

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2011-0061951 (43) 공개일자 2011년06월10일
(51) Int. Cl. <i>H04W 4/14</i> (2009.01) <i>H04W 48/02</i> (2009.01) (21) 출원번호 10-2009-0118504 (22) 출원일자 2009년12월02일 심사청구일자 2009년12월02일	(71) 출원인 고려대학교 산학협력단 서울 성북구 안암동5가 1 (72) 발명자 임해창 서울 서초구 서초1동 1642-34 래미안아파트 서초 7차 102-1002 손대능 서울 노원구 상계9동 보람아파트 207동 208호 이정태 서울 강서구 염창동 한화꿈에그린아파트 104동 1103호 (74) 대리인 특허법인남춘	

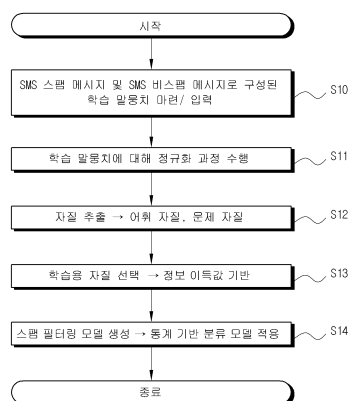
전체 청구항 수 : 총 17 항

(54) 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 장치

(57) 요약

본 발명은 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 단문 스팸 메시지 필터링 장치에 관한 것이다. 본 발명에 따른 스팸 필터링 모델 학습 방법은 (a) 복수의 스팸 메시지와 복수의 비스팸 메시지로 분류된 학습 말뚝치가 입력되는 단계와; (b) 상기 학습 말뚝치로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하는 단계와; (c) 상기 추출된 어휘 자질과 상기 추출된 문체 자질의 정보 이득 값(Information gain value)에 기초하여 복수의 학습용 자질을 선택하는 단계와; (d) 상기 복수의 학습용 자질에 대해 통계 기반 분류 모델을 적용하여 스팸 필터링 모델을 생성하는 단계를 포함하는 것을 특징으로 한다. 이에 따라, SMS 메시지와 같은 단문 메시지의 내용 뿐만 아니라, 그 내용이 어떤 방식으로 작성되었는지, 또는 어떤 언어 표현 양식으로 작성되었는지를 반영하게 되어, 비스팸 메시지를 스팸 메시지로 분류하는 오류를 현저히 감소시킬 수 있다.

대표도 - 도1



특허청구의 범위

청구항 1

단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법에 있어서,

- (a) 복수의 스팸 메시지와 복수의 비스팸 메시지로 분류된 학습 말뭉치가 입력되는 단계와;
- (b) 상기 학습 말뭉치로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하는 단계와;
- (c) 상기 추출된 어휘 자질과 상기 추출된 문체 자질의 정보 이득 값(Information gain value)에 기초하여 복수의 학습용 자질을 선택하는 단계와;
- (d) 상기 복수의 학습용 자질에 대해 통계 기반 분류 모델을 적용하여 스팸 필터링 모델을 생성하는 단계를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 2

제1항에 있어서,

상기 (b) 단계에서의 상기 문체 자질을 추출하는 단계는,

언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질을 추출하는 단계와;

언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질을 추출하는 단계; 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 3

제2항에 있어서,

상기 (b) 단계에서의 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 4

제2항에 있어서,

상기 (b) 단계에서의 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 5

제1항에 있어서,

- (e) 상기 (b) 단계의 수행 전에 상기 학습 말뭉치에 대해 철자 오류 교정 과정과 띄어쓰기 오류 교정 과정 중 적어도 어느 하나를 포함하는 정규화 과정을 수행하는 단계를 더 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 6

제1항에 있어서,

상기 (d) 단계에서 상기 통계 기반 분류 모델은 나이브 베이시언(Naive Bayesian) 분류 모델, 지지 벡터 머신(Support Vector Machine) 모델, 최대 엔트로피(Maximum Entropy Model) 모델 중 어느 하나를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 따른 스팸 필터링 모델 학습 방법을 통해 생성된 스팸 필터링 모델을 이용한

단문 스팸 메시지 필터링 방법에 있어서,

(a1) 필터링 대상 단문 메시지가 입력되는 단계와;

(b1) 상기 필터링 대상 단문 메시지로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하는 단계와;

(c1) 상기 추출된 어휘 자질 및 상기 추출된 문체 자질을 상기 스팸 필터링 모델에 적용하여 상기 필터링 대상 단문 메시지가 단문 스팸 메시지인지 여부를 결정하는 단계를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법.

청구항 8

제7항에 있어서,

상기 (b1) 단계에서의 상기 문체 자질을 추출하는 단계는,

언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질을 추출하는 단계와;

언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질을 추출하는 단계; 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법.

청구항 9

제8항에 있어서,

상기 (b1) 단계에서는 상기 (c) 단계에서 상기 학습용 자질로 선택된 자질이 상기 표층적 문체 자질 또는 상기 구조적 문체 자질로 추출되는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법.

청구항 10

제9항에 있어서,

상기 (b1) 단계에서의 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법.

청구항 11

제9항에 있어서,

상기 (b1) 단계에서의 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법.

청구항 12

통신망을 통해 전송되는 단문 메시지의 스팸 여부를 판단하는 단문 스팸 메시지 필터링 장치에 있어서,

제1항 내지 제6항 중 어느 한 항에 따른 스팸 필터링 모델 학습 방법에 따라 생성된 스팸 필터링 모델을 보유하고;

상기 단문 메시지로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하며;

상기 추출된 어휘 자질 및 상기 추출된 문체 자질을 상기 생성된 스팸 필터링 모델에 적용하여 상기 단문 메시지가 스팸 메시지인지 여부를 판단하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치.

청구항 13

제12항에 있어서,

상기 단문 스팸 메시지 필터링 장치는 언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질과, 언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질 중 적어도 어느 하나를 상기 문체 자질로 추출하는 것을 특징으로

로 하는 단문 스팸 메시지 필터링 장치.

