한국지능시스템학회 논문지 2009, Vol. 19, No. 3, pp. 343-349

사용자 클러스터별 이메일 반응 분포 계산 및 사용자 선호 스팸 메일 대응 시스템 구축

From Computing Distribution of Email Responses for Each User Cluster To Construct User Preference based Anti-spam Mail System

김종완
Jongwan Kim

대구대학교 컴퓨터·IT공학부
School of Computer and Information Technology, Daegu University
jwkim@daegu.ac.kr

요 약

본 논문은 전자메일 사용자별로 제공받은 사용자 선호 정보를 클러스터링하여 사용자 클러스터를 만든 후, 사용자 클러스터들의 전자메일 반응 분포를 계산함으로써 사용자 취향에 따라 동일한 전자메일에 대해서도 다른 반응을 가질 수 있다는 사실을 보이려고 한다. 본 논문에서는 사용자 선호도를 채용하여 보통의 내용기반 방식과는 다른 스팸 메일 대응 시스템을 구축하는 접근법을 제안한다. 제안된 방법은 전자메일 내용으로부터 유도된 전자메일 카테고리 정보뿐만 아니라 사용자 선호 정보도 고려한다. 데이터마이닝 프로세스로부터 유도된 중요한 개념과 규칙들을 정형적으로 표현하기 위하여 사용자 온톨로지를 구축하고, 규칙 최적화 방법을 적용하여 불필요한 규칙들을 제거한다. 실험결과는 제시된 사용자 선호 기반 시스템이 정확률과 시스템이 유도한 규칙, 사용자 이해도 면에서 좋은 결과를 제시한다.

키워드 : 사용자 클러스터, 사용자 선호 온톨로지, 스팸 메일 대응 시스템, 전자메일 카테고리 분류, 데이터마이닝

Abstract

In this paper, it would be shown that individuals can have different responses to the same email based on their preferences through computing the distributions of user clusters' email responses from clustering results based on email users' preference information. This paper presents an approach that incorporates user preferences to construct an anti-spam mail system, which is different from the conventional content-based ones. We consider email category information derived from the email content as well as user preference information. We also build a user preference ontology to formally represent the important concepts and rules derived from a data mining process and then apply a rule optimization procedure to exclude unnecessary rules. Experimental results show that our user preference based system achieves good performance in terms of accuracy, the rules derived from the system and human comprehensibility.

Key words : user cluster, user preference ontology, anti-spam mail system, email categorization, data mining.

1. 서 론

스팸 메일은 수신인이 원치 않는 전자메일(이하 메일로 사용)을 말하며, 수년전까지는 귀찮은 일이었으나, 오늘날에는 사용자 생산성과 시스템 안정성에 심각한 위협을 주는 스파이 또는 피싱(phishing) 기능을 수행하고 있다. 대부분 메일 클라이언트 소프트웨어는 블랙리스트나 키워드기반 필터 기술을 채용하여 초기에 효과적이었지만, 스패머가 사용하는 기술이 키워드 필터를 무력화하도록 지능화됨에 따라 정확도가 감소하고 있다[1]. 이러한 스팸 필터링 문제는 문서 분류의 특별한 종류로 볼 수 있다. 다양한 기계학습 알고리즘들인 나이브 베이즈 분류기(NBC: naive Bayesian classifier)와 서포트 벡터 머신(SVM: support vector machine)은 메타데이터에 메일 분류 작업을 위해 사용되어 왔다[2,3]. 최근 Cormack과 Lynam[4]은 온라인 지도 필터 평가(on-line supervised filter evaluation)를 수행하였다. 그 실험에서 특정인이 8개월간 받은 시간 순서로 정리된 49,086 메일 메시지를 대상으로 6가지 공개 소스 스팸 필터들의 11가지 변형들을 시험하였다. 이 필터는 테스트 집합이 대규모라는 점과 검열되지 않은 순수 메일들을 사용하고 각 필터에 순차적으로 피드백을 제시했다는 점에서 이전의 방법들과는 다르다. 이전의 연구들로부터, NBC와 SVM 사

접수일자 : 2009년 2월 27일
완료일자 : 2009년 6월 4일
이 논문은 2008학년도 대구대학교 학술연구비 지원에 의한 논문임

이에는 내용기반 필터링 응용에서 큰 차이가 없음을 알 수 있다.

이들 스팸 필터들이 통계적 정확성을 향상시키는 반면에, 여전히 비스팸이 스팸으로 분류되거나 스팸이 비스팸으로 분류되는 문제가 있다. 비스팸 용어를 대신하여 지겨운 스팸 통조림 광고가 아니라 맛있는 훈제고기를 의미하는 햄(ham)이란 용어가 함께 사용된다. 어떤 메일은 누군가에게는 스팸이지만 실제 상황에서 다른 사람들에게는 햄일 수 있고, 메일에 대한 사용자 반응은 개인 취향에 따라 다를 수 있다는 2가지 사실에서 연구를 시작하였다. 본 논문에서는 대학생 참여자로부터 사용자 취향과 메일 반응을 수집하고 연관성 분류 마이닝을 수행한 후, 정형 언어로 사용자 선호 온톨로지(user preference ontology)를 정의하는 규칙들을 생성하고, 구성된 온톨로지를 활용한 사용자 선호 기반 스팸 메일 대응 시스템 구축 방법을 설명한다. 실험 결과는 제시된 방법이 타당함을 보여준다.

