만, 수천 개의 패턴들로부터 수십 개의 규칙들을 찾아내기 때문에 이것이 쉽지는 않다. 그러므로 온톨로지에 있는 공리들과 일치되지 않는 패턴들은 나이브 베이지안 분류기나 서포트 벡터 머신 같은 일반적인 내용 기반 이메일 필터로 보내서 이들이 그 패턴들을 처리하도록 한다.

한편 규칙 최소화 방법을 소개하였으므로 추상적인 문장 대신에 사용자 이해도(user comprehensibility)를 위한 단순한 정량화된 척도인 매치 길이(match length: ml)를 제안한다. 수식 (1)에 각 규칙에 대한 매치 길이 ml이 정의되어 있다.

$$ml[i] = \frac{\text{각 패턴에 있는 속성들의 수}}{\text{규칙 } [i] \text{에 있는 전제조건들의 수}} \quad (1)$$

여기서 i는 i-번째 규칙의 색인이고, 분모인 각 규칙에서의 전제조건들의 수는 테스트 패턴들과 비교되는 속성들의 수이다. 모든 규칙들의 매치 길이에 대한 평균값이 커지면, 규칙 집합이 더 단순해져서 사람들이 이해하기가 더 쉬워진다. 예를 들어 테스트 패턴 (Age = FS and RHit = S and News = F and Adults = T and Games = F)이 제시되고 (Response=Spam)이라는 반응이 나타났다고 가정하자. 두 개의 규칙들 (R7: If Age=FS and News=F then Response=Spam)과 (R8: If Age=FS and RHit=S and News=F and Adults=T then Response=Spam)이 주어진다면, ml[7]=5/2=2.5이고, *ml*[8]=5/4=1.25이다. 그러므로 R7 규칙은 정량화된 이해도 관점에서 R8 보다 일치하는 길이가 더 크다. 이 척도는 간단하면서 정량화된 값이다. Chan과 Freitas[9]도 발견된 규칙들안에 있는 항들의 평균 개수로 규칙 이해도를 측정했으나, 입력 속성들의 개수는 고려하지 않았다.

공리의 신뢰도, 수용성 및 이해도 관점에서 위의 세 척도들은 성능 평가에 도움을 준다. 제안된 스팸 메일 대응 시스템의 사용자 선호 온톨로지를 평가하기 위해 우리는 90명의 대학생으로부터 총 40개의 샘플 이메일들을 대상으로 3600개의 레코드를 수집하였다. 모든 사용자가 샘플 이메일들에 반응했으므로, 3600(=40*90)개의 이메일 반응들이 작성되었으며, 또한 모든 사용자별로 90개의 선호들이 수집되었다. 3600개의 레코드들을 2400개의 학습용 패턴들과 1200개의 테스트 패턴들로 임의로 나눈 다음, 앞의 세 가지 척도들을 관찰하기 위한 실험을 수행하였다. ID3 마이닝을 적용하기 전에 3장에서 기술한 대로 모두 14개의 속성들에 대해 IG를 계산하였다. IG값이 높은 순으로 상위 5개의 속성들인 Age, RHit, News, Adults, Games을 선택하였고, Response를 목표변수로 하여 ID3에 의해 구성된 의사 결정 트리로부터 분류 규칙들을 생성하였다.

우리는 2400개의 학습용 패턴들에 대한 ID3 데이터 마이닝을 통해 처음에는 0%보다 큰 정확성을 가진 18개의 규칙들을 얻었고, 정확하게 매치되는 개체들의 수가 5보다 작은 2개의 규칙을 제거하여 16개의 규칙을 구했다. 이렇게 한 이유는 최소 정확도(minimum correct match)와 적용 범위(coverage)를 유지하면서(supporting) 그리고 최종 규칙 집합에서 적어도 하나 이상의 "Reply" 규칙들을 남기기 위함이었다. 짐작할 수 있듯이, 사용자들은 이메일에 대해서 다양하게 반응하는데, 특히 Reply 의견은 다른 반응에 비하여 다르

게 나타나는 경향이 있다. 따라서 다양한 흥미를 가지는 사용자들로부터 Reply 반응에 대한 공통된 규칙을 찾는 것은 어렵다. 마침내 ID3 마이닝 과정을 거쳐 16개의 규칙들이 생성되었다. 생성된 규칙들의 성능을 평가하기 위해 1200개의 테스트 패턴들을 16개의 규칙들에 적용하였다. 본 연구에서 제안된 규칙 최소화 방법(HLSRP)은 표 1에서 정확도가 감소하는 순으로 12개의 공리 규칙들을 생성하였다. 기호 "&"는 각 규칙들의 논리곱인 "and"를 표현한다.

