

전자메일 기반의 고객관계관리(CRM) 시스템 개발에 관한 연구

김 승 욱*, 양 광 민**

A Study on the Development of Electronic Mail-based Customer Relationship Management System

Seung Wook Kim* · Kwang Min Yang**

Abstract

This study designs and implements a new approach to the classification of e-mail requests from customer based on machine learning techniques. The work on building an electronic mail classifier can be cast into the framework of text classification, since an e-mail is viewed as a document, and judgement of interest is viewed as a class level given to the e-mail document. It is also implemented an e-mail based automated response system that integrate with Call Center in a practical use.

논문접수일 : 2003년 6월 12일

논문게재확정일 : 2003년 10월 11일

* 평택대학교 경상정보학부 전자상거래학과 전임강사

** 중앙대학교 경영대학 경영학과 교수

1. 서 론

전자 메일은 그 사용의 편리성과 보편성 때문에 일반인이 전화 다음으로 쉽게 이용할 수 있다는 장점과 더불어 기업의 입장에서는 고객 정보를 전자메일을 통하여 획득하고 다양한 측면에서 활용할 수 있다는 장점을 가지고 있다. 그러나 많은 국내 기업들의 경우 고객이 전자메일을 통하여 서비스 등과 관련된 문의를 하더라도 이에 대한 응답이 즉각적으로 이루어지는 경우가 매우 드물며 하루 이틀씩 응답이 지연되는 경우가 대부분이며 어떠한 경우에는 전혀 응답을 못 받았다는 경우도 많다는 결과가 있다(Frykholm, 2000). 이러한 경우 고객들은 관련기업의 서비스에 대하여 불만을 지니게 되며 어떤 경우에는 기존의 기업과 거래관계를 단절하게 되는 경우도 있어 기업의 입장에서는 장기적으로 기업의 수익성에 부정적인 영향을 미치게 된다(Jacob, 1995).

이러한 문제점을 해결하기 위해서 본 연구에서는 고객 문의의 전자메일을 실시간으로 모니터링 하면서 메일이 신규로 들어왔을 때에 기존에 학습된 분류방법에 의해서 사람 대신에 시스템이 전자메일의 내용을 분석하고 이해하여 단순한 질의메일일 경우에는 자동으로 관련내용을 회신하거나 이미 정의된 범주(category) 또는 담당 관리자에게 자동으로 고객의 전자메일을 전송하여 고객의 전자메일을 통한 문의사항 또는 요청사항에 대해서 신속하게 답변할 수 있는 고객관계관리 시스템을 개발하고자 하는 데에 본 연구의 목적이 있다.

2. 관련 문헌의 고찰

전자메일 자동분류 및 응답에 관한 연구는 기본적으로 대부분의 디지털 형태의 문서 자동분류

와 관련된 연구에 초점을 맞추고 있다. 따라서 본 연구에서는 기존의 문서 자동분류에 관한 연구에 대해서 살펴본 후에 전자메일 분류 시스템과 관련된 연구들에 대해서 살펴보고자 한다. 끝으로 기존 관련 문헌의 고찰을 통하여 본 연구의 문제점 및 시사점을 규명하고자 한다.

2.1 문서 자동분류에 관한 연구

웹 문서 등 디지털 문서의 폭발적인 증가에 따라 문서를 보다 효율적으로 문서 관리를 하기 위한 기술이 필요하며 수만 건의 문서를 사람이 수동으로 분류한다는 것은 시간적 경제적으로 불가능하다. 또한 급격히 변화하는 지식 정보 시대에 유연하게 대처하기 위해 문서 분류 체계를 수시로 조정할 수 있는 유연성이 필요하고 새롭게 창출된 지식을 습득하여 문서 분류 체계를 확장시킬 수 있는 확장성이 필요하다. 이와 같은 문제점을 해결하기 위해 최근 분류학습을 이용한 문서의 자동분류에 관한 연구가 다양하게 진행되고 있는데, 문서 자동분류란 통상 기계학습(machine learning)을 이용하여 미리 학습하여 둔 카테고리(category) 중 하나로 문서를 분류해 주는 처리를 말한다. 이러한 문서분류를 위한 방법으로는 이미 분류되어진 문서로부터 각 분류 범주에 나타나는 단어들의 출현빈도에 대한 정보를 추출하여 분류에 이용하는 통계적 분류 방법과, 문서가 가지고 있는 뜻을 파악하여 분류에 이용하는 지식기반 방법 등이 있다(Cohen and Singer, 1999).

통계적 분류 방법에는 사람에 의해 이미 분류되어 있는 훈련 셋(training set)으로부터 각 분류 카테고리에 나타나는 단어들의 출현 빈도에 대한 정보를 추출하고, 분류하고자 하는 문서로부터 주요 단어들을 추출한 후 이를 이용하여

적합한 카테고리를 찾거나 각 카테고리에 대하여 포함 여부를 판단하는 것이다(Salton, 1999). 베이지안 확률값을 이용하여 문서가 각 카테고리에 속할 확률을 계산하는 방법과, 분류하려는 문서와 각 카테고리에 포함된 문서들 간의 유사도를 계산하는 방법이 제안되었다.

