한국산학기술학회논문지
Vol. 11, No. 7 pp. 2595-2602, 2010

문자메시지의 특성을 고려한 한국어 모바일 스팸필터링 시스템

손대능[1], 이정태[2], 이승욱[2], 신중휘[3], 임해창[4*]
[1]이노에이스, [2]고려대학교 컴퓨터 · 전파통신공학과, [3]팬택, [4]고려대학교 컴퓨터 · 통신공학부

Korean Mobile Spam Filtering System Considering Characteristics of Text Messages

Dae-Neung Sohn[1], Jung-Tae Lee[2], Seung-Wook Lee[2], Joong-Hwi Shin[3] and Hae-Chang Rim[4*]
[1]**Innoace Co., Ltd.**
[2]**Department of Computer and Radio Communications Engineering, Korea University**
[3]**Pantech Co., Ltd.**
[4]**Division of Computer and Communications Engineering, Korea University**

요 약 본 논문에서는 휴대전화로 오는 짧은 문자메시지의 스타일을 반영하여 스팸 문자메시지를 검출해내는 한국어 모바일 스팸필터링 시스템을 소개한다. 제안하는 시스템은 내용어 어휘들의 출현에만 기반을 두는 기존 방법과 달리 제안하는 스타일 정보를 추가적으로 활용하여 스팸성 단어가 포함된 일반 문자메시지가 스팸으로 잘못 분류되는 치명적인 오류를 효과적으로 줄인다. 또한 띄어쓰기 및 철자 오류교정을 거쳐 문자메시지를 정규화 함으로써 스팸 분류성능을 향상시킨다. 실제 한국어 문자메시지를 이용한 실험 결과를 통해 제안하는 시스템이 한국어 스팸 문자메시지 검출에 효과적임을 보인다.

Abstract This paper introduces a mobile spam filtering system that considers the style of short text messages sent to mobile phones for detecting spam. The proposed system not only relies on the occurrence of content words as previously suggested but additionally leverages the style information to reduce critical cases in which legitimate messages containing spam words are mis-classified as spam. Moreover, the accuracy of spam classification is improved by normalizing the messages through the correction of word spacing and spelling errors. Experiment results using real world Korean text messages show that the proposed system is effective for Korean mobile spam filtering.

Key Words : Mobile spam filtering, Stylistic information, Text normalization

1. 서론

모바일 스팸이란 이윤을 목적으로 불특정 다수의 휴대전화 사용자에게 상품이나 서비스 등을 광고하는 문자메시지를 말한다. 2000년대 이후 급속한 이동통신기기의 대중화와 더불어 모바일 스팸의 양도 폭발적으로 증가하는 추세이다. 그 내용은 성인광고, 대출광고, 게임광고 등이 주를 이루며, 수신자로 하여금 불쾌감을 유발하고 불편을 가중시키고, 이로 인해 모바일 스팸이 사회적 이슈가 되고 있다. 방지책으로써, 스팸 문자메시지 발신 번호 차단, 문자메시지 일일 발송량 제한, 광고 목적 전화번호 차단 등 다양한 방지책이 시행되고 있다. 하지만 이러한 노력에도 불구하고 2008년 하반기 기준으로 국내에서만 하루 최소 2,000만 건 이상의 스팸 문자메시지가 발송되

본 연구는 교육과학기술부 한국연구재단(KRF-2007-361-AL0013) 및 2단계 BK21사업의 지원을 받아 수행되었음.
*교신저자 : 임해창(rim@nlp.korea.ac.kr)
접수일 10년 05월 12일 수정일 10년 06월 21일 게재확정일 10년 07월 06일

고 있으며, 1인당 1일 스팸 수신량이 0.46통에 이른다는 통계 조사가 발표되기도 했다[1].