이 논문은 다음과 같이 구성된다. 2장에서 관련 연구를 소개하고, 3장에서 시스템 구조 및 연구 수행동기를 설명한다. 4장은 사용자 선호 스팸 메일 대응 시스템 구축 방법을 기술하고, 5장에서 실험 환경 및 실험 결과를 제시하며, 결론은 6장에 기술된다.

2. 관련 연구

기존의 사용자 취향을 고려한 스팸 대응 시스템들은 다음과 같은 특징을 갖는다. 동일한 네트워크 그룹에 속한 다른 사용자들의 추천을 기반으로 하거나[5], 같은 사용자의 다른 메일 주소 정보로부터 판정한 정보를 활용[6], 또는 특정 시스템의 추천을 수락 또는 수락하지 않는 사용자의 판단에 의존한다[7]. 그러나 본 연구의 목표는 순수하게 사용자의 취향과 사용자의 반응에 의존한 사용자 선호 온톨로지 기반의 스팸 메일 대응 시스템을 개발하는 것이다. 특히 본 논문에서는 사용자 선호 정보에 기반한 클러스터링 결과로부터 사용자 그룹의 메일 반응 분포를 계산하여 본 연구자의 이전 연구인 사용자 선호 스팸 대응 시스템[8]이 타당함을 보여주고, 메일 카테고리를 결정하는 분류기를 구축하고 사용자 선호 온톨로지 공리들이 유용함을 보임으로써 제안된 방법을 검증하고자 한다.

전체 사용자와 개인 사용자를 동시에 고려하는 동적인 개인화(dynamic personalization) 접근법이 Segal에 의해 제안되었다[9]. 전체 또는 일반 스팸 대응 필터가 대규모 사용자 수집 메일로 스팸과 햄을 훈련하는데 비해, 개인 훈련 필터들은 개인별로 자신의 스팸 정의를 제공하는 장점을 가진다. 두 접근법들은 장단점이 있다. 동적인 개인화의 기본 아이디어는 개인화 데이터양이 증가하면 개인 데이터를 적용하고, 개인 데이터양이 적을 때는 전체 데이터를 사용하는 것이다. 이상적으로 테스트 집합에 있는 개별 메시지는 개별 사용자를 위해 별도의 식별자가 붙지만, 테스트 집합의 특정 사용자를 위한 사용자 지정 식별자는 없다. Segal은 사용자 기준으로 저장된 판단을 가진 공용 집합을 알지 못하기 때문에, 모든 사용자는 특정 메시지에 같은 식별자를 지정하고 같은 판단을 한다는 비현실적인 가정을 세운다[9]. 이러한 가정의 결과로서, 실험에서 개인화 분류의 가치를 낮게 평가하고 있는데, 이 점이 본 연구와 다르다. 본 연구에서는 메일 수신자를 고려하지 않고 메일 집합을 수집하였고, 모든 메일에 대해 사용자별로 자신의 선호도를 지정한 메일 집합 대상으로 데이터마이닝을 수행하였다. 그 후에 어떤 메일이 스팸인지 아닌지를 사용자 선호도와 메일 카테고리(email category) 정보에 의해 판단한다. 동적 개인화의 아이디어는 스팸에 대한 다른 의견을 가진 사용자 커뮤니티에 효과적일 수 있으며, 본 연구도 사용자별로 완전히 다른 메일 반응을 가정하므로 유사한 결과가 예상된다.

3. 시스템 구조

본 장에서는 사용자 선호 스팸 메일 대응 시스템 구조와 메일에 대한 사용자 반응이 사용자 그룹별로 다른 분포를 가진다는 사실을 보임으로써 본 논문의 연구동기가 충분하다는 사실을 제시한다.

3.1 시스템 구조

사용자들은 동일한 메일에도 다르게 반응을 할 수 있는데, 이러한 상황은 주로 개인의 취향과 잠재적인 행동양식에 의해 일어나며, 사용자들의 예측할 수 없는 행동양식들은 연구 범위를 벗어난다. 우리는 동일한 메일에 대한 사용자별 반응이 다를 수 있다는 가정에서 출발하여, 실제 상황에서 이 가정이 타당함을 보여주려고 한다.

제안된 시스템 구조는 그림 1에 나타난다. 그림 1에 보이는 대로, 참가한 사용자들로부터 사용자 프로파일(user profile)을 수집하였고, 사용자 로그 파일(user log file)은 샘플 메일들에 대한 사용자 반응으로 만들었다. 특히 주어진 메일에 대한 사용자 반응은 사용자 선호와 메일 내용 기반의 메일 카테고리에 의존하므로, 메일 카테고리(이하 ECat) 정보가 시스템에 포함되어야 한다. 일반적으로 성인, 금융, 취업 같은 카테고리로 메일을 분류하는데 NBC나 SVM이 사용된다. 1장에서 언급했듯이, SVM은 이런 종류의 분류에 유용한 접근법이므로 본 연구에서는 메일 카테고리 분류기로서 표준 SVM[3]을 사용한다.