표 1. HLSRP 방법으로 유도된 공리 규칙들과 성능.
Table 1. Axiom rules derived by HLSRP and their performance.

순번	공리 규칙	매치 길이	공리 정확도
1	Age=JS & RHit=N & News=F & Adults=F & Games=F => Response=Delete	1.00	82.4%
2	Age=FS & RHit=VW & Adults=T => Response=Spam	1.67	53.0%
3	Age=FS & News=F => Response=Spam	2.50	51.9%
4	Age=JS & RHit=N & News=T & Adults=T => Response=Store	1.25	43.8%
5	Age=FS & RHit=N & News=T & Adults=T => Response=Delete	1.25	41.7%
6	Age=JS & RHit=S & News=F & Adults=F & Games=F => Response=Spam	1.00	40.0%
7	Age=JS & RHit=VS & News=T & Adults=T => Response=Reply	1.00	40.0%
8	Age=FS & RHit=VS & Games=F => Response=Spam	1.67	36.4%
9	Age=JS & News=F & Adults=T => Response=Delete	1.67	33.1%
10	Age=JS & RHit=S & News=F & Adults=F & Games=T => Response=Delete	1.00	28.0%
11	Age=JS & RHit=N & News=F & Adults=F & Games=T => Response=Reply	1.00	22.0%
12	Age=JS & RHit=N & News=T & Adults=F & Games=T => Response=Store	1.00	21.4%

표 1로부터, Age=JS and RHit=Neutral and News=False and Adults=False and Games=False 선호를 가진 용자들은 이메일을 받았을 때 보통 "Delete"로 반응했다는 것을 알려준다. 나머지 규칙들도 사용자가 받은 이메일에 대한 반응을 선택하는 이유를 같은 방식으로 설명해준다. 우리는 각 규칙에 대해 낮은 정확도가 얻어진 두 가지 이유가 있다고 생각한다. 첫 번째 이유는 News부터 RealEstates까지의 모든 가능한 11개의 카테고리들을 균등하게 고려하지 않고 임의의 샘플 이메일들을 선택해서 이메일 속성에 대한 분포가 치우쳐 있다는 점이다. 11가지 카테고리의 일부가 샘플 이메일의 대부분을 차지하고 나머지 카테고리는 없거나 극히 소수의 이메일을 가지고 있기 때문이다. 두 번째 이유는 사용자들의 반응이 같은 종류 이메일들에 대해서도 다를 수 있기 때문이다. 따라서 테스트 개체들에 대한 높은 공리 신뢰도를 가질 수 없었다. 그러나 데이터 마이닝에 의해 발견된 논리 규칙들로부터 사용자 선호 온톨로지를 구축하는 것은 다양한 종류의 이메일에 대한 사용자의 행동양식을 추정하고 어떤 이메일이 스팸인지 아닌지로 분류되는 이유를 설명하는데 기여한다.