청구항 14

제13항에 있어서,

상기 단문 스팸 메시지 필터링 장치는 상기 (c) 단계에서 상기 학습용 자질로 선택된 자질을 상기 표층적 문체 자질 또는 상기 구조적 문체 자질로 추출하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치.

청구항 15

제14항에 있어서,

상기 단문 스팸 메시지 필터링 장치에 의해 추출되는 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치.

청구항 16

제14항에 있어서,

상기 단문 스팸 메시지 필터링 장치에 의해 추출되는 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치.

청구항 17

제12항에 있어서,

상기 단문 메시지는 SMS 메시지와 댓글 메시지 중 어느 하나를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 장치에 관한 것으로서, 보다 상세하게는 스팸 필터링 모델에 문체 자질 요소를 고려하여 비스팸 메시지를 스팸 메시지로 분류하는 오류를 현저히 감소시킬 수 있는 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 장치에 관한 것이다.

배경 기술

[0002] 과거에는 이동 통신 단말기로 수신되는 SMS(Short Message Service) 메시지 중 스팸 메시지가 수신되면, 수신된 메시지의 내용과 무관하게 무조건적으로 수신하여 사용자에게 알림 기능을 통해 알리도록 되어 있었다. 이와 같은 스팸 메시지의 무분별한 수신은 이동 통신 단말기의 사용자에게 불필요한 메시지 확인이나 음란성 또는 사행성 내용을 제공하는 등 많은 개인적, 사회적 문제를 야기시켰다.

[0003] 이러한 문제점을 해결하기 위해 스팸 메시지를 분리하여 이동 통신 단말기로의 수신을 차단하고자 하는 기술들이 제안되어 왔으며, 그 대표적인 예가 특정 번호를 스팸 메시지를 전송하는 번호로 등록하여 해당 번호로 수신되는 SMS 메시지를 차단하는 것이다.

[0004] 근래에는 SMS 메시지 내에 내포되어 있는 특정 단어를 분석하고, 등록된 단어가 포함된 SMS 메시지를 스팸 메시지로 분류하여 필터링하는 스팸 필터링 기술이 제안되었다.

[0005] 그러나, SMS 메시지와 같은 단문 메시지의 내용은 단문 메시지를 구성하는 정보가 적다는 것을 의미하며, 이는 단문 메시지 중 스팸 메시지의 분류를 어렵게 만드는 요인으로 작용한다.

[0006] 이와 같은 단문 메시지 중 스팸 메시지의 분류의 어려움을 해결하고자 하는 노력으로, 근래에 내용 기반 SMS 필터링 방법이 정보 역과 분야에서 제안되고 있으며, 스팸 필터링시 사용하는 어휘를 확장하고자 하는 기술이 『J.M. Gomez, G. Cajigas, Puertas Sanz E., and Carrero Garcia, F. "Content Based SMS Spam Filtering", In Proceedings of the 2006 ACM Symposium on Document Engineering, pp. 107-114, 2006.』과 『G. V. Cormack, J.M. Gomez, G. Cajigas, and Puertas Sanz E. "Spam Filtering for Short Messages", In Proceedings of CIKM '07, pages 313-320, 2007』을 통해 제안된 바 있다.

[0007] 상기 논문에서 제안하는 스팸 필터링 기술은 스팸 메시지와 비스팸 메시지가 분류되어 있는 학습 데이터에서 스팸 필터 구축에 필요한 자질을 추출하고 있다. 이 때 자질 추출 단위로, [표 1]에 나타난 바와 같이, 단어 뿐만 아니라 음절-bigram(영어의 문자-bigram을 한글에 맞게 적용시킨 기법), 음절-trigram, 직교 희소 단어(Orthogonal sparse)-bigram을 함께 사용하고 있다.

[0008] [표 1]

추출 단위	추출된 자질 - "시간되면 연락주세요!!" -
단어(형태소 단위)	"시간", "되", "면", "연락", "주", "시", "어요"
음절-bigram	"시간", "간되", "되면", "면", "연", "연락", ...
음절-trigram	"시간되", "간되면", "되면", "면연", "면연락", ...
직교 희소 단어-bigram	"시간(1)되", "시간(2)면", "시간(3)연락", "시간(4)주", ...

[0009]

[0010] 상기와 같은 기존의 SMS 스팸 필터링 관련 연구들은 상대적으로 짧은 SMS 메시지의 내용에서 '대출', '이자', 또는 '대리 운전'과 같은 광고성 어휘나 구(Phrase)를 잘 추출하는데 주력하고 있었음을 보여준다.

[0011] 그러나, 상기 기술들은 비스팸 메시지에도 '대출', '이자', 또는 '대리 운전'과 같은 스팸 메시지로 분류될 기준이 되는 광고성 어휘나 구(Phrase) 등의 표현이 나타나지 않는 것을 전제할 때 그 효과가 크다 할 것이다. 그러나, 비스팸 메시지인 경우에도 광고성 어휘나 구(Phrase) 등의 표현, 예컨대, [표 2]에 나타난 바와 같이, '대출', '이자', 또는 '대리 운전'등과 같은 단어가 나타날 수 있으며, 상기 기술들에 의할 경우 이를 스팸 메시지로 분류하는 오류가 발생할 여지가 크다.

[0012] [표 2]

비스팸 메시지	스팸 메시지
철수야 오늘 술 마셨으니 대리 운전 불러..	바로 콜 대리 운전 정성껏 모시겠습니다!!
돈 없어.. 대출 이자 내야되...	최저 대출 이자! 현금 필요시 지금 전화주세요☎

[0013]

발명의 내용

해결 하고자하는 과제

[0014] 이에 본 발명은 상기와 같은 문제점을 해결하기 위해 안출된 것으로써, 스팸 필터링 모델에 문체 자질 요소를 고려하여 비스팸 메시지를 스팸 메시지로 분류하는 오류를 현저히 감소시킬 수 있는 단문 스팸 메시지의 필터링

을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 장치를 제공하는데 그 목적이 있다.