선호도와 반응 사이의 연관성 분류 규칙들을 발견하기 위해 WEKA[10]를 사용하였으며, 사용자 선호 온톨로지는 데이터마이닝과 규칙 최소화 작업 후에 만들어졌다. 온톨로지를 사용하는 추론 엔진 OntoEngine[11]은 사용자 선호, 메일 카테고리, 개인 정보를 기반으로 전향 추론(forward chaining inference)을 통해서 메일들을 4개의 카테고리(Spam, Reply, Delete, Store)로 분류할 수 있다.

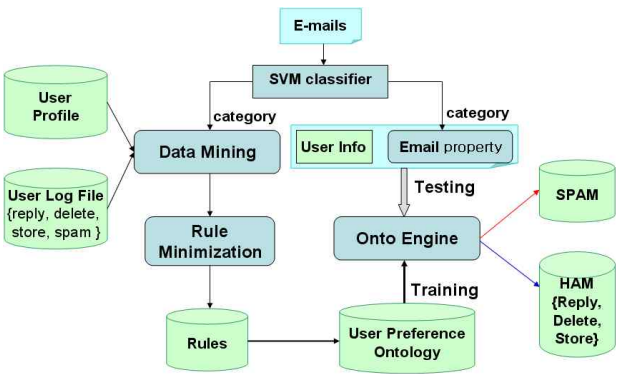

그림 1. 제안된 사용자 선호 온톨로지 기반 스팸 메일 대응 시스템의 구조

Fig. 1. Architecture of the proposed user preference ontology-based anti-spam mail system

3.2 사용자 정보/반응 수집

메일에 대한 사용자 반응을 조사하기 위해서 사용자 선호와 메일에 대한 사용자 반응 클래스를 표현하도록 사용자 프로파일을 설계한다. 상용 웹 메일들에서 사용자의 개인 정보를 요구하는 것처럼, 사용자 프로파일에 포함할 특성들로 15가지 {Age, Gender, RHit, Adults, Entertainment, Finance, Jobs, Shopping, Travel, News, Sports, Kids, IT, RealEstates, Etc}를 선택하였다. 나이(Age) 속성은 실험 참여자가 대학생이므로 두 그룹 FS(1-2학년)와 JS(3-4학년)로 나누었다. RHit는 어떤 메일이 스팸으로 간주되기 전에 스팸과 관련된 용어들이 본문에 얼마나 많이 출현하는지를 나타내는데, 메일 내용을 고려하지 않고 언어 표현이 사람에게 익숙하기 때문에 약함(W), 보통(N), 강함(S)의 언어 항으로 구별하였다. 만약 사용자들이 강한 스팸 필터를 원한다면, RHit 값으로 약함(W)을 선택할 것이고, 약한 필터를 원할 때는 RHit=S를 선택하게 된다. 사용자 선호도가 사용자의 메일 반응에 영향을 미치는 가를 알기 위해, 사용자 선호도 및 반응과 함께 기계학습에 의해 분류된 ECat 정보도 데이터마이닝 프로세스에 이용한다.

메일 수신자들은 받은 메일에 대해 대개 4가지, Reply(응답), Delete(삭제), Store(보관), Spam(스팸)으로 반응한다. 이러한 사실에 근거하여 샘플 메일들, 실험에 참여한 사용자들의 개인 정보와 취향, 샘플 메일에 대한 반응들을 수집하였다. 개인 정보에 해당하는 속성들인 Age는 {FS, JS}로, Gender는 {M, F}의 이진으로 주어지며, 대부분의 다른 취향 속성들도 이진으로 표현된다. 예를 들어, 만약 사용자가 Entertainment에 관심이 있다면 그 속성에 대해 true(T)를 선택하며, 관심을 가지고 있지 않으면 false(F)를 선택한다. 반면에, RHit에서는 다진 속성값 (W, N, S)을 가지고, ECat 속성은 메일 카테고리를 나타내기 위해 Adults부터 RealEstates까지의 11가지 카테고리 속성과 카테고리가 불분명한 기타(Etc) 속성을 포함해서 12가지 값을 사용한다. 사용자 반응 분류의 목표 변수 또는 출력 변수인 Response는 4가지 범주형 값을 가진다.