지식기반 방법은 분류 대상 문서의 샘플들을 분석하여 분류 규칙들을 만들고 이러한 규칙을 이용하여 문서 분류를 수행하는 것으로, 문서의 내용에 따라 분류 규칙을 만드는 방법과, 문서 내용 외의 정보들을 이용하는 방법이 있다(Hayes and Weinstein, 1990).

2.2 메일 분류에 관한 연구

NFO Worldgroup(2002)의 조사에 의하면 미국 소비자들의 83% 이상이 최소한 하루에 한번 정도는 업무와 관련된 일로 전자메일에 접속한다고 보고하고 있으며 포레스터 리서치(2000)에 따르면 1997년의 경우 상업용 전자메일 건수가 30억 건을 초과하였으며, 2002년이면 약 2,500억 건으로 증가할 전망이다. 또한 프라이스워터하우스쿠퍼스에서 실시한 조사에서는 응답자의 83%가 인터넷을 이용하는 가장 근본적인 목적이 전자메일 때문이라고 했다. 이러한 전자메일 사용 급증은 개인과 기업의 입장에서 전자메일을 자동으로 분류하고 응답하려는 필요성이 점차로 증가하고 있다.

먼저 개인의 입장에서 볼 때 전자메일의 사용이 다른 사람과의 의사소통에 있어 중요한 의사소통의 수단이기 때문에 전자메일의 사용이 증가한 영향도 일부 있으나 정크(junk)나 스팸(spam) 메일의 증가도 한몫을 한 결과이다. 이러한 전자메일 증가는 사용자들이 이런 전자메일을 읽고 필요한 메일과 필요치 않은 스팸 메

일을 처리하는 데에 많은 시간을 필요로 하게 되었다. 1990년대 중반부터 정크 메일의 급격한 증가는 메일 사용자들에게 귀찮고 성가신 일이 되었으며 오히려 스팸 메일과 본인 업무에 필요한 메일을 구분하는데 많은 시간과 노력이 필요하게 되므로 스팸 메일은 사회적으로 시간과 비용에 문제점으로 대두되고 있다. 또한 Cranor와 Lamacchia(1998)의 스팸 비율(spam rate)라는 보고서를 보면 1997년의 초반기 AOL 시스템의 전자메일들 중에서 30%가 스팸 메일이었으며 기업 네트워크에 들어오는 전체 메일 중에 10% 정도가 스팸 메일이었고 그 중 임의로 400개의 스팸 메일을 조사한 결과 11%가 성인용 광고였다.

또한 기업의 입장에서 볼 때, 미국의 경우 55%의 고객들이 기업에 전자메일을 보낸 후 6시간 이내에 그에 대한 정확한 응답을 기대하지만, 20%의 기업들만이 고객의 기대를 충족시켜 준다고 한다. 실제로 기업들의 경우 웹사이트에 들어가 보면 제일 아래에 웹 마스터의 전자메일 주소만이 표시되어 있을 뿐, 여기에 아무리 메일을 보내도 신속한 응답을 기대하기가 어렵다.

2.3 기존 연구의 한계점

정크 메일 분류와 관련된 연구에서는 다른 여러 가지 방법 중에서 나이브 베이즈 방법에 의한 전자메일 분류가 가장 좋은 성과를 나타내었으며, 정확도 측면에서도 95% 내외의 결과를 나타내었다(Cohen, 1999). 그러나 정크 메일 분류와 관련된 연구에서 나타난 높은 정확도는 전자메일에 대해서 범주 기준 자체가 '삭제' 또는 '허용' 2가지 정도의 분류 기준만을 가지고 전자메일을 분류하기 때문에 높은 정확도가 나타났다고 판단된다. 개인 차원에서 전자메일을

구분	메일 분류 시스템	주요 내용	결과
메일 범주화 연구	i-ems(2002)	Sender, Decision Tree, Keyword, TF-IDF 4가지 다른 방법을 이용하여 전자메일 분류 정확도 측정	키워드 탐지 방법이 가장 좋은 결과
	ifile(2000)	개인 전자메일 관리를 돕기 위한 시스템으로서 나이브 베이즈 방법을 이용하여 89%의 정확도를 보임	단순 범주 적용
	Ripper(1999)	키워드탐지 방법과 TF-IDF 방법을 비교하여 실험한 결과 84-94%의 비슷한 정확도를 나타냄	키워드 탐지 방법이 더 적합하다고 강조
	MailCat(1999)	TF-IDF 방법을 이용하여 40-60%의 정확도를 나타냄	단순 범주 적용
	Re:Agent(1998)	TF-IDF 방법을 이용하여 98%의 정확도를 나타냄	단순 범주 적용
	Information Lens (1987)	기존에 미리 정해놓은 규칙에 의하여 전자 메시지를 분류하는 방식으로, 반구조화된 템플릿을 사용자에게 제공하여 입력방식을 권장함	수작업에 의하여 템플릿을 관리, 자연어 처리과정을 배제
정크 메일 연구	Keyword vs Bayes (2000)	2가지 방법을 비교하였으며, Keyword 방법은 정확도 95%, 재현율 53%를 보였으며 나이브 베이즈 방법은 정확도 98%, 재현율 78%를 나타냄	2가지 연구방법 결과 큰 차이 없음
	Naive Bayes (1999)	스팸 메일 분류에 있어 정확도는 나이브 베이즈가 95%, Ripper는 90%를 나타냄	단순 스팸 메일 분류에 한정됨
	Genetic(1999)	유전자 프로그램을 이용 베이즈 분류기와 비교, 정확도는 95%와 96%, 재현율은 70%와 77%를 나타냄	베이즈 분류기가 비교적 정확한 결과
	SpamCop(1998)	나이브 베이즈 방법을 사용하여 하나의 메일 안에서 4개 이상 같은 단어의 반복이나 특정 단어의 사용이 비슷한 비율로 나타날 경우 스팸 메일로 간주	나이브 베이즈 방법이 94%의 높은 정확도