과제 해결수단

- [0015] 상기 목적은 본 발명에 따라, 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법에 있어서, (a) 복수의 스팸 메시지와 복수의 비스팸 메시지로 분류된 학습 말뭉치가 입력되는 단계와; (b) 상기 학습 말뭉치로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하는 단계와; (c) 상기 추출된 어휘 자질과 상기 추출된 문체 자질의 정보 이득 값(Information gain value)에 기초하여 복수의 학습용 자질을 선택하는 단계와; (d) 상기 복수의 학습용 자질에 대해 통계 기반 분류 모델을 적용하여 스팸 필터링 모델을 생성하는 단계를 포함하는 것을 특징으로 하는 스팸 필터링 모델 학습 방법에 의해서 달성된다.
- [0016] 여기서, 상기 (b) 단계에서의 상기 문체 자질을 추출하는 단계는, 언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질을 추출하는 단계와; 언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질을 추출하는 단계; 중 적어도 어느 하나를 포함할 수 있다.
- [0017] 또한, 상기 (b) 단계에서의 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함할 수 있다.
- [0018] 그리고, 상기 (b) 단계에서의 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함할 수 있다.
- [0019] 또한, (e) 상기 (b) 단계의 수행 전에 상기 학습 말뭉치에 대해 철자 오류 교정 과정과 띄어쓰기 오류 교정 과정 중 적어도 어느 하나를 포함하는 정규화 과정을 수행하는 단계를 더 포함할 수 있다.
- [0020] 그리고, 상기 (d) 단계에서 상기 통계 기반 분류 모델은 나이브 베이지언(Naive Bayesian) 분류 모델, 지지 벡터 머신(Support Vector Machine) 모델, 최대 엔트로피(Maximum Entropy Model) 모델 중 어느 하나를 포함할 수 있다.
- [0021] 한편, 상기 목적은 본 발명의 다른 실시 형태에 따라, 상기 스팸 필터링 모델 학습 방법을 통해 생성된 스팸 필터링 모델을 이용한 단문 스팸 메시지 필터링 방법에 있어서, (a1) 필터링 대상 단문 메시지가 입력되는 단계와; (b1) 상기 필터링 대상 단문 메시지로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하는 단계와; (c1) 상기 추출된 어휘 자질 및 상기 추출된 문체 자질을 상기 스팸 필터링 모델에 적용하여 상기 필터링 대상 단문 메시지가 단문 스팸 메시지인지 여부를 결정하는 단계를 포함하는 것을 특징으로 하는 단문 스팸 메시지 필터링 방법에 의해서도 달성된다.
- [0022] 여기서, 상기 (b1) 단계에서의 상기 문체 자질을 추출하는 단계는, 언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질을 추출하는 단계와; 언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질을 추출하는 단계; 중 적어도 어느 하나를 포함할 수 있다.
- [0023] 그리고, 상기 (b1) 단계에서는 상기 (c) 단계에서 상기 학습용 자질로 선택된 자질이 상기 표층적 문체 자질 또는 상기 구조적 문체 자질로 추출될 수 있다.
- [0024] 여기서, 상기 (b1) 단계에서의 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함할 수 있다.
- [0025] 또한, 상기 (b1) 단계에서의 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함할 수 있다.
- [0026] 한편, 상기 목적은 본 발명의 또 다른 실시 형태에 따라, 통신망을 통해 전송되는 단문 메시지의 스팸 여부를 판단하는 단문 스팸 메시지 필터링 장치에 있어서, 상기의 스팸 필터링 모델 학습 방법에 따라 생성된 스팸 필터링 모델을 보유하고; 상기 단문 메시지로부터 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출하며; 상기 추출된 어휘 자질 및 상기 추출된 문체 자질을 상기 생성된 스팸 필터링 모델에 적용하여 상

기 단문 메시지가 스팸 메시지인지 여부를 판단하는 것을 특징으로 하는 단문 스팸 메시지 필터링 장치에 의해서도 달성된다.

- [0027] 여기서, 상기 단문 스팸 메시지 필터링 장치는 언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질과, 언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질 중 적어도 어느 하나를 상기 문체 자질로 추출할 수 있다.
- [0028] 또한, 상기 단문 스팸 메시지 필터링 장치는 상기 (c) 단계에서 상기 학습용 자질로 선택된 자질을 상기 표층적 문체 자질 또는 상기 구조적 문체 자질로 추출할 수 있다.
- [0029] 여기서, 상기 단문 스팸 메시지 필터링 장치에 의해 추출되는 상기 표층적 문체 자질은 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함할 수 있다.
- [0030] 그리고, 상기 단문 스팸 메시지 필터링 장치에 의해 추출되는 상기 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함할 수 있다.
- [0031] 그리고, 상기 단문 메시지는 SMS 메시지와 댓글 메시지 중 어느 하나를 포함할 수 있다.

효 과

- [0032] 본 발명에 따르면, 문체 자질 요소를 고려하여 스팸 필터링 모델을 생성함으로써, SMS 메시지와 같은 단문 메시지의 내용 뿐만 아니라, 그 내용이 어떤 방식으로 작성되었는지, 또는 어떤 언어 표현 양식으로 작성되었는지를 반영하게 되어, 비스팸 메시지를 스팸 메시지로 분류하는 오류를 현저히 감소시킬 수 있는 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 방법 및 이를 이용한 단문 스팸 메시지 필터링 방법 및 장치가 제공된다.
- [0033] 또한, 본 발명에 따르면 SMS 메시지 뿐만 아니라 단문 메시지의 형태를 띄는 웹 페이지 상의 댓글 메시지에서 상기 스팸 필터링 모델을 적용하여 광고성 댓글과 같은 스팸 메시지 형태의 댓글의 등록을 차단하거나 삭제할 수 있다.

발명의 실시를 위한 구체적인 내용

- [0034] 이하에서는 첨부된 도면들을 참조하여 본 발명에 따른 실시예에 대해 상세히 설명한다.
- [0035] 본 발명에 따른 단문 스팸 메시지 필터링 방법은 통계 기반 문서 분류의 관점에서 접근하며, 이에 따라 단문 스팸 메시지의 필터링을 위한 스팸 필터링 모델 학습 과정과, 실제 단문 메시지에서 단문 스팸 메시지를 필터링하는 단문 스팸 메시지 필터링 과정을 수행한다. 여기서, 본 발명에서는 단문 메시지가 SMS(Short Message Service) 메시지인 것을 예로 하여 설명한다.