3.3 연구동기 점검

본 연구의 동기(motive)인 동시에 가정인 '사용자 선호도가 메일 카테고리 정보와 함께 메일에 대한 개별 사용자의 반응에 영향을 미칠 수 있다'의 유효성을 확인하기 위하여 Witten과 Frank[10]가 개발한 데이터마이닝 도구인 WEKA의 EM 클러스터링 기법을 수집 데이터에 적용한다. 기대-최대로 불리는 EM(expectation maximization) 알고리즘은 1977년에 Dempster 등이 제안하였으며[12], 기계학습과 컴퓨터비전 등에서 광범위하게 사용되는 클러스터링 방법으로서, 관찰되지 않고 잠재된 변수들에 의존하는 확률모델에서 가장 큰 기대치를 갖는 패러미터 추정치를 발견할 목적으로 통계학에서 사용된다. 실험에 참여한 n명의 사용자(실험에서는 n=74)의 11가지 속성이 포함된 선호 레코드들 대상으로, EM 클러스터링을 수행한 결과 단지 2개의 클러스터가 최적으로 발견되었다. 그러나 두 개의 클러스터는 너무 작아서, 사용자 개체인 n을 클러스터의 수로 설정하고 상향식으로 EM 클러스터링을 한 결과 57개의 클러스터가 생성되었다. 하지만 실제로는 n개의 개체가 6개의 클러스터로 분류되므로 상향식 클러스터링 결과는 6개의 클러스터를 생성한다. 본 연구에서는 이 두 가지 경우에 사용자 선호 정보에 따른 사용자 클러스터의 메일 반응 분포를 계산하였다. 표 1은 각 메일 카테고리별로 개별 반응인 Spam, Delete, Store, Reply의 수를 전체 반응수로 나눈 개별 분포값을 보여준다. 대부분의 메일들은 성인(adults), 오락(entertainment), 금융(finance), 취업(jobs), 쇼핑(shopping), 여행(travel)의 6가지 카테고리로부터 수집되었다. 표 1에서 사용자들이 일부 메일을 기타(etc)로 분류한 것은 해당 메일이 자신이 생각하기에는 미리 명시된 6가지 카테고리에 속하지 않는다고 판정한 경우이다. 어떤 개별 클러스터안의 각 사용자 선호에 대한 반응의 4가지 백분율 합이 가끔 100%가 되지 않는 이유는 소수의 메일에 대하여

표 1. 개별 사용자 클러스터에 대하여 주관적인 사용자 선호도에 따른 사용자 반응 분포

Table 1. Distribution of user responses according to different user subjective preferences for each user cluster

클러스터 수	클러스터 번호	사용자 반응	주관적인 사용자 선호도 (%)						
			성인	오락	금융	취업	쇼핑	여행	기타
2	0	Spam	67.3	12.3	54.0	16.8	19.1	12.2	25.8
		Delete	18.7	54.7	22.7	33.0	48.6	43.1	45.0
		Store	10.8	29.5	15.5	29.6	29.4	40.4	18.3
		Reply	3.2	3.54	7.8	20.7	2.9	4.4	10.9
	1	Spam	85.7	11.3	33.7	3.9	14.0	17.3	11.7
		Delete	14.0	56.7	51.2	40.2	55.7	41.3	61.7
		Store	0.4	30.7	12.7	52.0	28.4	38.3	15.6
		Reply	0	1.3	2.4	3.9	1.7	3.1	7.5
6	0	Spam	73.5	8.8	58.2	13.0	6.1	4.3	23.9
		Delete	14.0	56.5	25.9	29.0	48.3	41.9	47.2
		Store	9.9	32.5	9.6	37.4	42.9	50.8	16.3
		Reply	2.6	2.2	6.0	20.6	2.7	3.1	10.4
	1	Spam	86.7	3.5	41.4	4.3	48.6	11.0	9.1
		Delete	13.3	61.2	17.2	26.1	39.4	37.0	63.6
		Store	0	31.8	34.5	65.2	10.1	48.3	0
		Reply	0	3.5	6.9	4.3	1.8	3.7	2.6
	2	Spam	61.2	15.8	33.3	20.0	35.9	22.2	37.8
		Delete	25.2	60.4	30.8	28.6	49.8	46.2	36.5
		Hold	9.7	19.6	23.9	34.3	11.9	26.2	16.6
		Reply	3.9	4.2	12.0	17.1	2.4	5.4	9.1
	3	Spam	80.6	11.5	34.1	0	16.1	21.4	16.1
		Delete	19.4	70.0	35.2	36.0	66.2	37.7	63.1
		Store	0	18.3	27.5	56.0	17.5	40.2	13.3
		Reply	0	0.2	3.3	8.0	0.1	0.7	7.3
	4	Spam	86.8	3.5	34.4	5.3	5.2	12.5	2.1
		Delete	12.5	32.7	59.6	64.9	51.7	47.6	61.7
		Store	0.7	60.4	5.5	29.8	39.0	33.7	23.0
		Reply	0	3.5	0.5	0	4.0	6.2	9.0
	5	Spam	100	97.6	100	0	4.0	100	4.8
		Delete	0	0	0	0	0	0	83.3
		Store	0	2.4	0	0	96.0	0	7.1
		Reply	0	0	0	0	0	0	4.8

일부 사용자들이 자신의 반응을 제공하지 않았기 때문이다. 표 1의 결과들은 개별 사용자 그룹이 사용자 선호에 따라 반응에서 다른 분포를 갖는다는 것을 보여준다. 결과적으로 본 연구동기인 사용자의 주관적인 선호도가 주어진 메일과 메일 카테고리 정보를 고려할 경우 사용자 반응에 영향을 미친다는 사실이 타당함을 입증하였다.