<표 1> 전자메일 분류에 관한 연구

관리하기 위하여 개발된 정크 메일을 분류하기 위한 수준의 연구결과는 정확도 측면에서는 매우 높은 수준을 나타내었지만 기업을 대상으로 적용하기에는 한계가 있다.

또한 전자메일 범주화와 관련된 연구들에서는 전자메일이 일반적인 자동 문서분류의 문제가 아니라 고정된 특징을 가지는 문서에 대해서 분류를 수행하는 경우이기 때문에 핵심단어 탐지방법에 기반을 둔 학습방법을 사용하는 것이

더 효과적이라고 파악되었다. 그 이유는 전자 메일이 보낸 사람(from), 받는 사람(to), 제목(subject), 본문(body)와 같은 네 개의 고정된 속성을 가지는 문서이기 때문에 분류를 수행하는 다른 정보검색 기법보다 더 적합하다는 설명이며 또한 전자메일의 내용이 다른 문서들에 비해서 비교적 데이터양이 적고 규칙성을 가지고 있기 때문에 전자메일 분류를 위해서는 키워드 탐지규칙 방법이 우수하다고 할 수 있다.

그러나 메일 범주화 연구에서 나타난 문제점은 대부분의 연구에서 그다지 많지 않은 2-3개 내외의 분류 범주를 기준으로 연구를 수행하였기 때문에 현실적인 측면에서 보면 여러 가지 측면에서 한계점을 가지고 있다.

3. 시스템 설계 및 구현

3.1 아키텍처

본 연구에서 설계 및 구현한 응답관리 시스템을 우리는 전자메일 기반 고객관계관리 시스템 즉, EbCRMS(Electronic mail based Customer Relationship Management System)이라 명명하였다. EbCRMS는 메일서버, 데이터베이스 서버, 웹 서버, 애플리케이션 서버 등 다양한 서버 데몬들 과 이것들의 서비스 라이브러리를 이용하여 프로그래밍 한 세부모듈들로 매우 복잡하게 구성되어있다. 이 시스템은 크게 고객에 의해 받아들인 메일을 실시간으로 관리하는 부분과 메일이 메시지를 형태소로 분석 및 이것을 분류 및 학습하는 부분 그리고 유의수준에 의한 메일 응답 및 전송 부분 3 가지로 나눌 수 있다. 이와 관련된 시스템 아키텍처를 살펴 보면 <그림 1>과 같다.

한편, 본 시스템을 설계하기에 앞서 고려되어야 할 설계 원칙(principle of design) 또는 고려사항은 다음과 같다.

첫째, 정확성

고객으로부터의 전자메일을 신속하게 응답하기 위해서 가장 기본적인 조건은 전자메일의 내용을 정확하게 분류해내는 것이다. 본 연구에서는 언어학적인 접근방법의 가장 기본적인 방법이