최근 이러한 모바일 스팸을 자동으로 구별하기 위해 문자메시지의 내용을 이용한 스팸 필터링 기법들이 주목받고 있다. 스팸 필터링의 주된 대상이 되었던 전자 우편(e-mail)과는 달리, 문자메시지는 한 번에 전송 가능한 글자 수의 제한과 단말기의 하드웨어적 제약으로 인해 그 내용이 짧다는 특징이 있다. 이는 스팸 분류 시 활용 가능한 정보량의 부족으로 이어져 결과적으로 분류 작업을 더 어렵게 만드는 원인이 된다[2]. 이러한 문제를 극복하기 위해 기존의 내용 기반 모바일 스팸 필터링 관련 연구들은 문자메시지의 짧은 내용을 어떻게 표현하느냐에 주목하였다. 다시 말해 짧은 문자메시지의 내용에서 광고성 어휘나 구를 정확히 탐지하는데 주력했다고 볼 수 있다. 이에 해당되는 한국어의 어휘들은 “대출”, “이자”, “대리”, “운전” 등이 있다. 하지만 표 1에 나타난 예제에서와 같이, 비스팸 메시지에서도 이러한 단어들은 사용될 수 있다. 이로 인해 비스팸 메시지를 스팸으로 분류하는 심각한 분류 오류를 유발할 수 있다.

[표 1] 스팸 메시지와 비스팸 메시지 예제

비스팸 메시지	스팸 메시지
철수야오늘술마셨으니 **대리 운전** 부르셩	바로 콜 **대리 운전** 정성껏 모시겠슴다!!
돈없엉.. 대출 받아야되..	☎최저 대출 이자! 현금 필요시 지금전화 콜☎

또한, 모바일 기기의 특성 상 축약어나 철자 오류 및 띄어쓰기 오류 등이 문자메시지 내에 빈번하게 발생할 수 있다. 이러한 문자메시지의 특성은 내용에 기반을 두는 스팸 필터링 시스템의 성능 저하시키는 주요 요인으로 꼽힐 수 있다.

이에 착안하여, 본 연구는 모바일 스팸 필터링 시 스타일 정보[5]의 사용과 메시지 오류 교정 과정을 포함하는 시스템을 제안한다. 여기서 스타일 정보란 문자메시지의 내용뿐만 아니라, 그 내용이 어떤 식으로 쓰여 있는지를 나타내는 정보를 말한다. 제안하는 방법의 평가를 위해 대량의 실제 한국어 문자메시지 말뭉치를 이용하여 기존 내용 기반 모바일 스팸 필터링 기법과의 비교실험을 수행하였다. 실험을 통해서 제안하는 시스템이 기존의 내용 기반 스팸 필터보다 성능이 뛰어남을 보였다.

본 논문의 구성은 다음과 같다. 2장에선 모바일 스팸 필터링과 관련된 연구와 스타일 정보가 주로 사용되는 저자분류 연구에 대해 다룬다. 3장에서는 본 연구에서 제안하는 시스템인 문자메시지 교정 과정과 스타일 정보를 이용한 내용 기반 한국어 모마일 스팸 필터에 대해 살펴본다. 4장에서는 실험 데이터와 실험 환경, 실험 결과 및 분석에 대해 기술한다. 마지막으로 5장에서 결론 및 향후 연구에 대해 기술한다.

2. 관련 연구

2.1 모바일 스팸 필터링

2000년대 들어 스팸 문자메시지 문제가 이슈화 되면서부터 모바일 스팸 필터링에 관한 연구가 진행되었다. [2]에서는 전자우편 스팸 필터링 분야에서 널리 사용되었던 기계학습 기법을 이용하여 내용 기반 모바일 스팸 필터링을 시도하였다. 이 연구는 문자메시지의 짧은 내용을 효과적으로 표현하기 위해 어휘 자질 추출 단위로 단어뿐만 아니라 문자-bigram, 문자-trigram, 단어-bigram을 함께 사용하였다. 그러나 이러한 표현 방법이 실제 모바일 스팸 필터링에 효과적인지를 검증하는 실험을 수행하지 않았다. 후속 연구로써 기계학습기법을 이용한 모바일 스팸 필터 구축 시 자질 표현 방식에 따른 성능 변화를 보여준 사례가 있다[3,4]. 이 연구들에선 단어로만 이루어진 어휘 자질집합과 단어, 문자-bigram, 문자-trigram, 직교 희소 단어-bigram으로 구성된 확장된 어휘 자질 집합을 비교 실험하여 그 효용성을 보였다. 확장된 어휘 자질 집합의 목적은 문자메시지의 내용에서 광고성 자질 또는 스팸과 비스팸 분류에 효과적인 어휘나 구(phrase)를 더 많이 추출해 내기 위함이다. 그러나 어휘관련 자질만으로는 짧은 문자메시지를 표현하는데 역부족일 수 있으며, 분류 작업에 있어서 중요한 역할을 하는 있는 스타일 정보를 표현하기 어렵다. 더불어, 이들 연구에선 문자메시지에 자주 나타나는 철자오류, 두음문자, 축약어와 같은 어휘 변형을 고려하지 못한다는 한계점이 있다.