4. 스팸 메일 대응 시스템 구축

사용자 선호를 고려한 스팸 메일 대응 시스템을 구축하기 위해서 사용자 취향 기반으로 행동양식을 형식적으로 정의한 도메인 온톨로지가 도움이 될 수 있다. 스팸 대응 온톨로지 생성은 본 연구자의 이전 연구[8]에서 설명된 바와 같이 3단계로 수행한다.

4.1 분류 규칙 발견

사용자 선호도가 기록된 프로파일과 메일 카테고리 정보, 그리고 샘플 메일에 대한 사용자 반응을 포함한 로그 파일의 세 가지 정보들 사이의 알려지지 않은 상관관계를 밝혀야 한다. 이를 위한 마이닝 도구로 대표적인 의사결정 트리(decision tree) 알고리즘인 ID3[10]을 선택하였다. 3장에서 설명했듯이 샘플 데이터들은 대부분이 이진 속성을 가지고 일부 범주형 속성들로 구성되어 있으므로, ID3 알고리즘이 데이터 집합으로부터 대표 규칙들을 발견하기에 적합하다. ID3 마이닝을 수행하여 의사결정 트리를 생성시킨 후, 트리의 각 경로를 하나의 규칙으로 기술함으로써 의사결정 트리를 규칙들로 변환한다. 어떤 규칙이 좋은 지를 평가하기 위해서 그 규칙과 일치하는 테스트 패턴들의 비율을 계산함으로써 정확성이 계산된다.

4.2 논리합성 기반 규칙 최소화

의사결정 트리에서 얻어진 규칙들 가운데 중복된 규칙들을 제거하고 이해하기 쉬운 규칙들을 선택하기 위하여 규칙 최소화 과정을 적용한다. 본 연구에서는 이전 연구에서 제안한 불리언 변수와 범주형 변수들을 함께 다룰 수 있고 논리합성(logic synthesis)으로부터 영감을 얻은 일종의 규칙 가지치기 접근법인 혼합형 논리합성 규칙 가지치기(hybrid logic synthesis rule pruning: HLSRP)을 사용한다[8]. 카르노 맵[13]은 불리언 논리 단순화를 이해하기 위한 간단한 방법으로서, 동일한 함수에 대하여 2개 이상의 단순화된 논리 표현들을 발견하는 것이 가능하다. 예를 들면, F(A, B, C)=$\sum(0,1,2,3,5,7)$는 세 개의 입력 변수 A, B, C로 구성되고, 함수 F는 6개의 최소항(min term) {000, 001, 010, 011, 101, 111}을 가진다. 이때 함수 F는 F=C+A′C와 F=A′+AC의 두 종류 논리 최소화가 가능하다. 즉 함수 F가 고정되어도 두 개의 서로 다른 논리 표현이 가능하지만, 두 표현은 논리 관점에서 동등하다.

데이터마이닝으로부터 유도된 규칙 집합에는 몇 가지 변수들이 있다. 대부분의 변수들은 불리언 또는 이진 변수이지만, RHit와 ECat 같은 다진 변수도 있다. HLSRP에서는 특정 변수의 둘 또는 그 이상의 논리가 다르다면, 해당 변수를 포함한 규칙들은 해당 변수가 삭제된 단순한 규칙으로 유도된다. ID3 마이닝으로 유도된 규칙들을 최소화하는 자세한 HLSRP 알고리즘은 다음과 같다.

```
1. 사용자 프로파일, 로그파일, 메일 카테고리에 ID3 마이닝
   수행하여 규칙 생성
2. 4가지 반응별로 단계 3의 루프를 수행
3. 각 반응그룹에 대하여 주어진 규칙 모두를 대상으로
   검사
   3.1 정확률이 가장 높은 피봇 규칙을 발견
   3.2 피봇 규칙과 비교하여 하나의 조건만 다른 규칙들을
       검사하고 발견
       if (해당 규칙이 발견) {
         if (한 쌍만 발견)
         then 두 개의 규칙을 하나의 규칙으로 병합;
         else if (둘 이상의 쌍이 발견)
              then {
                정보획득량 감소순으로 후보 속성을 정렬;
                낮은 정보획득량 속성을 병합후보로 선정;
                해당 속성 삭제한 하나의 규칙으로 병합;
         }
       }
   3.3 정확률이 높은 규칙을 피봇 규칙으로 바꿈
   3.4 다음 반응그룹에 대하여 단계 3을 반복 수행함
4. 분류 규칙을 도메인 온톨로지로 변환
```

4.3 온톨로지 표현

분류 규칙들을 온톨로지와 온톨로지 사이의 매핑을 표현하는데 적합한 강한 형식의 일차 논리 언어(strongly-typed first-order logic language)인 Web-PDDL[11]을 사용하여 온톨로지로 표현한다. 이전의 결과들을 기초로 Web-PDDL로 사용자 선호 온톨로지를 정의하고, 클래스(class)와 프로퍼티(property)로서 아래의 개념들(concepts)을 선택한다.