[0036] 스팸 필터링 모델 학습 과정

- [0037] 이하에서는 도 1을 참조하여 본 발명에 따른 스팸 필터링 모델 학습 방법에 대해 상세히 설명한다.
- [0038] 먼저, 본 발명에 따른 스팸 필터링 모델의 생성을 위해, 복수의 SMS 스팸 메시지와 복수의 SMS 비스팸 메시지로 분류된 학습 말뭉치를 마련한다(S10). 여기서, SMS 스팸 메시지와 SMS 비스팸 메시지는 스팸 필터링 모델의 생성에 적합하도록 일정 비율로 마련되며, 본 발명에서는 SMS 스팸 메시지와 SMS 비스팸 메시지의 비율을 1:9의 비율로 하는 것을 예로 한다. 이 때, 준비된 학습 말뭉치는 본 발명에 따른 스팸 필터링 모델 학습 방법을 수행하는 컴퓨터 등과 같은 정보처리장치에 입력된다(S10).
- [0039] 그런 다음, 준비된 학습 말뭉치에 대해 정규화 과정을 수행한다(S11). 여기서, 학습 말뭉치에 대한 정규화 과정은 빈번히 나타나는 SMS 메시지 내에서의 철자 오류를 교정하거나, 띄어쓰기 오류를 교정하는 과정을 포함할 수 있다. 이를 통해, 후술할 자질 추출 과정에서 추출되는 자질들의 정확성을 향상시키게 된다.
- [0040] 본 발명에서는 절차 오류 교정 모델로 규칙기반 방식이 적용되고, 띄어쓰기 오류 교정 모델로 Hidden Markov 모

텔이 적용되는 것을 예로 하며, S10 단계에서 준비된 학습 말뭉치를 철자 오류나 띄어쓰기 오류가 존재하지 않은 검증된 SMS 메시지를 마련하는 경우 S11 단계인 정규화 과정은 생략될 수 있음은 물론이다.

[0041] 상기와 같이 학습 말뭉치에 대한 정규화 과정이 완료되면, 학습 말뭉치로부터 스팸 필터링 모델의 생성을 위한 자질을 추출한다(S12). 자질 추출 과정은 학습 말뭉치를 구성하는 SMS 메시지들을 컴퓨터와 같은 정보처리장치가 처리 가능한 형태의 정보로 표현하는 과정으로 정의될 수 있으며, 본 발명에 따른 스팸 필터링 모델 학습 방법을 통해 추출되는 자질은 어휘 자질과 문체 자질을 포함한다.

[0042] 어휘 자질은 SMS 메시지의 내용에 기반하는 것으로, [표 1]에 나타난 바와 같이, 단어(또는 형태소 단위), 음절-bigram, 음절-trigram, 직교 회소 단어-bigram을 적용하여 추출된다. 여기서, 어휘 자질의 추출 방법은 기 공지된 다양한 방식이 적용 가능한 바 그 상세한 설명은 생략한다.

[0043] 문체 자질은 학습 말뭉치를 구성하는 각 SMS 메시지의 언어 표현 양식에 기반으로 문체 자질을 추출한다. SMS 스팸 메시지를 발송하는 스팸머는 광고, 권유, 설득 등과 같은 특정 의도를 가지고 메시지를 작성하기 때문에, 스팸머가 사용하는 문체는 비스패머와 차이가 날 수 있다. 문체 자질의 추출은 스팸머와 비스패머의 문체, 즉 언어 표현 양식을 고려하여 SMS 스팸 메시지와 SMS 비스패머 메시지를 구분하는데 사용된다.

[0044] 본 발명에 따른 문체 자질은 언어 표현 양식의 형식적 특징에 기반한 표층적 문체 자질과, 언어 표현 양식의 실질적 특징에 기반한 구조적 문체 자질을 포함할 수 있다.

[0045] 여기서, 표층적 문체 자질은 학습 말뭉치를 구성하는 각 SMS 메시지의 전체 길이, 메시지 내의 단어 평균 길이, 기능어별 빈도수, 품사-trigram, 메시지의 전체 길이 중 특수 기호의 비율, 이모티콘의 빈도수, 및 광고성 기호의 빈도수 중 적어도 어느 하나를 포함할 수 있다.

[0046] 그리고, 구조적 문체 자질은 메시지 내 문장에서 추출된 구(Phrase)의 종류별 빈도수, 문맥-자유-문법 생성 패턴(Context-free-grammar productions pattern), 및 메시지 내 문장의 구문 구조 파싱 트리의 깊이 중 적어도 어느 하나를 포함할 수 있다.

[0047] [표 3]은 본 발명에 따른 스팸 필터링 모델 학습 방법을 통해 추출된 문체 자질, 즉 표층적 문체 자질과 구조적 문체 자질의 예를 나타낸 것이다. [표 3]에 나타난 예에서는 SMS 문자의 내용이 "시간되면 연락주세요!!"이며, 표층적 문체 자질 중 SMS 메시지의 전체 길이가 21(byte)이고, 단어의 평균 길이가 9(byte)가 된다. 도 4에 도示된 바와 같이 각 문체 자질에 대해 추출된 값들이 문체 자질의 값들이 된다.

[0048] [표 3]

문체 자질	추출된 자질값 “시간되면 연락주세요!!”
SMS 메시지의 전체 길이	21
SMS 메시지 내 단어의 평균 길이	9
기능어별 빈도수	(되=1)(면=1)(시=1)(어요=1)
품사-trigram	(NNG XSV EC)(XXV EC NNG)...
SMS 메시지 전체 길이 중 특수기호의 비율	0.095
이모티콘 빈도수	0
광고성 기호 패턴 빈도수	0
추출된[명사구/동사구/형용사구/...]빈도수 전체 추출된 구빈도수	(명사구(NP)=0.4) (동사구(VNP)=0.6)
문맥-자유-문법 생성 패턴	("NP->NNG XSV EC"=1) ("VNP->NNG VCP EF"=1) ...
SMS 메시지 내 문장의 구문구조 파싱트리 깊이	2

[0049]

[0050] 상기의 문체 자질 중 SMS 메시지의 전체 길이의 경우, 통상적으로 SMS 스팸 메시지의 전체 길이가 SMS 비스패머 메시지의 전체 길이보다 짧은 것을 반영한 것이고, SMS 메시지 내의 단어의 평균 길이도 SMS 스팸 메시지가 SMS

비스팸 메시지보다 짧은 것을 반영한다.