표 2는 ID3 마이닝에 의해 생성된 규칙들을 제안된 규칙 가지치기에 의해 유도된 공리 규칙 집합과 비교한 결과를 보여준다. 표 2에서 보는 것처럼, 평균 공리 정확도는 43.81%에서 41.14%로 2.67% 낮아졌다. 그러나 규칙들의 25%가 감소되고 평균 규칙 매치 길이는 6.4% 향상되었다. 또한 제안된 HLSRP 방법은 1200개 이상의 테스트 패턴 수용능력을 보여주었다. 이것은 일부 테스트 개체들이 몇 개의 규칙들에 겹쳐 수용되었으며, 제안된 방법은 규칙 병합에 의해 거의 모든 테스트 개체들을 처리할 수 있다는 것을 의미한다. 따라서 제안된 규칙 가지치기 방법은 정확도의 큰 손실없이 두 개 또는 그 이상의 규칙들을 하나의 규칙으로 병합할 수 있다고 여겨진다. 하지만 이것으로 제안된 시스템이 내용기반 필터로 개체들을 보낼 필요가 없다고 단정하기는 어렵다. 이러한 문제를 해결하기 위해서 정확도가 높은 확실한 공리를 사용해서 1차 분류하고, 정확도가 낮은 공리에 해당하는 패턴들과 공리의 전제조건과 일치하지 않는 패턴들은 내용기반 필터 시스템으로 판단을 넘기는 하이브리드 형태로 최종 시스템을 구성하려고 한다. 한편, 가능한 짧은 규칙으로 사용자 선호 온톨로지를 구축하는 것은 바람직하다. 왜냐하면 각 사용자에게 규칙들을 전달하면 사용자는 자신의 개인적 흥미에 따라 규칙 집합을 쉽게 수정하여 자신의 사용자 선호 온톨로지를 시스템으로 피드백할 수 있게 된다.

표 2. ID3와 HLSRP 방법에 의해 유도된 두개의 공리 규칙 집합에 대한 실험결과 비교.
Table 2. Comparison on experimental results for two axiom rule sets derived by the ID3 mining and the proposed HLSRP.

방법	공리 규칙의 수				공리 정확도	수용성	매치 길이
	Reply	Delete	Store	Spam			
ID3	2	4	2	8	43.81%	1180/1200	1.25
HLSRP	2	4	2	4	41.14%	1254/1200	1.33
Improv.	0	0	0	50%	-2.67%	6.3%	6.4%

6. 결 론

본 논문에서는 스팸 메일 대응 시스템을 위한 사용자 선호 온톨로지를 구성하는 방법을 제안하였다. 제안된 방법의 중요한 특성은 사용자 선호에 따라 같은 이메일에 대해서 개별 사용자들의 다른 반응을 허용한다는 것이다. 이것은 메일의 스팸 유무를 판정할 때 모든 사용자가 같은 내용의 이메일에 대한 응답이 같을 것이라고 예상하는 현재의 스팸 메일 필터링 시스템과는 다르다. 물론 제안된 방법도 사용자들의 과거 반응과 선호에 상관없이 임의의 이메일에 대한 사용자들의 일시적인 변덕은 다룰 수 없다. 그러나 본 연구는 이메일의 내용뿐만 아니라 사용자 선호와 과거 반응 기록을 고려함으로서 개인화된 스팸 메일 대응 서비스 지원을 향한 진척을 이루었다. 본 연구의 중요한 기여도로는 사용자의 선호 온톨로지가 어떤 메일이 왜 스팸인지 또는 비스팸인지 의미

있는 방법으로 설명할 수 있다는 것이다. 또한 데이터 마이닝 기법으로 논리 규칙들을 발견하는 점과 제안된 혼합형 논리합성 규칙 가지치기 방법이 사람이 이해하고 수정하기 쉬운 규칙들을 유도한다는 것도 중요한 연구결과이다. 향후에는 각 사용자들에게 제공되는 실시간 이메일을 처리하고 스팸 메일 분류 평가도 할 수 있도록 시스템을 확장해야 한다. 또한 온톨로지내 공리들의 정확도가 낮은 문제를 해결하기 위해서, 샘플 이메일의 종류와 개수를 늘려서 사용자의 다양한 기호에 좀 더 부합하는 규칙들을 생성하고, 이를 토대로 데이터 마이닝 방법도 개선하여 정확도를 향상시키는 연구도 수행할 필요가 있다.

Acknowledgements: We are thanks to Prof. Dejing Dou at U of Oregon for his comments and help to use OntoEngine in the proposed system.