2.2 저자분류

스타일 자질을 주로 사용하는 분야는 저자분류(Authorship Classification) 연구를 들 수 있다[5]. 저자분류란 주어진 텍스트나 문서의 저자를 찾아내는 작업을 의미한다. 어떤 내용을 표현하는데 있어서 저자마다 사용하는 언어표현양식 혹은 문체가 다르다는 것이 이 연구 분야의 주요한 가정이다. 이를 바탕으로 문체를 고려할 수 있는 자질들이 무엇이며, 어떻게 추출해 낼 것인지가 저자분류 연구의 핵심 고려사항이 된다. 그 추출방식에 따라 크게 표층 언어학적 분석방법(Shallow Linguistic

Analysis)과 구조 언어학적 분석방법(Structural Linguistic Analysis) 두 부류로 나눌 수 있다. 표층 언어학적 분석에 의해 추출되는 주요 자질들로는 단어의 평균 길이[6], 문장길이[7,8], 기능어의 빈도[9], 품사-trigram[5] 등이 있다. 구조 언어학적 분석방법에서 사용되는 주요 자질들은 구문구조분석 결과로부터 추출된 문맥자유문법 생성패턴[10] (CFG productions), 문장 내 구문 종류 별 빈도[10] (예: 명사구, 동사구 등의 빈도), 구문구조트리의 깊이[11] 등이 있다. 기존 연구에 따르면 이러한 자질들은 저자의 개성이나 특징을 표현할 수 있다[5].

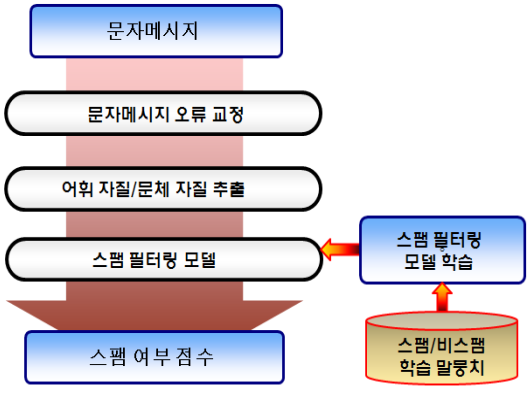
모바일 스팸 필터링을 위해서 이러한 연구를 활용할 수 있다. 본 연구에서는 스팸 문자메시지를 보낸 사람과 일반 메시지를 보낸 사람의 문체는 다르다고 가정하여 저자분류에 사용되었던 자질들을 모바일 스팸 필터링 시스템에서도 활용한다.

3. 제안하는 방법

본 논문에서는 문자메시지 오류 교정 과정과 스타일 자질을 이용한 내용 기반 한국어 모바일 스팸 필터링 프레임워크를 제안한다. 문자메시지 오류 교정은 스팸 필터링에 필요한 자질 추출 시 영향을 줄 수 있는 철자 오류나 띄어쓰기 오류 교정을 뜻한다. 3.1장에서 제안하는 스팸 필터링 시스템 구조대해 기술하고, 3.2장에서 스타일 자질에 대한 설명을 기술한다.

3.1 시스템 구조

본 논문에서 제안하는 스팸 필터링 시스템의 전반적인 구조는 그림 1과 같다.



[그림 1] 제안하는 시스템의 구조

입력은 문자메시지이고 출력은 해당 문자메시지의 스팸 여부이다. 입력된 문자메시지는 철자오류 및 띄어쓰기 오류 교정 과정을 거친다. 이러한 오류 교정에 관련된 다양한 기법들 중에서도 문자메시지에 대해 높은 정확률을 보이는 [12]를 활용하였다. 교정을 거친 메시지의 텍스트로부터 [3,4]에서 제안한 어휘 자질과 함께, 추가적으로 본 논문에서 제안하는 스타일 자질을 추출한다. 통계 기반 스팸 필터링 모델은 추출된 자질을 이용해 메시지가 스팸일 가능성을 계산한다. 통계 기반 학습 모델로는 최대엔트로피 모델[13]을 사용하였다. 최대엔트로피 모델은 스팸성 여부를 나타내는 확률을 추정할 시에 서로 다른 성격의 자질을 통합하여 사용할 수 있다는 장점이 있다. 전자우편 스팸 필터링 시 베이지안 분류기보다 최대 엔트로피 모델이 좋은 성능을 보여주었던 연구가 있으며[14], 유사 태스크인 문서 분류 작업 시에도 비교적 안정적인 성능을 보인다고 알려져 있다[15].