- [0051] 또한, 문체 자질 중 기능어별 빈도수의 경우, SMS 스팸 메시지의 존대말의 빈도수가 SMS 비스팸 메시지보다 상대적으로 높을 수 있고, 조사 사용의 빈도수가 SMS 비스팸 메시지가 SMS 스팸 메시지보다 상대적으로 높을 수 있음을 반영하게 된다.
- [0052] 그리고, SMS 메시지 전체 길이 중 특수기호의 비율의 경우, 비스팸의 발신자는 주로 감정을 표현하기 위한 특수기호, 예를 들어 즐거움을 나타내는 '^'나, 슬픔을 나타내는 'ㅠ'와 같은 특수기호를 사용하게 된다. 반면, 스팸머는 \$\$\$나 %와 같이 금전적인 내용의 특수기호를 사용하는 경향이 있다.
- [0053] [표 3]에 나타난 나머지 문체 자질들도 상기와 같이, 스팸 메시지와 비스팸 메시지 사이에서 발생할 수 있는 문체적 차이를 반영하게 된다.
- [0054] 상기와 같은 과정을 통해 학습 말뭉치를 구성하는 SMS 메시지들, 즉 SMS 스팸 메시지들과 SMS 비스팸 메시지들로부터 어휘 자질 및 문체 자질이 추출되면, 추출된 어휘 자질 및 문체 자질들 중 스팸 필터링 모델의 생성에 도움이 되는 자질, 즉 SMS 스팸 메시지와 SMS 비스팸 메시지를 구분하는데 도움이 되는 자질을 선택한다(S13). 이하에서는 어휘 자질 및 문체 자질 중 스팸 필터링 모델의 생성에 사용될 자질을 학습용 자질로 정의하여 설명한다.
- [0055] 본 발명에서는 학습용 자질의 선택을 위해 정보 이득 값(Information gain value)에 기반한 자질 선택 과정이 수행되는 것을 예로 한다. 여기서, 정보 이득 값은 어떤 자질이 SMS 스팸 메시지와 SMS 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지 여부를 나타내는 값으로, 스팸 필터링에 도움이 되지 않는 자질이 존재할 수 있으므로 스팸 필터링 모델 학습 과정에서 정보 이득 값이 높은 상위 자질들만을 학습용 자질로 선택하여 사용하게 된다. 여기서, 학습용 자질의 수는 학습 말뭉치의 크기나 분류 작업에 따라 달라질 수 있다.
- [0056] 상기와 같은 과정을 통해 학습용 자질의 선택이 완료되면, 선택된 학습용 자질에 대해 통계 기반 분류 모델을 적용하여 스팸 필터링 모델을 생성한다(S14). 본 발명에서는 통계 기반 분류 모델로 나이브 베이지언(Naive Bayesian) 분류 모델, 지지 벡터 머신(Support Vector Machine) 모델, 최대 엔트로피(Maximum Entropy Model) 모델 중 어느 하나가 적용되는 것을 예로 한다.
- [0057] **단문 스팸 메시지 필터링 과정**
- [0058] 이하에서는 도 2를 참조하여, 본 발명에 따른 스팸 필터링 모델 학습 방법을 통해 생성된 스팸 필터링 모델을 이용한 단문 스팸 메시지 필터링 방법에 대해 상세히 설명한다.
- [0059] 먼저, 본 발명에 따른 단문 스팸 메시지 필터링 방법을 수행하기 위한 단문 스팸 메시지 필터링 장치에 상기의 스팸 필터링 모델이 등록된 상태(S20)에서, 필터링 대상 단문 메시지, 예를 들어 상술한 바와 같이 SMS 메시지가 입력된다(S21).
- [0060] 그런 다음, 필터링 대상 SMS 메시지에 대해 정규화 과정이 수행된다(S22). 여기서, 필터링 대상 SMS 메시지에 대한 정규화 과정은 철자 오류 교정 과정이나 띄어쓰기 오류 교정 과정이 진행되는 것으로, 상술한 스팸 필터링 모델 학습 과정에서의 정규화 과정(도 1의 S11)에 대응하는 바 그 설명은 생략한다.
- [0061] 필터링 대상 SMS 메시지에 대한 정규화 과정이 완료되면, 필터링 대상 SMS 메시지에서 내용에 기반한 어휘 자질과 언어 표현 양식에 기반한 문체 자질을 추출한다(S23). 여기서, 본 발명에 따른 단문 스팸 메시지 필터링 방법에서 어휘 자질과 문체 자질의 추출은 상술한 스팸 필터링 모델 학습 과정에서의 어휘 자질 및 문체 자질의 추출 과정(도 1의 S12)에 대응하는 바 그 상세한 설명은 생략한다.
- [0062] 본 발명에 따른 단문 스팸 메시지 필터링 과정에서 추출되는 어휘 자질 및 문체 자질은 상술한 스팸 필터링 모델 학습 과정에서 학습용 자질로 선택(도 1의 S13)된 자질, 즉 어휘 자질, 표층적 문체 자질 및 구조적 문체 자질 중 학습용 자질로 선택된 자질 만이 단문 스팸 메시지 필터링 과정에서 필터링 대상 SMS 메시지가 SMS 스팸 메시지인지 SMS 비스팸 메시지인지를 결정하기 위한 자질로 추출되는 것을 예로 한다.
- [0063] 상기와 같은 과정을 통해 추출된 자질, 즉 어휘 자질, 표층적 문체 자질, 및/또는 구조적 문체 자질을 본 발명에 따른 스팸 필터링 모델 학습 방법을 통해 생성된 스팸 필터링 모델에 적용하여(S24), 필터링 대상 SMS 메시지가 SMS 스팸 메시지인지 SMS 비스팸 메시지인지 여부를 결정하게 된다(S25).