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

Properties (Predicates): name, sex, age, prefer, category, respond

Web-PDDL에서는 위의 개념들을 다음과 같이 표현할 수 있다.

```
(define (domain anti_spam)
   (: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 RHit ECat Age - @xsd:string Response - Action)
   (:Objects Reply Store Delete Spam - Response
           Adults Entertainment Finance ... - Preference)
   (:predicates (name c - Client n - @xsd:string)
           (sex c - Client s - Gender)
           (age c - Client a - Age)
           (prefer c - Client p - Preference)
           (category e - Email e - ECat)
           (respond c - Client e - Email r - Response)))
```

여기에서 사용자의 반응들과 선호들은 Response 클래스와 Preference 클래스의 개체들로 정의할 수 있으며, 데이터마이닝을 통해 얻은 규칙들은 공리로서 사용자 선호 온톨

로지로 표현된다. 예를 들어 if Age = FS and ECat = Adults and Adults = T then Response = Store 규칙은 하나의 공리로서 Web-PDDL을 이용해 아래와 같이 표현될 수 있다.

```
(axioms:
   (forall (c - Client e - Email)
      (if (and (age c "FS") (prefer c Adults)
               (Category e "Adults"))
          (respond c e Store)))
```

5. 실험 환경 및 결과

이 장에서는 전통적인 내용기반 메일 분류기들이 제안된 환경에서 적합하지 않은 이유를 설명하고, 실험 환경과 실험 결과도 기술한다.

5.1 본 연구에 내용기반 필터들이 적합하지 않은 이유

전통적인 내용기반 필터들이 사용자 선호를 고려하는 본 연구에 적합하지 않은 이유를 설명하겠다. 표 2는 사용자들이 메일들에 대하여 다양한 반응을 제공한다는 사실을 보여주는 반응 분포를 나타낸다. 표 2안의 각 M_i는 i번째 메일을 나타내는 식별자이다. 사용자에 따라 같은 메일도 다르게 반응하므로 본 연구의 시스템과 전통적인 시스템을 비교할 직접적인 수단이 없다는 사실이 유추된다.

표 2. 샘플 메일들에 대한 사용자 반응 분포
Table 2. Distribution of users' responses to sample mails

	M_1	M_2	M_3	M_4	M_5	...
Reply	5	2	1	17	1	
Spam	31	29	24	6	25	...
Store	10	3	13	22	9	
Delete	21	30	28	18	31	

5.2 실험 환경

사용자 선호도 고려 스팸 메일 대응 시스템을 평가하기 위하여, 성인, 오락, 금융, 취업, 쇼핑 및 여행의 6가지 카테고리를 주로 포함한 메일들을 수집하였다. 무명의 사용자들에게 전송된 메일들로부터 각 카테고리별로 150개씩 모두 900개의 한국어와 영어 메일을 수집하였다. 하지만 실험에서는 SVM으로 메일 카테고리 분류가 이루어진 249개의 영문 메일만 사용하였으며, 실험에 참여한 74명의 대학생들로부터 자신들의 선호 정보와 함께 249개 메일에 대한 반응 정보를 수집하였다. 소수의 사용자는 어떤 메일에 대한 자신의 반응을 결정하는데 어려움이 있어서 전체 18,426 (= 74 * 249) 레코드 가운데 91개의 레코드는 비어 있으므로 실험에서 배제하였다. 결과적으로 18,335 레코드를 가진 데이터 집합이 얻어졌고, 각 레코드는 메일 카테고리, 사용자 선호 정보, 반응 정보를 포함한다. 실험에서는 전체 레코드의 약 2/3인 12,223 레코드들을 학습과정에서 사용하고, 나머지는 규칙 성능평가를 위한 테스트 과정에서 사용하였다.

스팸 대응 시스템에 사용된 사용자 선호 온톨로지의 유용성을 보이기 위한 온톨로지 관점의 성능척도들이 요구된다. 내용기반 스팸 메일 대응 시스템 분야에서는 오분류율, 정확률, 재현율 등의 척도들이 사용된다[4]. 이 척도들은 메일 내용으로부터 계산되므로 메일 중심 척도라 할 수 있지만, 본 연구는 사용자 중심 서비스가 목적이므로 다른 척도가 필요하다. 본 연구자가 이전에 제안한 척도[8]을 사용하여 제시된 시스템의 성능 우위를 보이겠다.

첫째, 공리 정확성(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 점수는 1 증가한다. 동시에 테스트 패턴의 Response와 공리의 결론이 같다면 공리의 결론 매칭 점수도 1 증가한다. 그러면 공리의 신뢰도는 결론 매칭 점수를 조건 매칭 점수로 나눔으로서 계산된다. 당연히 신뢰도가 높은 공리들을 가진 온톨로지가 선호된다.