참 고 문 헌

[1] Cormack, G. V., Overview of the TREC 2005 Spam Track, http://plg.uwaterloo.ca/~gvcormac/trecspam-track05
[2] Wolfe, P., Scott, C., and Erwin, M., Anti-Spam Tool Kit, McGraw Hill, 2004.
[3] Gray, A. and Haahr, M, "Personalized Collaborative Spam Filtering," in Proc. of the First Conference on Email and Anti-Spam, 2004.
[4] Ravi, J., Shi, W., and Xu, C., "Personalized Email Management at Network Edges," IEEE Internet Computing, Vol. 9, No .2, pp. 54-60, 2005.
[5] Anti-Spam Firewall, http://www.barracuda-networ-ks.com/ns/products/ anti_spam_tech.php.
[6] Kang, S., Semi-Automatic Construction of Practical Ontology and its Application for Word Sense Disambiguation, POSTECH, PhD. Thesis, 2002.
[7] Maedche, A., "Ontology Learning for the Semantic Web," The Kluwer International Series. in Engineering and Computer Science, Vol .665, 2003.
[8] Files, C. M. and Perkowski, M. A., "Multi-Valued Functional Decomposition as a Machine Learning Method," in Proc. of ISMVL, pp. 173-178, 1998.
[9] Chan, A. and Freitas, A., "A New Classification Rule Pruning Procedure for an Ant Colony Algorithm," LNCS 3871, pp. 25-36, 2005.
[10] 강신재, 김종완, "텍스트정보와 하이퍼링크에 기반한 지능형 스팸 메일 필터링", 한국 퍼지 및 지능시스템학회 논문지, 제14권 제7호, pp.895-901, 2004.
[11] Witten, I. H. and Frank, E., Data Mining: practical machine learning tools and techniques, 2nd ed, Morgan Kaufmann, 2005.
[12] http://en.wikipedia.org/wiki/Karnaugh_map
[13] McDermott, D. and Dou D., "Representing disjunction and quantifiers in RDF," in Proc. Int'l Semantic Web Conference, pp. 250-263, 2002.
[14] Dou, D,, McDermott, V., and Qi, P., "Ontology translation on the semantic web," Journal of Data Semantics, Vol. 2, pp. 35-57, 2004.

저 자 소 개

김종완 (Jong-Wan Kim)
1987년 : 서울대학교 컴퓨터공학과 공학사
1989년 : 서울대학교 컴퓨터공학과 공학석사
1994년 : 서울대학교 컴퓨터공학과 공학박사
1995년~현재 : 대구대학교 컴퓨터·IT 공학부 교수
1999년~2000년 : 미국 U. of Mass at Amherst 방문교수
2006년~현재 : 미국 U. of Oregon 객원교수

관심분야 : 인공지능, 스팸 대응 시스템, 데이터마이닝, 퍼지 시스템, 온톨로지
Phone : 053-850-6575
E-mail : jwkim@daegu.ac.kr

김희재 (Hee-Jae Kim)
1992년 : 대구가톨릭대학교 통계학과 이학사
1994년 : 대구가톨릭대학교 전산통계학과 이학석사
2002년 ~ 2005년 : 대구한의대학교 멀티미디어학부 초빙교원

2005년 : 대구대학교 컴퓨터정보공학과 박사수료
2006년 ~ 현재 : 대구대학교 컴퓨터정보공학과 겸임교수

관심분야 : 퍼지시스템, 인공지능, 데이터마이닝, 온톨로지
Phone : 019-520-1022
E-mail : heejae0305@daegu.ac.kr

강신재 (Sin-Jae Kang)
1995년 : 경북대학교 컴퓨터공학과 공학사
1997년 : 포항공과대학교 컴퓨터공학과 공학석사
2002년 : 포항공과대학교 컴퓨터공학과 공학박사
1997년~1998년 : SK Telecom 정보기술 연구원 주임연구원
2002년~현재 : 대구대학교 컴퓨터·IT공학부 조교수
2007년~현재 : 오스트리아 U. of Innsbruck, DERI 연구소 방문연구원

관심분야 : 자연어처리, 온톨로지, 시맨틱 웹, 웹 서비스
Phone : 053-850-6584
E-mail : sjkang@daegu.ac.kr