[0064] 본 발명에 따른 스팸 필터링 모델 학습 방법 및 단문 스팸 메시지 필터링 방법을 통한 SMS 메시지의 스팸 필터링의 우수성을 검증하기 위해, 상술한 『J.M. Gomez, G. Cajigas, Puertas Sanz E., and Carrero Garcia, F. "Content Based SMS Spam Filtering", In Proceedings of the 2006 ACM Symposium on Document Engineering, pp. 107-114, 2006.』에서 제안하는 필터링 방법을 한국어에 맞게 적용한 Baseline과의 비교 실험을 수행하였다.

[0065] 실험에는 총 20,000개의 한국어 SMS 메시지를 수집하고, 이중 18,000 개를 스팸 필터링 모델 학습 과정에 적용하고, 나머지 2,000개를 단문 스팸 필터링 과정에 적용하기 위한 샘플로 사용하였다. 그리고, SMS 스팸 메시지와 SMS 비스팸 메시지의 비율은 상술한 바와 같이 1:9가 되도록 구성하였다.

[0066] 평가 척도로는 1-AUC(%)를 사용하였으며, 수치가 작을수록 스팸 필터링 성능이 좋은 것으로 판단할 수 있다. 여기서, AUC(%)는 ROC 곡선(Receiver Operating Characteristic curve)을 이용하며, 비스팸이 스팸으로 분류되는 경우를 false-positive로 스팸이 비스팸으로 분류되는 경우를 false-negative로 설정한 후 false-negative의 비율에 대한 false-positive의 비율을 ROC 곡선(Receiver Operating Characteristic curve)으로 하여 AUC(%)을 결정하였다.

[0067] 아래의 [표 4]는 실험 결과를 나타낸 것으로 본 발명에 따른 스팸 필터링 모델 학습 방법 및 단문 스팸 메시지 필터링 방법을 통한 SMS 메시지의 스팸 필터링이 효과적임을 확인할 수 있다.

[0068] [표 4]

	1-AUC(%)
Baseline	9.9651
본 발명	2.7444

[0069]

[0070] 단문 스팸 메시지 필터링 장치 - SMS 메시지

[0071] 이하에서는 상술한 스팸 필터링 모델 학습 방법과 단문 스팸 메시지 필터링 방법을 이용하는 단문 스팸 메시지 필터링 장치에 대해 설명한다.

[0072] 본 발명에 따른 단문 스팸 메시지 필터링 장치는 통신망을 통해 전송되는 단문 메시지를 관리한다. 여기서, 단문 스팸 메시지 필터링 장치는 상술한 스팸 필터링 모델 학습 방법에 따라 스팸 필터링 모델을 보유하며, 상술한 단문 스팸 메시지 필터링 방법에 따라 단문 메시지의 스팸 여부를 판단한다.

[0073] 도 3 내지 도 5는 본 발명에 따른 단문 스팸 메시지 필터링 장치가 이동 통신망(70)을 통해 송수신되는 SMS 메시지의 스팸 여부를 판단하는 예를 설명하기 위한 도면으로, 제1 이동 통신 단말기(10a)와 제2 이동 통신 단말기(10b)가 이동 통신망(70)을 통해 상호 통신하고, SMS 중계 서버(30)는 제1 이동 통신 단말기(10a)와 제2 이동 통신 단말기(10b) 간의 SMS 메시지의 송수신을 중계한다.

[0074] 여기서, 도 4는 본 발명에 따른 단문 스팸 메시지 필터링 장치가 SMS 중계 서버(30)에 적용되어 동작하는 예를 도시한 도면으로, SMS 중계 서버(30)는 상술한 스팸 필터링 모델 학습 방법에 따라 스팸 필터링 모델을 보유하게 된다(S40). 그리고, SMS 중계 서버(30)는 생성된 스팸 필터링 모델을 스팸 필터링 데이터베이스(50)에 저장한다. 여기서, SMS 중계 서버(30)가 스팸 필터링 모델을 생성하는 과정은 상술한 바와 같은 바, 그 설명은 생략한다.

[0075] 한편, 제2 이동 통신 단말기(10b)의 사용자가 자신의 단말기를 이용하여 SMS 문자 메시지를 작성한 후(S41) SMS 메시지를 제1 이동 통신 단말기(10a)로 전달하기 위해 전송하면, 해당 SMS 메시지가 SMS 중계 서버(30)로 전송된다(S42).

[0076] 제2 이동 통신 단말기(10b)로부터 전송된 SMS 메시지가 SMS 중계 서버(30)에 수신되면, SMS 중계 서버(30)는 본 발명에 따른 단문 스팸 메시지 필터링 방법을 수행하여 스팸 여부를 판단한다. 즉, SMS 중계 서버(30)는, 도 4에 도시된 바와 같이, 수신된 SMS 메시지에 대해 정규화 과정을 수행한 후(S43), SMS 메시지에서 어휘 자질 및 문체 자질을 추출한다(S44). 그리고, SMS 중계 서버(30)는 추출된 자질을 스팸 필터링 모델에 적용하여

(S45), 제2 이동 통신 단말기(10b)로부터 수신된 SMS 메시지가 스팸 메시지인지 여부를 판단한다(S46).

[0077] 그리고, SMS 중계 서버(30)는 S47 단계에서 SMS 메시지가 스팸 메시지로 판단되면 제1 이동 통신 단말기(10a)로의 SMS 메시지의 전송을 차단하고(S47), SMS 메시지가 비스팸 메시지로 판단되면 해당 SMS 메시지를 제1 이동 통신 단말기(10a)로 전송함으로써(S48) 제1 이동 통신 단말기(10a)가 비스팸 메시지만을 수신 가능하게 된다(S49).

[0078] 한편, 도 5는 본 발명에 따른 단문 스팸 메시지 필터링 장치가 도 3에 도시된 제1 이동 통신 단말기(10a)에 적용되어 동작하는 예를 도시한 도면이다.