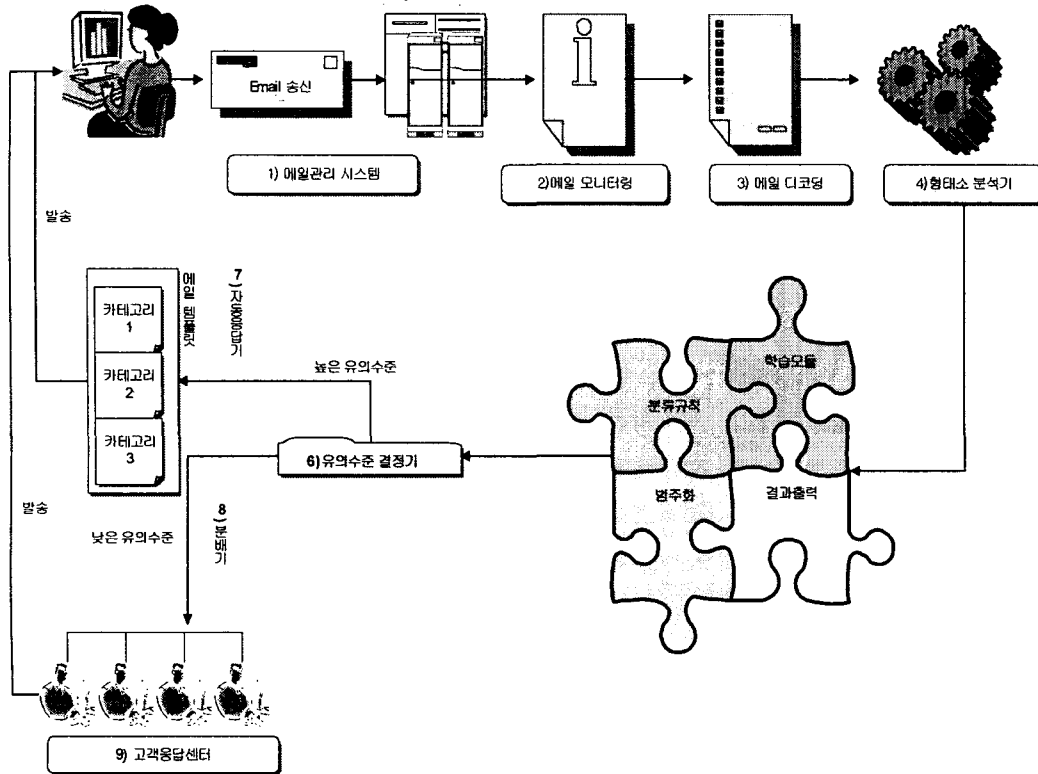
형태소 분석을 사용하고 통계적인 접근방법인 키워드탐지방법을 복합적으로 사용하고 있다.

둘째, 자동화

자동화된 전자메일 응답관리 시스템을 활용할 경우에는 기존의 수작업으로 분류되던 전자메일을 실시간으로 모니터링하고 이를 분류규칙에 의하여 분류하고 유의수준에 따라서 전자메일을 자동 응답할 수 있다면 기존의 단순하고 반복적인 콜 센터의 상담원들의 업무를 대폭적으로 경감해줄 수 있기 때문에 자동화라는 측면이 본 연구의 시스템을 설계하기 위한 주요 고려사항이다.

셋째 편이성

최근 들어 웹을 통한 고객 접촉이 많아지면서 웹을 통한 고객과의 실시간 의사소통이 중요해지고 있다. 일반적으로 콜 센터에 전화를 통한 질의는 특정 업무시간 내에서만 서비스를 받을 수 있다는 단점이 있지만 웹을 통한 질의 및 이에 대한 답변은 업무시간이외의 시간에도 서비스가 가능하다. 특히 전자메일을 통해 응답을 원하는 사람들은 언제 어디에서든지 인터넷을 이용하여 자신의 질의내용을 파악하고자 하기 때문에 웹을 기반으로 한 고객관계관리 시스템 구현을 고려하였다.

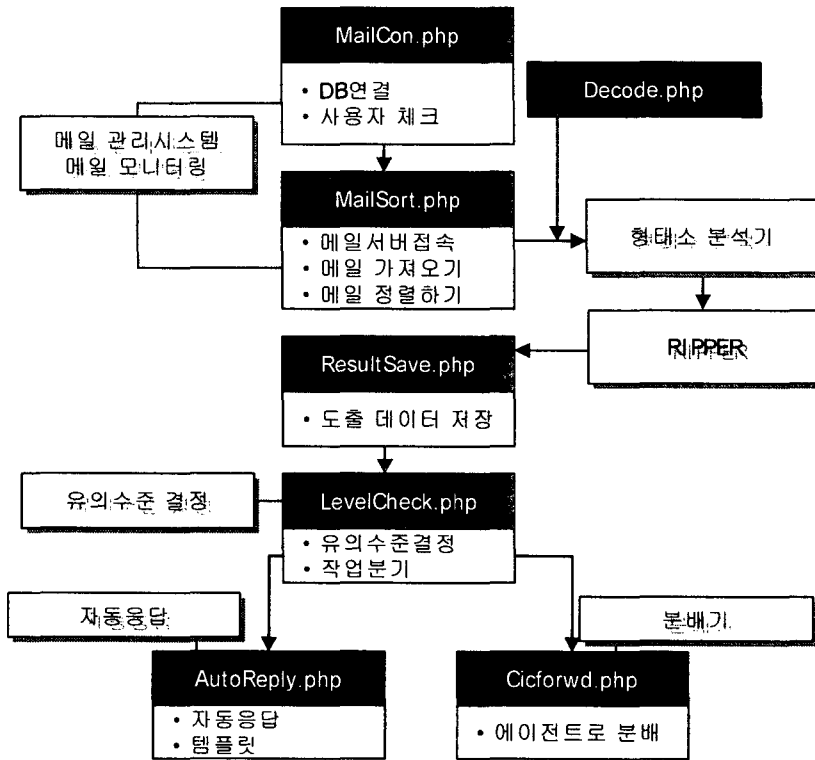


<그림 1> EbCRMS의 아키텍처

3.2 시스템 구성 및 환경

EbCRMS의 구현모듈은 DB접속부분과 사용자 접속 체크를 위한 전역 인클루드 파일로 사용될 MailCon.php와 메일서버에 접속하고 메일서버에서 메일메시지들을 가져와 정렬하는 MailSort.php, 정렬된 메일메시지를 차례로 형태소분석에 입력하기 위한 Decode.php, Ripper에 의해 도출된 분류집단 및 핵심단어를 저장하는 Resultsave.php, 유의수준을 결정하고 나온 결과에 따라 다음 프로세스로 보내는 Levelcheck.php, 자동응답처리를 해주는 AutoReply.php, Call Center에이전트에게 메일