둘째, HLSRP 규칙 최소화 방법으로 유도된 규칙의 사용자 이해도(user comprehensibility)를 측정하기 위한 단순 정량화 척도인 매치항 비율(matched term ratio: mtr)을 제시한다. 각 규칙별 mtr은 각 패턴에 있는 속성들의 수를 규칙의 전제조건에 있는 속성들의 수로 나누어 구해진다. 모든 규칙들의 매치항 비율에 대한 평균값이 커질수록, 규칙 집합이 더 단순하고 사람들이 이해하기가 쉬워진다. 예를 들어, 테스트 패턴 (Age=FS and ECat=Adults and RHit=S and News=F and Adults=F and Jobs=T)이 제시되고 (Response=Spam)이라는 반응이 나타났다고 가정하자. 두 개의 규칙들 (R1: If Age=FS and ECat=Adults and News=F and Adults=F then Response=Spam)과 (R2: If Age=FS and ECat=Adults and RHit=S and News=F and Adults=F then Response=Spam)이 주어진다면, mtr[1]=6/4=1.5이고, mtr[2]=6/5=1.2이다. 그러므로 R1 규칙은 정량화된 이해도 관점에서 R2 보다 매치항 비율이 더 크다. Chan과 Freitas[14]도 발견된 규칙들안에 있는 항들의 평균 개수로 규칙 이해도를 측정했으나, 입력 속성들의 개수는 고려하지 않는 점이 다르다.

5.3 실험 결과

사용자 선호 온톨로지의 공리규칙 유용성 실험은 아래와 같이 수행하였다. ID3 마이닝 실험에서 304개의 규칙들을 얻었고, 규칙 최소화 방법인 HLSRP를 적용하여 238개의 규칙들을 얻었으며, 6,112 테스트 개체 대상으로 이들 규칙의 성능을 평가하였다. 높은 정확률을 갖는 12개의 대표 규칙들을 공리로 해석한 내용이 표 3에 주어진다. 표 3의 각 규칙은 사용자가 자신의 선호도에 따라 특정한 메일에 반응하는 방식을 설명하고 있다. 예를 들면, 첫 번째 공리규칙은 3-4학년 남학생이 성인, 오락, 아이를 싫어하고 스포츠를 좋아하는 선호도를 가진다면, 섹스 같은 성인 내용을 포함하는 메일을 100% 스팸으로 분류하는 반응을 보일 것이라는 사실을 나타낸다. 사실 거의 모든 성인 메일들은 전통적인 내용기반 필터들에서 스팸으로 분류되는 반면에, 본 실험에서는 소수의 사용자들은 다르게 생각할 수 있다는 가능성을 제시한다. 표 3의 8번째 규칙을 보면, 성인 카테고리 메일을 100% 스팸으로 분류하지 않고 있다. 바로 이점이 제안된 사용자 선호 기반 접근법이 사용자 특성에 부합하는 솔루션이 될 수 있다는 점을 납득시킨다.

표 4는 HLSRP에 의해 생성된 규칙 집합과 ID3 마이닝 수행 후 얻어진 규칙 집합의 성능을 요약한 것이다. 실험결과로부터 두 방법의 온톨로지 정확성은 동일하고, 규칙 최

표 3. HLSRP 알고리즘에 의해 유도된 공리 규칙들과 성능
Table 3. Axiom rules derived by HLSRP algorithm and their performance

	공리 규칙	정확률	매치항 비율
1	Age=JS & Gender=M & ECat=Adults & Adults=F & Entertainment=F & Sports=T & Kids=F ==> Response=Spam	1	1.86
2	Age=FS & ECat=Adults & Adults=F & Travel=F ==> Response=Spam	1	3.25
3	RHit=N & ECat=Entertainment & Entertainment=F ==> Response=Delete	1	4.33
4	RHit=W & ECat=IT & Finance=T & Sports=F & Kids=F ==> Response=Delete	1	2.6
5	RHit=W & ECat=News & Married=F & Finance=T & Shopping=F & Sports=F ==> Response=Delete	1	2.17
6	RHit=W & ECat=Travel & Adults=F & Entertainment=F & Finance=F & Shopping=F & Travel=T & Kids=F ==> Response=Store	0.978	1.625
7	RHit=N & ECat=Travel & IT=T & Travel=F & Sports=T ==> Response=Delete	0.95	2.6
8	Age=FS & ECat=Adults & Adults=F & Shopping=T & Travel=T ==> Response=Spam	0.947	2.6
9	RHit=W & Gender=M & ECat=Shopping & Adults=F & Finance=F & Shopping=F & Travel=T & Sports=T ==> Response=Delete	0.938	1.625
10	RHit=W & Gender=F & ECat=Jobs & RealEstate=F & Entertainment=T & Shopping=T & Sports=F & Kids=F ==> Response=Delete	0.867	1.625
11	RHit=N & ECat=Etc & IT=F & Finance=F ==> Response=Delete	0.853	3.25
12	RHit=N & ECat=Entertainment & Entertainment=T & Shopping=T & Sports=T ==> Response=Store	0.810	2.6

소화를 수행하면 규칙수가 21.7% 줄어들고, 평균 매치항 비율은 13.2% 개선됨을 알 수 있다. 짧은 길이의 감소된 규칙 집합이 사용자 선호 규칙 구성에 바람직하다는 것은 직관적으로 타당하다.