[0079] SMS 중계 서버(30)는 스팸 필터링 모델 학습 방법에 따라 스팸 필터링 모델을 생성하여 보유한다(S50). 그리고, SMS 중계 서버(30)는 생성된 스팸 필터링 모델을 스팸 필터링 데이터베이스(50)에 저장한다. 여기서, SMS 중계 서버(30)가 스팸 필터링 모델을 생성하는 과정은 상술한 바와 같은 바, 그 설명은 생략한다.

[0080] 그리고, SMS 중계 서버(30)는 자신이 보유한 스팸 필터링 모델을 제1 이동 통신 단말기(10a)로 전송하고(S51), 제1 이동 통신 단말기(10a)는 SMS 중계 서버(30)로부터 전송된 스팸 필터링 모델을 수신하여 등록(S52)함으로써 본 발명에 따른 단문 스팸 메시지 필터링 장치로 동작하는 제1 이동 통신 단말기(10a)가 스팸 필터링 모델을 보유하게 된다. 여기서, 제1 이동 통신 단말기(10a)가 SMS 중계 서버(30)로부터 스팸 필터링 모델을 수신하는 방법은 제1 이동 통신 단말기(10a)를 통해 무선 인터넷(70a)에 접속하여 전송받거나, 컴퓨터 등을 통해 전송받은 스팸 필터링 모델을 다시 제1 이동 통신 단말기(10a)로 등록하는 방법 등 다양한 형태로 구현 가능하다.

[0081] 상기와 같은 과정을 통해 제1 이동 통신 단말기(10a)에 본 발명에 따른 스팸 필터링 모델이 등록된 상태에서, 제2 이동 통신 단말기(10b)의 사용자가 자신의 단말기를 이용하여 SMS 문자 메시지를 작성한 후(S53) SMS 메시지를 제1 이동 통신 단말기(10a)로 전달하기 위해 전송하면, SMS 중계 서버(30)는 SMS 메시지가 제1 이동 통신 단말기(10a)에 수신 가능하도록 SMS 메시지의 전송을 중계한다(S55).

[0082] 그런 다음, SMS 중계 서버(30)의 중계 과정을 통해 SMS 메시지가 제1 이동 통신 단말기(10a)에 수신되면, 제1 이동 통신 단말기(10a)는 S52 단계에서 등록된 스팸 필터링 모델을 이용하여 본 발명에 따른 단문 스팸 메시지 필터링 방법을 수행하여 스팸 여부를 판단한다.

[0083] 즉, 제1 이동 통신 단말기(10a)는, 도 5에 도시된 바와 같이, 수신된 SMS 메시지에 대해 정규화 과정을 수행한 후(S56), SMS 메시지로부터 어휘 자질 및 문체 자질을 추출한다(S57). 그리고, 제1 이동 통신 단말기(10a)는 추출된 자질을 스팸 필터링 모델에 적용하여(S58), 제2 이동 통신 단말기(10b)로부터 수신된 SMS 메시지가 스팸 메시지인지 여부를 판단한다(S59).

[0084] 그리고, 제1 이동 통신 단말기(10a)는 S59 단계에서 SMS 메시지가 비스팸 메시지로 판단되면, 수신된 SMS 메시지를 화면 상에 표시하거나(S60) 메시지 수신 알람 기능을 수행하는 등 SMS 메시지의 수신에 따른 동작을 수행한다.

[0085] 반면, 제1 이동 통신 단말기(10a)는 S59 단계에서 SMS 메시지가 스팸 메시지로 판단되면, 스팸 메시지의 수신에 따른 기 설정된 동작을 수행한다(S61). 예를 들어, 제1 이동 통신 단말기(10a)는 메시지 수신 알람을 동작시키지 않은 상태에서 수신된 SMS 메시지를 스팸 보관함에 저장하거나 해당 SMS 메시지를 전송한 제2 이동 통신 단말기(10b)의 전화번호를 스팸 전화번호로 등록할 수 있다.

[0086] 단문 스팸 메시지 필터링 장치 - 댓글 메시지

[0087] 이하에서는 상술한 스팸 필터링 모델 학습 방법과 단문 스팸 메시지 필터링 방법을 이용하는 단문 스팸 메시지 필터링 장치의 다른 실시 형태를 도 6 및 도 7을 참조하여 설명한다.

[0088] 도 6을 참조하여 설명하면, 클라이언트 단말기(10c, 10d, 10e)와 웹 서버(30a)는 인터넷(70a)과 같은 네트워크를 통해 상호 통신 가능하게 마련된다. 여기서, 클라이언트 단말기(10c, 10d, 10e)는 인터넷(70a)을 통해 웹 서버(30a)에 접속 가능한 노트북과 같은 컴퓨터나 PDA 등 다양한 형태로 마련된다.

[0089] 웹 서버(30a)는 다양한 콘텐츠를 제공하기 위한 웹 페이지를 인터넷(70a)을 통해 제공하고, 클라이언트 단말기(10c, 10d, 10e)의 사용자는 인터넷(70a)을 통해 웹 서버(30a)에서 제공하는 웹 페이지에 접속한다. 그리고, 웹 서버(30a)에서 제공되는 웹 페이지에는 단문 형태의 댓글(이하, '댓글 메시지'라 함)의 등록이 가능하게 마련된다.