을 전달하는 Cicforwd.php가 있다. 또한, 형태소분석기는 앞에서 언급한 국민대학교 자연어 정보검색 연구실에서 개발한 HAM을 이용했고, 학습 및 분류기는 분류학습 엔진인 Cohen의 Ripper를 한글을 수용할 수 있는 8비트 체계로 수정하고 입출력 부분을 변경한 후 컴파일 하여 다른 시스템과 연동 하였다. 이와 같은 내용은 다음의 <그림 2> 전체 모듈 구성도에서 보는 바와 같다.



<그림 2> 전체 모듈 구성도

3.3 시스템의 구현

첫째, 메일서버 접속과 사용자 체크
MailCon.php는 전체 시스템의 입구에 해당하는 부분으로 데이터베이스 접속을 위한 전역변수 선언과 함수 그리고 사용자정보의 체크를 책임지고 있는 부분이다. 또한 Mailsort.php는 설계시 구상했던 메일모니터링부분을 담당하고 있다. PHP에서 지원하는 IMAP함수를 이용하여 IMAP서버에 접속하여 최신의 날짜별로 정렬하는데 정렬할 때 제목, 메시지번호, 정렬번호, 메일날짜, 보낸 사람의 ID, 메일주소 등을 분류하여 정렬한다.

둘째, 분류결과 및 유의수준 결정
LevelCheck.php는 Ripper에 의해 도출된 분류 규칙, 분류집단 그리고 그 집단에 속하는 핵심 단어들을 mail_category에 저장하면 그것과 각각의 메일메시지에 있는 단어들을 비교하여 유의수준을 백분율로 도출한다. 또한, LevelCheck.php의 메커니즘은 현재 검사하고 있는 메일메시지의 단어와 연관된 mail_category 테이블 안의 레코드 셋을 위의 function_check_ratio함수에 담아 이를 유의수준 백분율 계수로 나타내는 것이다.

```

function check_ratio( $mail_content , $category_word )
{
    $count = 0 ;
    $sword = explode(" ", $category_word ) ;

    for ( $i = 0 ; $i < count($sword) ; $i++ )
    {
        if ( ereg( $sword[$i] , $mail_content ) == 1 ) $count ++ ;
    }
    if ( $count > 0 ) return round( $count / count($sword) , 2 ) ;
    else return $count ;
}

```

<그림 3> 유의수준 결정계수의 도출

셋째, 자동응답 및 메시지 전달

AutoReply.php는 유의수준 결정에 의해 가장 높은 유의수준 값이 70% 이상이면 해당카테고리 담당자의 메일 주소를 배열에 담고 요청자의 주소 및 내용을 변수에 저장한다. 그리고 LevelCheck.php에서 유의수준이 70% 미만인 경우에는 가장 높은 값을 가진 카테고리 담당자에게 요청 고객에게 받은 전자메일 내용을 그대로 전달하게 된다. 또한 받은 메시지 자체가 어느 카테고리에도 해당되지 않은 내용을 담고 있을 수 있기 때문에 기타내용을 담당하는 콜 센터 상담원을 두어 기타담당자가 이를 해결한다.

또한 Cicforwd.php도 AutoReply.php와 작동방식은 유사하다. 그러나 카테고리에 의한 템플릿내용이 아니라 요청자의 해당 메시지만을 카테고리 담당자에게 전달하는 것이기 때문에 더 간단한 작업이라 할 수 있겠다.

끝으로 고객으로부터 접수된 질의메일에 대해서 콜 센터의 관리자는 전자메일 관리를 위해서 전체 접수된 메일 중 자동응답을 통합 전달된 메일결과와 콜 센터 상담원에게 전달된 결과 그리고 분류되지 않은 메일의 분류결과를 확인하고 지속적으로 관리할 수 있다.

4. 효율성 분석

DeLone과 McLean(1992)은 일반적으로 정보시스템의 성공을 평가하기 위한 요소로 시스템 품질과 정보 품질을 성공요소로 보고 있으며 Pitt(1995)와 Seddon(1999)은 DeLone과 McLean의 모형에 각각 서비스의 질(service quality)과 사용자 참여(user involvement)가 첨가된 형태로 정보시스템 성과평가 모형을 확장하였다. 또한 전자메일 분류의 정확도를 평가하는 측정지표로는 Lewis(1995), Cohen(1996), Hecherman(1998)등이 이용한 2x2 분할표(contingency table)를 이용하여 측정지표를 도출하였다.