3.2 스타일 자질

스타일 자질의 사용 목적은 스팸과 비스팸간의 문체의 차이를 고려하기 위함이다. 본 연구에서는 스타일 자질 사용을 위해 다음과 같이 가정한다. 하나, 문자메시지의 발신자로는 스팸 발신자와 비스팸 발신자가 존재한다. 둘, 스팸 발신자는 비스팸 발신자와는 구분되는 언어표현 양식을 일관되게 사용하는 경향이 있다. 셋, 각 문자메시지의 내용은 그 메시지를 작성한 발신자의 특성을 담고 있다.

스팸 발신자들의 목적은 일반적으로 상품, 서비스 광고 등이기 때문에 이들이 사용하는 문체는 일상생활에서 비스팸 발신자들이 사용하는 문체와는 상이할 것이라고 예상할 수 있다. 더군다나, 스팸 발신자들은 스팸 메시지를 불특정 다수에게 보내기 때문에, 각각의 메시지마다 내용이나 문체에 변화를 주기는 힘들 것으로 사료된다. 이러한 사항을 비추어 볼 때 위의 가정은 설득력이 있다고 볼 수 있다. 본 연구는 이 가정을 바탕으로 표층적 스타일 자질과 구조적 스타일 자질을 활용한다.

3.2.1 표층적 스타일 자질

본 논문에서는 표층 언어학적 분석(Shallow linguistic analysis)을 통해 자동으로 추출 가능한 자질의 집합을 편의상 Shallow라고 명명하며, 구성은 표 2와 같다.

[표 2] 표층적 스타일 자질 (Shallow)

자질 이름	자질 구성
길이자질 (LEN)	문자메시지 전체길이
	단어의 평균길이
기능어자질 (FW)	기능어별 빈도수
품사-trigram 자질 (POS)	문자메시지에서 추출된 품사-trigram
특수기호자질 (SC)	메시지 내 특수기호 비율
	이모티콘 빈도 (예: “^^”, “ㅠㅠ”)
	광고성 기호 빈도 (예: “$”, “☎”)

길이자질의 경우 스팸 발신자와 비스팸 발신자가 작성하는 메시지의 내용이 다르기 때문에, 그 내용을 구성하는 어휘의 길이분포도 차이를 보일 것이라는 가정에 근거하고 있다. 기능어자질의 사용 이유는 기능어의 문법적인 역할과 잦은 빈도로 인해 저자의 특징이 반영되어 있을 가능성이 높기 때문이다. 품사-trigram은 문장의 통사구조를 표층적 수준에서 고려할 수 있고[5], 추출이 용이하기 때문에 사용하였다. 특수기호자질은 특수기호가 문자메시지에 빈번하게 사용되며, 그 용도가 기호 별로 각기 다르다는 것에 착안하여 고안되었다. 비스팸 발신자의 경우 일상 대화에서 그들의 느낌이나 감정 표현을 위해 이모티콘을 자주 사용하는 반면, 스팸 발신자의 경우 상품이나 서비스 강조를 위해 "$" 기호나 “☎” 와 같은 기호를 많이 사용하는 경향을 보인다.

자질추출 시 표 2의 모든 자질은 해당 문자메시지의 길이로 정규화 되어 0~1사이의 소수 값으로 추출된다. 길이자질은 문자메시지의 최대 길이인 160바이트로 정규화하였다. 단어의 평균길이 계산 시 특수기호는 제외하였다. 품사정보와 기능어추출을 위해 형태소/품사 부착기[16]를 사용하였다.

3.2.2 구조적 스타일 자질

스타일 자질 중에는 구문구조 분석을 통해 얻어진 결과를 이용한 것들 또한 존재 한다. 본 논문에서는 그러한 스타일 자질집합을 편의상 Structural 이라고 칭하며, 해당 자질들은 표 3과 같다.