표 4. ID3와 HLSRP로부터 유도된 두 공리 규칙 집합의 성능 비교
Table 4. Compared performance of two axiom rule sets derived from the ID3 and the HLSRP method

비교 방법	공리 규칙의 수					평균 매치항 비율
	Reply	Delete	Store	Spam	계	
ID3	7	152	64	81	304	2.19
HLSRP	6	110	55	67	238	2.48
개선도(%)	14.3	27.6	14.1	17.3	21.7	13.2

6. 결 론

현재의 내용기반 스팸 메일 필터는 동일한 메일에 대해서 같은 반응을 보일 것이라고 가정하는 범용 필터링 방식을 채택하고 있다. 본 연구는 사용자별 개인 성향과 메일 반응 행태를 반영한 스팸 메일 대응 시스템 구축 방법을 제안하였으며, 알려진 범위에서 사용자 선호도를 고려하여 스팸 문제를 해결하려는 첫 시도이다. 데이터 수집 과정과 제안된 시스템 구현의 제한점에도 불구하고, 적응적이며 쉽게 개인화할 수 있는 스팸 대응 필터의 수요는 증가하고 있다. 또한 메일 사용자별로 제공받은 사용자 선호 정보 대상으로 클러스터링을 수행하고 몇몇 사용자 그룹의 메일 반응 분포를 계산함으로써 사용자 취향에 따라 동일한 메일에 대해서도 서로 다른 반응을 가질 수 있다는 사실을 실험으로 확인하였다. 제안된 방법의 중요한 특징은 같은 메일에 대해서도 사용자들의 선호에 따라 다른 반응을 허용함으로써, 메일 내용 기반으로 모든 사용자가 같은 메일에 대해 동일하게 응답할 것으로 판정하는 현재 시스템과는 차별화된다. 향후에는 현존 내용 기반 필터들과 비교하기 좋도록 실시간으로 많은 수의 메일을 처리하도록 하고, 사용자의 주기적인 반응을 계속 관찰하여 사용자 선호 온톨로지를 학습하도록 시스템을 확장할 필요가 있다.

참 고 문 헌

[1] P. Wolfe, C. Scott, and M. Erwin, *Anti-Spam Tool Kit*, McGraw Hill, 2004.

[2] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A bayesian approach to filtering junk e-mail." *Proc. of AAAI Workshop on Learning for Text Categorization*, pp.55-62, 1998.

[3] H. Drucker, D. Wu, and V. Vapnik, "Support Vector Machines for Spam Categorization," *IEEE Trans. on Neural Networks*, Vol.10, No.5, pp.1048-1054, 1999.

[4] G. Cormack and T. Lynam, "On-line Supervised Spam Filter Evaluation," *ACM Trans. on Information Systems*, Vol.25, No.3, article 11, 2007.

[5] A. Gray and M. Haahr, "Personalized, Collaborative Spam Filtering," *Proc. of the First Conf. on Email and Anti-Spam*,

http://www.ceas.cc/papers-2004/, 2004.
[6] J. Ravi, W. Shi, and C. Xu, "Personalized Email Management at Network Edges," *IEEE Internet Computing*, Vol.9, No.2, pp.54-60, 2005.
[7] Anti-Spam Firewall, http://www.barracudanetworks.com/ns/products/anti_spam_tech.php.
[8] 김종완, 김희재, 강신재, "데이터 마이닝 기술을 적용한 사용자 선호 스팸 대응 온톨로지 구축," *한국퍼지및지능시스템학회논문지*, Vol.17, No.2, pp.160-166, 2007.
[9] R. Segal, "Combining Global and Personal Anti-Spam Filtering," *Proc. of the 4th Conf. on Email and Anti-Spam*, http://www.ceas.cc/papers-2007/, 2007.
[10] I. Witten and E. Frank, *Data Mining: practical machine learning tools and techniques 2nd ed.* Morgan Kaufmann, 2005.
[11] D. Dou, V. McDermott, and P. Qi, "Ontology translation on the semantic web," *Journal of Data Semantics*, Vol.2, pp.35-57, 2004.
[12] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, Vol.39, No.1, pp.1-38, 1977.
[13] T. Sasao, *Switching Theory for Logic Synthesis*, Kluwer Academic Publishers, 1999.
[14] A. Chan and A. Freitas, "A New Classification-Rule Pruning Procedure for an Ant Colony Algorithm," *Proc. of Artificial Evolution*, pp.25-36, 2005.

저 자 소 개

김종완(Jongwan Kim)
1987년 : 서울대학교 컴퓨터공학과(공학사)
1989년 : 서울대학교 컴퓨터공학과(공학석사)
1994년 : 서울대학교 컴퓨터공학과 졸업 (공학박사)
1995년～현재 : 대구대학교 컴퓨터·IT 공학부 교수
1999년～2000년 : 미국 U. of Massachusetts 방문교수
2006년～2007년 : 미국 U. of Oregon 방문교수

관심분야 : 인공지능, 데이터마이닝, 퍼지시스템, IT융합서비스
Phone : 053-850-6575
Fax : 053-850-6589
E-mail : jwkim@daegu.ac.kr