따라서 본 연구에서는 기존의 정보시스템 성과평가를 위한 측정지표들에서 성과측정 지표를 도출하고 전자메일의 분류 정확도를 평가하기 위해서는 문서분류 및 전자메일 분류의 정확성을 측정하는 지표들을 활용하여 본 시스템의 효율성을 분석하고자 한다.

첫째 정확성 지표는 EbCRMS에 대한 가장 기본적인 평가로서 고객으로부터의 전자메일 내용을 파악하여 얼마나 정확하게 이미 정의된 특정 범주에 분류시키느냐에 달려 있다. 또한 더 나아가서 특정 주제 범주에 분류된 모든 전자메일들에 대해서 시스템이 특정 주제 범주에 맞게 분류할 수 있는 능력을 필요로 하게 된다. 대부분의 문서 분류 또는 전자메일 분류와 관련된 연구들은 정확도(precision)와 재현율(recall)을 성과 측정지표로 사용하고 있다.

분류 정확도를 분석하기 위하여 국내 S생명사 콜 센터에 2002년 3월에서 5월 사이에 접수된 2893의 전자메일 가운데에서 무작위로 500개의 전자메일을 추출하였다.

먼저 콜 센터에서 전자메일 관리를 담당하고 있

는 관련 분야의 전문가들을 통해 500개의 데이터를 가장 빈번하게 문의하는 질문에 대한 유형을 분류하게 하였고 이를 토대로 하여 6개의 범주로 분류되었다. 다음에는 동일한 500개의 데이터에 대해서 EbCRMS를 사용하여 자동분류를 수행하였다. 총 경우의 수는 실험 데이터 500개에 대해서 각 실험데이터가 한번에 6개의 범주 중에서 어느 범주에 속하는지에 대한 실험을 수행하였다. 위의 기준에 따라서 분석한 결과를 분할표(contingency matrix)에 표시한 결과 <표 2>와 같이 결과가 나타났다.

<표 2>에서 보는바와 같이 EbCRMS의 전자메일 분류 정확도를 측정한 결과 83%로 나타났으며, 실제로 특정 주제 범주에 정확하게 분류된 모든 전자메일들에 대해서 시스템이 사전에 정확하게 특정 주제 범주에 맞게 분류할 수 있는 능력을 나타내는 재현율은 73%로 나타났다.

		실제 분류값(전문가)	
		아니오	예
예상 분류값(시스템)	아니오	A (2365개, 78.8%)	B (135개, 4.5%)
	예	C (135개, 4.5%)	D(365개, 12.1%)

<표 2> 분류 정확도 결과

이러한 수준은 Elisabeth, Judy와 Eric(2002)의 연구에서 핵심단어 탐지방법을 이용한 정확도가 약 72%, MailCat(1999)의 경우 50% 내외의 정확도를 나타낸 것과 비교해 보면 본 연구에서 개발한 EbCRMS가 더 높은 정확도를 나타

났다. 또한 Ripper(1996)와 Re:Agent(1998)에 비해서는 정확도가 비슷하거나 다소 떨어지는 것으로 나타났으나 이들 연구에서는 2-3개의 범주만을 기준으로 성과를 측정하여 나타난 결과이기 때문에 보다 많은 범주를 대상으로 한 본 시스템의 성과가 더 좋다고 해석할 수 있다.

둘째 신뢰성은 일정한 수준의 훈련데이터를 확보하고 실험데이터의 크기를 늘려감으로서 전자메일 분류의 정확도가 어떻게 변하는지에 대한 결과를 측정하는 지표이다.

다음의 <표 3>은 훈련데이터와 실험데이터를 100개에서 1,000개까지 증가할 때에 각각의 정확도를 측정하기 위해서 10회씩 실험한 결과를 평균한 표이다.

실험데이터 \ 훈련데이터	100(개)	200	400	600	800	1000
100(개)	82.2	81.7	80.2	80.0	79.9	79.0
200	82.2	81.8	80.2	81.0	80.2	80.8
400	84.4	82.8	82.9	82.4	82.2	81.6
600	84.4	83.2	83.5	82.4	82.2	81.6
800	84.4	83.2	83.5	82.8	82.4	82.2
1000	84.5	83.3	83.7	82.9	82.5	82.2

<표 3> 신뢰성 측정결과

앞의 <표 3>에서 나타난 분석결과는 실험데이터 100개부터 1,000개의 경우 훈련데이터 200개에서 분류 정확도가 증가하고 있으며 대략 400개 이후 1,000개까지는 정확도의 변화가 거의 없이 일정하게 나타났다. 이러한 분석 결과는 본 연구에서 개발한 EbCRMS의 학습을 위하여 필요로 하는 전자메일 훈련데이터의 크기는 200개에서 400개 사이가 적절한 것으로 나타났

다.

또한 실험 데이터의 변화에 따른 정확도는 훈련 데이터 300개를 기준으로 분석하였을 때 실험데이터 400개 미만까지는 83%이상의 정확도를 나타내다가 400개 이후에는 정확도에 변화가 없었으며 실험데이터 1,000개일 경우에도 정확도가 80% 이상인 것으로 나타났다.