[표 3] 구조적 스타일 자질 (Structural)

자질 이름	자질 구성
구문빈도자질 (PhraseC)	추출된 명사구, 동사구, 형용사구 등의 상대빈도
문맥자유문법 생성패턴 자질 (SyntacticP)	구문구조 생성 패턴 (예: NP->AP NNG)
파싱트리 깊이자질 (Depth)	메시지 내 문장의 구문구조 파싱트리의 깊이

구문빈도자질과 문맥자유문법 생성패턴 자질은 저자가 사용하는 문체의 통사적 구조를 반영하는데 효과적이라고 알려져 있는 자질들이다[10]. 파싱트리 깊이자질의 경우 스팸 발신자들은 상품이나 서비스 설명을 분명하고 자세하게 하기 위해, 비스팸 발신자들에 비해 길고 잘 쓰인 문장을 구사할 것이라 가정하고 사용하였다.

구조적 스타일 자질 추출을 위해 구문구조 분석기[17]를 사용하였다. 구문구조생성패턴 자질 값은 해당 문자메시지의 길이로 정규화 했으며, 파싱트리 깊이자질의 경우 학습 데이터 내 문자메시지에서 나타난 최댓값으로 정규화 하였다.

4. 실험

4.1 실험 데이터

본 연구에서는 실제 문자메시지를 수집하여 실험에 사용하였다. 스팸과 비스팸 구분은 1장의 모바일스팸의 정의에 따라 수작업으로 이루어 졌다. 실험을 위해 총 20,000개의 문자메시지를 이용하여 이중 18,000개를 학습 집합으로, 2,000개를 실험 집합으로 사용하였다. 학습 집합과 실험 집합에서의 비스팸과 스팸의 구성 비율은 9:1로 이루어져 있다. 본 연구가 제안하는 시스템에서 문자메시지 철자 오류나 띄어쓰기 오류 교정 과정이 전체적인 성능에 어느 정도 영향을 주는 지를 파악하기 위해 20,000개의 메시지에 대한 오류 교정 전의 데이터와 오류 교정 후의 데이터, 두 집합으로 구분하여 실험에 사용하였다.

4.2 자질 구성

본 연구가 제안하는 시스템의 특성인 스타일 자질의 사용의 효용성 검증을 위한 비교 실험을 위해 아래와 같은 자질 집합을 사용한다.

Baseline: 기존의 내용 기반 스팸 필터링 연구[3,4] 방법에서 사용한 자질 구성이다. 단어(형태소 단위), 음절-bigram, 음절-trigram, 직교 희소 단어-bigram (Orthogonal Sparse Bigram)을 포함하고 있으며 어휘 자질로만 이루어져 있다.

Shallow: 3.2.1장에서 기술한 표층적 스타일 자질로 구성된 집합이며 길이자질, 기능어자질, 품사-trigram 자질, 특수기호자질을 포함한다.

Structural: 이 집합은 3.2.2장에서 기술한 구조적 스타일 자질을 나타내며, 구문빈도자질, 문맥자유문법 생성패턴 자질, 파싱트리 깊이자질을 포함한다.

Allstyle: 표층적 스타일 자질과 구조적 스타일 자질을 모두 포함하고 있는 집합이다.

Proposed: 내용기반 스팸 필터링 연구에서 사용되었던 어휘 자질과 문체 자질 전체를 포함하고 있는 집합이다. 이 집합의 의도는 제안하는 스타일 자질이 내용 기반 한국어 모바일 스팸 필터링 성능 향상에 어느 정도 기여하는지를 알아보기 위함이다.

각각의 자질 구성에 따라 실험데이터에서 추출되는 자질 수가 많고, 그 중에는 도움이 되지 않는 자질도 존재한다. 따라서 스팸 필터링 모델 학습과 실험에는 추출 된 자질 중 정보이득(Information gain) 값이 높은 상위 200개 자질 만을 사용한다. 정보이득에 의한 자질 선택은 자질 수를 효과적으로 감소시켜주면서도 분류 정확도는 떨어뜨리지 않고, 오히려 성능 향상에 기여하는 것으로 알려져 있다[18].

4.3 실험 평가 방법

모바일 스팸 필터링 작업의 특성으로 인해 스팸 필터가 얼마나 많은 오류를 내는지를 나타내는 false-positive (비스팸을 스팸으로 판단하는 경우)와 false-negative(스팸을 비스팸으로 분류하는 경우)가 성능의 중요한 척도로 알려져 있다. 또한 스팸 분류 시 임계값에 따라 이 두 개의 값은 서로 상충관계(trade-off)에 있으며 이를 반영하기 위한 평가 방법이 필요하다[19]. 이를 위해 본 연구는 전자우편 스팸 필터링 분야의 연구와 TREC 스팸필터링 대회 등에서 공식 성능평가 척도로 사용되고 있는 1-AUC(%) [19]를 사용한다. 1-AUC(%) 평가 방법은 가능한 모든 분류 임계값에 대해 false-positive와 false-negative를 동시해 고려해 하나의 실수 값을 내어준다. 이 값은 스팸 필터의 오류율을 측정한 것이기 때문에 낮을수록 스팸 필터의 성능이 좋다고 볼 수 있다. 모든 실험은 결과의 우연성을 최대한 배제하기 위해 10-fold 교차 검증과 결과 값 간의 통계적으로 유의미함을 보기 위해 t-test 검증을 수행하였다.

4.4 실험결과 및 분석

표 4는 각각의 자질 구성에 대해 문자메시지 오류 교정 과정을 포함한 경우(교정 후)와 포함하지 않은 경우(교정 전)로 나누어 실험을 진행한 결과를 보여준다. 낮은 값일수록 정확한 분류를 의미하고, † 표시는 Baseline 대비 성능변화 수치가 통계적으로 유의미함(신뢰도 레벨 $p < 0.01$)을 나타낸다.

[표 4] 자질 구성 별 성능

자질 구성	교정 전	교정 후
Baseline	13.5674	9.9651
Shallow	21.6026†	13.0012†
Structural	24.3581†	19.1899†
Allstyle	**10.8283**†	**7.9212**†
Proposed	**3.9601**†	**2.7444**†

어휘 자질만을 사용한 Baseline에 대비하여 어휘 자질과 스타일 자질을 동시에 사용하는 Proposed의 1-AUC(%) 값이 월등히 낮아지는 것을 볼 수 있다. 이 결과는 스타일 자질이 내용 기반 모바일 스팸 필터링 시 도움을 주는 것을 보여 준다. 더불어, 어휘 자질을 전혀 사용하지 않고 오로지 스타일 자질만을 사용한 AllStyle의 1-AUC(%) 값이 Baseline의 1-AUC(%) 값 보다 더 낮음을 볼 수 있다. 이는 일반 사용자와 스패머의 문체에 차이가 있음을 간접적으로 볼 수 있는 결과이다.

표 5는 각각의 스타일 자질의 성능 기여도를 알아보기 위해 제안하는 방법인 어휘 자질과 모든 스타일 자질을 함께 사용하는 Proposed에서 한 번에 하나의 문체 자질을 제외해서 얻은 실험 결과이다. LEN, SC, SyntacticP를 제거 했을 시 1-AUC(%) 값이 상승하는 것을 확인할 수 있다. 이는 LEN, SC, SyntacticP가 스팸 필터링 시 오류 감소에 큰 기여를 하고 있는 것으로 해석할 수 있다.

[표 5] Proposed에서 각 스타일자질 제거 시 성능

자질 구성	교정 전	교정 후
Proposed	3.9601	2.7444
- LEN	**5.7284**†	**3.3804**†
- FW	4.3443†	3.2424†
- POS	3.9919	2.9347
- SC	**4.3854**†	**3.6901**†
- PhraseC	4.0030	3.0824†
- SyntacticP	**5.4808**†	**4.2381**†
- Depth	4.0854†	2.9839†

[표 6] Baseline에서 각 스타일자질 추가 시 성능

자질 구성	교정 전	교정 후
Baseline	13.5674	9.9651
+ LEN	**11.7354**†	6.7961†
+ FW	12.3583†	7.2147†
+ POS	13.4419	9.8888
+ SC	**11.7346**†	**6.4614**†
+ PhraseC	11.9229†	9.0454†
+ SyntacticP	**6.1291**†	**5.0050**†
+ Depth	11.2543†	9.9241