셋째 처리 속도는 본 연구에서 개발한 EbCRMS 시스템이 전자메일을 처리할 때에 걸리는 CPU 처리속도를 말하고 있으며 전자메일의 수량 또는 크기가 변함에 따라서 전자메일을 분류하는데 소요되는 CPU 처리속도를 초단위로 측정하여 분석해 볼 수 있다. 처리 속도 측정지표는 EbCRMS 시스템이 처리할 수 있는 처리 용량에 대한 성과를 측정할 수 있다.

훈련 데이터 300개를 고정하고 이때, 실험데이터가 100개에서 1,000개까지 늘어날 때에 EbCRMS 시스템이 한번에 각 처리단위에 대해서 처리시간을 측정하였다. 이때 사용한 처리속도는 본 연구에서 연구의 환경으로 설정한 시스템의 CPU 처리 시간을 기준으로 측정하였으며, 약 10회에 걸쳐서 측정한 결과의 평균값에 대하여 결과분석을 시도하였다. 그 결과는 아래의 <표 4>에서 보는 바와 같다.

실험데이터 (건수)	처리시간 (초)
100	2.2
200	3.9
400	5.7
600	6.3
800	7.1
1000	9.7

<표 4> EbCRMS의 처리속도

<표 4>에서 보는 바와 같이 본 시스템은 동시에 약 400개미만의 전자메일 분류에 있어서는 약 5초미만의 시간이 소요되었으며, 1000개까지는 시간은 약 10초 이내의 처리속도가 나타났다. 이러한 분석 결과가 의미하는 바는 EbCRMS이 일반적인 노이즈가 많고 복잡한 내용이 많은 전자메일을 분류 처리하는 데에도 적절한 수준의 처리속도를 유지하고 있었다.

5. 결론

본 연구에서는 인터넷의 이용이 보편화됨에 따라 점차 늘어가고 있는 전자메일의 사용증대와 이에 따른 고객의 전자메일을 통한 질의에 대해서 보다 효과적으로 고객에게 대응할 수 있는 전자메일 기반의 고객관계관리 시스템을 개발하였다. 또한 그동안 컴퓨터 공학이나 인공지능에서 연구되어오던 기계학습을 활용한 문서분류 기술을 전자메일을 분류하는 데에 적용하였으며 특히 본 연구는 단순한 스팸메일을 차단하는 목적의 기존 연구에서 더 나아가 전자메일의 내용을 파악하여 그가 속한 내용의 범주를 결정하는 수준까지의 연구를 수행하였다.

그러나 본 연구가 가지고 있는 한계점 및 향후의 연구과제는 다음과 같다.

첫째, 국내 기업들에서 전자메일을 관리, 응답하는 대부분의 부서에서는 그다지 많지 않은 범주(category)를 가지고 관리하고 있는 것이 현실이다. 전자메일을 관리해야할 범주의 수가 증가 할 경우 본 연구에서 구축한 시스템에서 성과 평가한 결과가 유지 될 것인지에 대해서는 향후의 연구과제로 남겨있다.

둘째, 고객의 질의 및 간단한 요청사항에 대해서 기존에 웹 사이트에 있는 게시판이나 FAQ

등을 우선적으로 이용하고 FAQ 등에서 해결할 수 없는 내용들은 전자메일을 통하여 대응하게끔 하는 방안에 대한 논의가 향후 계속적으로 연구되어야 할 것이다. 이러한 방법의 이점은 급속하게 늘어나고 있는 고객의 질의 또는 서비스 요청 메일을 대응하기 위한 노력은 기업의 입장에서 볼 때 엄청난 관리비용이 필요로 하기 때문이다. FAQ 등과 연동된 고객 대응방법은 간단하고 단순한 질문내용은 FAQ에서 해결하고 보다 시간이 요구되는 질문에 대해서는 콜 센터에서 대응하는 것이 고객에게 보다 높은 수준의 응대품질이 가능하기 때문이다.

셋째, 본 연구의 EbCRMS 시스템의 성과를 분석하기 위한 하나의 측정 자료로서 EbCRMS의 처리속도를 측정하기 위하여 본 연구에서는 수작업을 통하여 CPU의 처리속도를 여러 차례의 실험을 통하여 평균 결과값을 분석 자료로서 활용하였으나 향후의 연구에서는 보다 정밀한 CPU의 연산처리 속도를 측정하기 위하여 CPU 연산처리 속도를 측정할 수 있는 소프트웨어 활용하여 보다 정확한 분석결과를 기반으로 하여 EbCRMS 시스템의 성과를 분석하여야 할 것이다.