자질간의 의존 관계 및 서로 영향을 줄 수 있음을 고려하여 Baseline에 한 번에 하나씩의 스타일 자질을 추가한 실험도 수행하였다. 표 6은 이 실험에 대한 1-AUC(%) 결과를 보여주고 있다. 표 5의 결과와 마찬가지로 LEN, SC, SyntaticP를 추가 했을 시 1-AUC(%) 값 하락 폭이 제일 크기 때문에 이들 스타일 자질이 성능 기여도가 가장 큰 것으로 결론지을 수 있다. 이러한 결과를 바탕으로 스팸과 비스팸간의 문체는 문장 길이나 단어의 길이, 특수 기호 사용 여부와 큰 관련이 있다고 볼 수 있다. 또한, 스팸과 비스팸은 구문구조 상에서도 차이를 보인다고 해석이 가능하다. 반면, 표 5, 6의 결과에서 품사 trigram을 지칭하는 POS의 성능 기여도가 가장 적은 것을 확인할 수 있다. 이는 어휘와 품사간의 강한 상호 연관성 때문인 것으로 추정된다.

표 4의 결과에서 “교정 전”의 결과가 “교정 후”의 결과보다 1-AUC(%) 값과 비교했을 때 필터링 시 더 많은 오류를 포함하고 있는 것을 볼 수 있다. 이러한 결과는 표 5와 표 6에서도 공통적으로 관찰할 수 있으며, 이는 문자메시지의 오류 교정 과정이 스팸 필터링 시 성능에 도움을 준다는 것을 보여준다. 문자메시지 오류 교정 과정이 어휘 자질 추출 시 철자 오류로 인한 미등록어(Unseen data) 비율을 줄이는데 도움을 줄 수 있고, 스타일 자질 추출 시 사용되는 태거와 파서 등 자연어처리 툴의 분석 성공률을 높이는데 좋은 영향을 미쳤다고 미루어 짐작 할 수 있다. 실제로 교정과정을 거친 “교정 후”의 실험 말뭉치에서 추출된 자질의 사전 크기(Vocabulary Size)가 “교정 전” 의 실험 말뭉치에서 추출된 자질의 사전 크기보다 12% 가량 줄었다.

5. 결론

본 논문에서는 한국어 문자메시지 스팸 필터링 시 문자메시지의 짧은 길이를 고려하여 어휘 자질만을 사용했을 경우 나타날 수 있는 오류를 줄이기 위해 스타일 자질을 사용하고, 문자메시지의 철자오류나 띄어쓰기 오류가 성능에 미치는 영향을 고려하여 메시지 교정 과정을 포함하는 시스템을 제안하였다. 실제 한국어 문자메시지를 이용한 실험을 통해 문자메시지 교정 과정이 자질 추출 과정의 신뢰도를 높이고 스팸 필터의 전반적인 성능향상에 기여함을 보였다. 또한, 본 연구는 형태소/품사 부착기, 구문구조 분석기 등을 이용한 스타일 자질 추출 및 사용을 통해 모바일 스팸 분류 시 사용자에게 치명적일 수 있는 오류를 줄였다는 점에서 기여점이 있다고 사료된다.

더불어, 본 논문에서 제안하는 시스템은 최근 사회적으로 문제가 되고 있는 블로그 댓글 스팸, 포털 사이트 내 뉴스 기사의 댓글 스팸 필터링 방법으로도 사용이 가능하다. 향후 연구로는 스팸 발신자가 지인인 것처럼 문자메시지 수신자를 속이려는 의도로 보내는 피싱(Phishing) 형태의 스팸 메시지를 어떤 식으로 판별해야 할 것인가를 들 수 있다. 이러한 피싱 메시지는 그 내용과 문체가 일련의 비스팸 메시지와도 매우 유사하여 자동으로 분류해 내기 매우 어려운 측면이 있다. 이를 막기 위한 방법으로는 매우 정교하게 고안된 규칙을 사용하는 것을 들 수 있으며, 자동으로 구축된 통계 기반 스팸 분류 모델과 통합하는 방법을 고안하고자 한다.

참고문헌

[1] 정보통신부 뉴스, “이메일 스팸 계속 감소 추세”, 7월, 2007.

[2] J. M. Gómez et al., "Content Based SMS Spam Filtering", Proc. of the 2006 ACM Symposium on Document Engineering, pp. 107-114, 2006.

[3] G. V. Cormack et al., "Spam filtering for short messages", Proc. of ACM Sixteenth Conference on Information and Knowledge Management, pp. 313-320, 2007.

[4] G. V. Cormack et al., "Feature engineering for mobile (SMS) spam filtering", Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 871-872, 2007.

[5] M. Koppel et al., "Automatically categorizing written texts by author gender", Literary and Linguistic Computing, Vol. 17, No. 4, pp. 401-412, 2002.

[6] T. C. Mendenhall, "The characteristic curves of composition," Science, pp. 237-246, 1887.

[7] G. U. Yule, "On sentence-length as a statistical characteristic of style in prose: with application to two cases of disputed authorship", Biometrika, Vol. 30, No. 3-4, pp. 363-390, 1939.

[8] A. Q. Morton, "The authorship of greek prose", Journal of the Royal Statistical Society Series A (General), pp. 169-233, 1965.

[9] F. Mosteller et al., Applied Bayesian and classical inference: the case of the Federalist papers, Springer Verlag, 1984.

[10] E. Stamatatos et al., "Automatic text categorization in terms of genre and author", Computational

Linguistics, Vol. 26, No. 4, pp. 471-495, 2000.
[11] O. Uzuner et al., "A comparative study of language models for book and author recognition", Proc. of 2nd International Joint Conference on Natural Language Processing, pp. 969-980, 2005.
[12] J.-H. Byun et al., “Three-Phase Text Error Correction Model for Korean SMS Messages”, IEICE Transactions on Information and Systems, Vol. E92-D, No. 5, pp. 1213-1217, 2009.
[13] A. L. Berger et al., "A maximum entropy approach to natural language processing", Computational Linguistics, Vol. 22, No. 1, pp. 39-71, 1996.
[14] L. Zhang et al., “Filtering junk mail with a maximum entropy model”, Proc. of 20th International Conference on Computer Processing of Oriental Languages, pp. 446-453, 2003.
[15] K. Nigam et al., "Using maximum entropy for text classification", Proc. of the IJCAI-99 Workshop on Machine Learning for Information Filtering, pp. 61-67, 1999.
[16] 이상주 외, "품사태깅을 위한 어휘문맥 의존규칙의 말뭉치기반 중의성주도 학습", 한국정보과학회 논문지(B), 제 26권, 제 1호, pp. 178-189, 1999.
[17] 박소영 외, "문장성분의 다양한 자질을 이용한 한국어 구문분석 모델", 한국정보처리학회 논문지(B), 제 11권, 제 6호, pp. 743-748, 2004.
[18] Y. Yang et al., "A comparative study on feature selection in text categorization", Proc. of 14th International Conference on Machine Learning, pp. 412-420, 1997.
[19] G. V. Cormack et al., "TREC 2005 spam track overview", Proc. of 2005 Text REtrieval Conference, 2005.

손 대 능(Dae-Neung Sohn) **[정회원]**

- 2010년 2월 : 고려대학교 대학원 컴퓨터·전파통신공학과 (컴퓨터 학석사)
- 2010년 3월 ~ 현재 : 이노에이스 사원

<관심분야>
정보검색, 문서분류

이 정 태(Jung-Tae Lee) **[정회원]**

- 2008년 2월 : 고려대학교 대학원 컴퓨터·전파통신공학과 (컴퓨터 학석사)
- 2008년 3월 ~ 현재 : 고려대학교 대학원 컴퓨터·전파통신공학과 박사과정

<관심분야>
정보검색, 자연어처리

이 승 욱(Seung-Wook Lee) **[정회원]**

- 2008년 2월 : 고려대학교 대학원 컴퓨터·전파통신공학과 (컴퓨터 학석사)
- 2008년 3월 ~ 현재 : 고려대학교 대학원 컴퓨터·전파통신공학과 박사과정

<관심분야>
정보검색, 자연어처리

신 중 휘(Joong-Hwi Shin) **[정회원]**

- 2009년 2월 : 고려대학교 대학원 컴퓨터·전파통신공학과 (컴퓨터 학석사)
- 2009년 3월 ~ 현재 : 팬택 중앙연구소 연구원

<관심분야>
자연어처리, 언어모형

임 해 창(Hae-Chang Rim) **[정회원]**

- 1983년 12월 : University of Missouri-Columbia 대학원 전산학과 (전산학석사)
- 1990년 12월 : University of Texas-Austin 대학원 전산학과 (전산학박사)
- 1991년 9월 ~ 현재 : 고려대학교 컴퓨터학과(현 컴퓨터·통신공학부) 교수

<관심분야>
자연어처리, 정보검색