참 고 문 헌

- [1] Androutsopoulos, I., J. Koutsias, K. Chandrinos and C. Spyropoulos, "An Experimental Comparison of Naive Bayesian and Keyword-based Anti-spam Filtering with Personal e-mail Message," *Proceeding of the 34th Annual International SCM SIGIR Conference on Research and Development in Information Retrieval*, 2000, pp. 160~167.
- [2] Apte, C., F. Damerau and S. M, Weiss, "Towards Language Independent Automated Learning of Text Categorization Models," *Proceeding of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, 1994, pp. 23~30.
- [3] Boone, G., "Concept Features in Re:Agent, an Intelligent Email Agent," *Second International Conference on Autonomous Agents*, Minneapolis, 1998, pp. 141~148.
- [4] Cohen, W. and Y. Singer, "Context-sensitive Learning Methods for Text Categorization," *ACM Transactions on Information Systems*, Vol. 17, No. 2, 1999, pp. 141~173.
- [5] _____, "Learning Rules that classify email," *AAAI Spring Symposium on Machine Learning in Information Access*, 1996, pp. 18~25.
- [6] Cooper, G. F. and E. Herskovits, "A Bayesian Method for Induction of Probabilistic Networks from Data," *Machine Learning*, Vol. 9, 1992, pp. 309~347.
- [7] Elisabeth, C., K. Judy and M. Eric, "Automatic Induction of Rules for e-mail Classification," *Proceeding of Australian Document Computing Symposium*, 2002, pp. 13~20.
- [8] Florac, W. A., R. E. Park and A. D. Carleton, "Practical Software Measurement: Measuring for Process Management and Improvement," *CMU/SEI-97-HB-003*, 1997, pp. 19~23.
- [9] Hayes, P. and S. Weinstein, "Adding

- Value to Financial News by Computer," *Proceeding of the First International Conference on Artificial Intelligence Applications on Wall Street*, 1991, pp. 2~8.
- [10] Jacobs, P., "Using Statistical Methods to Improve Knowledge Based News Categorization," *IEEE Expert*, Vol. 8, No. 2, 1993, pp. 13~23.
- [11] Joachims, T., "A Probabilistic Analysis of the Rocchio Algorithm with TF-IDF for Text Categorization," *Proceeding of ICML, 14th International Conference on Machine Learning*, 1997, pp. 143~151.
- [12] Lang, K., "Newsweeder: Learning to Filter Netnews. In Prieditis and Russell(Eds.)," *Proceedings of the 21th International Conference on Machine Learning*, 1995, pp. 331~339.
- [13] Lewis, D. D. "An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task," *SIGIR*, 1992, pp. 37~50.
- [14] Malone, T. W., K. R. Grant and F. A. Turbak, "The Information Lens: An Intelligent System for Information Sharing in Organizations," *Proceedings of the Conference on Human Factors in Computing Systems*, 1986, pp. 1~8.
- [15] NFO Worldgroup, "The Year 2001 Corporate Web and Email Usage Study," *Annual NFO Worldgdoup Research Report*, 2002, pp. 9~10.
- [16] Parasuraman, A., V. A. Zeithaml and L. L. Berry, "SERVQUAL: A Multiple- item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, Vol. 64, No. 1, 1988, pp. 12~40.
- [17] Pazzani, M., "Representation of Electronic Mail Filtering Profiles: An User Study," *Proceeding ACM Conference Intelligent User Interface*, ACM Press, 2000, pp. 1~5.
- [18] Provost, J., "Naive-bayes vs. Rule-learning in Classification of Email," *Artificial Intelligence Lab Technical Report AI-TR-99-284*, University of Texas at Austin, 1999, pp. 1~4.
- [19] Rennie, J., "ifile: An Application of Machine Learning to E-mail Filtering," *KDD-2000 Text Mining Workshop*, Boston, 2000. pp. 1~6.
- [20] Sahami, M., S. Dumais, D. Heckerman and E. Horvits, "A Bayesian Approach to Filtering Junk E-mail," *AAAI-98 Workshop on Learning for Text Categorization*, 1998, pp. 1~8.
- [21] Segal, R. and M. Kephart, "Mailcat: An Intelligent Assistant for Organizing E-mail," *Proceedings of the Third International Conference on Autonomous Agents*, Seattle, WA, 1999, pp. 276~282.
- [22] Whittaker, S. and C. L. Sidner, "Email Overload: Exploring Personal Information Management of Email," *CBI*, 1996, pp. 276~283.

■ 저자소개



김 승 욱

중앙대학교 경영학과에서 학사, 석사 그리고 경영학박사를 취득하였으며 현재 평택대학교 경상정보학부 전자상거래학과 전임강사로 재직 중이다. 이전에는 SAP Korea CRM 컨설턴트, 삼일회계법인(PWC) CRM 컨설턴트, 델로이트 & 투슈 코리아 (Deloitte & Touche Korea)에서는 비즈니스 컨설턴트 그리고 연세대학교 산업경영연구소에서 선임연구원을 역임했다.



양 광 민

서울대학교 재료공학과 학사, 서울대학교와 Minnesota에서 경영학 석사, UCLA에서 경영학박사를 취득하였으며 현재 중앙대학교 정교수로 재직하고 있다. 중앙대학교 경영대 학장과 한국경영학회 회장(2002)을 역임했다.

이 논문은 2003년 6월 12일에 접수하여 2차 수정을 거쳐 2003년 10월 11일에 게재확정 되었습니다.