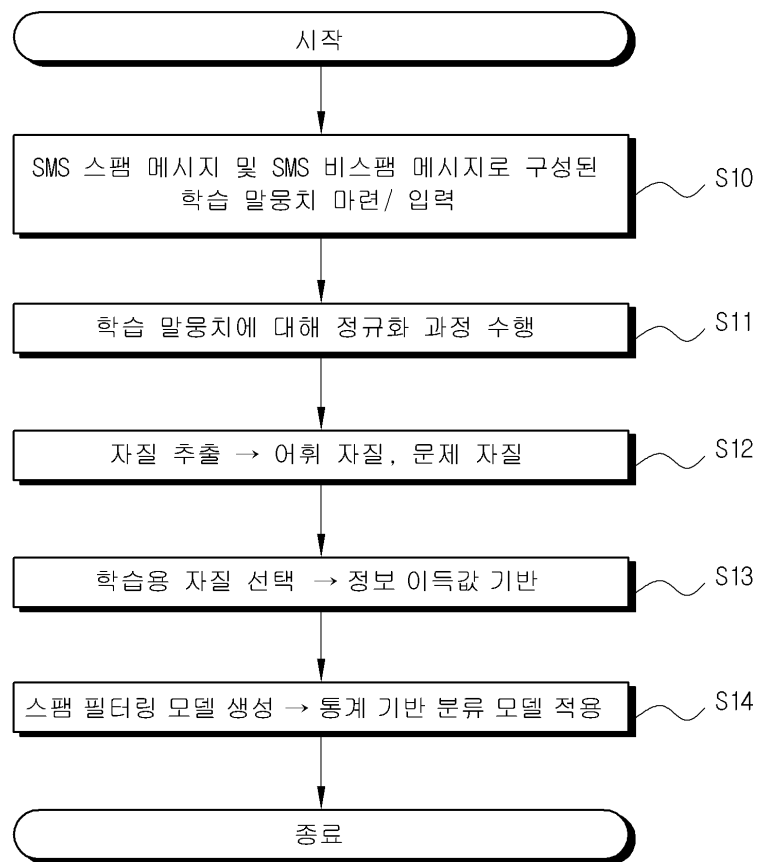
- [0090] 도 7은 본 발명에 따른 단문 스팸 메시지 필터링 장치가 도 6에 도시된 웹 서버(30a)에 적용되어 동작하는 예를 도시한 도면이다. 도 7을 참조하여 설명하면, 웹 서버(30a)는 상술한 스팸 필터링 모델 학습 방법에 따라 스팸 필터링 모델을 생성하여 보유한다(S70). 그리고, 웹 서버(30a)는 생성된 스팸 필터링 모델을 스팸 필터링 데이터베이스(50a)에 저장한다. 여기서, 웹 서버(30a)가 스팸 필터링 모델을 생성하는 과정은 상술한 바와 같은 바, 그 설명은 생략한다.
- [0091] 한편, 클라이언트 단말기(10c, 10d, 10e)의 사용자가 자신의 단말기를 이용하여 웹 페이지에 접속하고(S71), 웹 페이지에 마련된 댓글창에 댓글 메시지를 작성하면(S72), 웹 서버(30a)는 해당 댓글 메시지를 등록한다(S73).
- [0092] 이 때, 웹 서버(30a)는 본 발명에 따른 단문 스팸 메시지 필터링 방법을 수행하여 해당 댓글 메시지에 대한 스팸 여부를 판단한다. 즉, 웹 서버(30a)는, 도 7에 도시된 바와 같이, 등록된 댓글 메시지에 대해 정규화 과정을 수행한 후(S74), 댓글 메시지로부터 어휘 자질 및 문체 자질을 추출한다(S75). 그리고, 웹 서버(30a)는 추출된 자질을 스팸 필터링 모델에 적용하여(S76), 클라이언트 단말기(10c, 10d, 10e)로부터 등록된 댓글 메시지가 스팸 메시지인지 여부를 판단한다(S77).
- [0093] 그리고, 웹 서버(30a)는 S77 단계에서 댓글 메시지가 스팸 메시지로 판단되면 해당 댓글 메시지를 스팸으로 처리한다(S79). 예를 들어, 웹 서버(30a)는 댓글 메시지가 스팸 메시지로 판단되면 해당 댓글 메시지를 임의로 삭제할 수 있다. 반면, 웹 서버(30a)는 댓글 메시지가 비스팸 메시지로 판단되면 댓글 메시지의 등록 상태를 유지하는 등, 해당 댓글 메시지를 비스팸으로 처리하게 된다(S78).
- [0094] 전술한 실시예에서는, 도 7에 도시된 바와 같이, 웹 서버(30a)가 댓글 메시지의 등록 후에 스팸 여부를 판단하는 것을 예로 하고 있으나, 댓글 메시지의 등록 전에 스팸 여부를 판단한 후 등록 여부를 결정할 수 있음은 물론이다.
- [0095] 이상에서 본 발명의 바람직한 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면의 간단한 설명

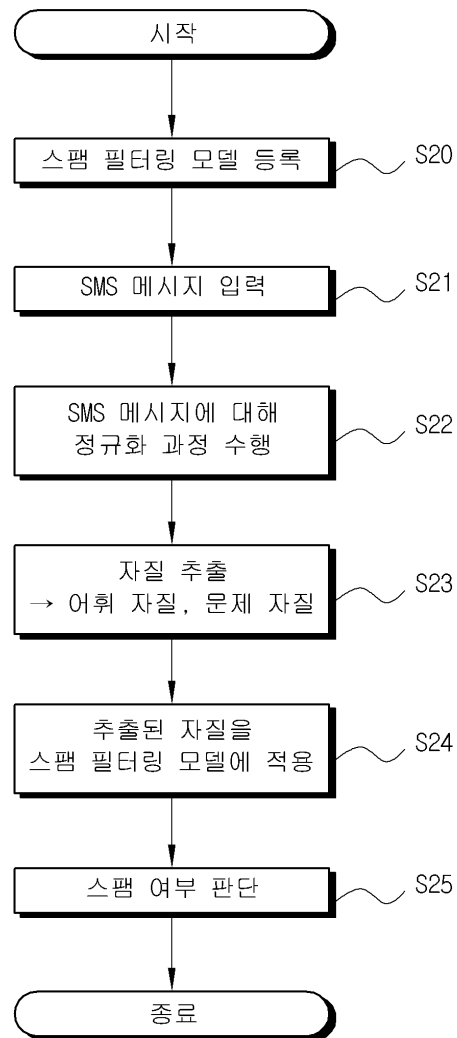
- [0096] 도 1은 본 발명에 따른 스팸 필터링 모델 학습 방법을 설명하기 위한 도면이고,
- [0097] 도 2는 본 발명에 따른 단문 스팸 메시지 필터링 방법을 설명하기 위한 도면이고,
- [0098] 도 3 내지 5는 본 발명의 실시예에 따른 스팸 메시지 필터링 시스템을 설명하기 위한 도면이고,
- [0099] 도 6 및 도 7은 본 발명의 다른 실시예에 따른 스팸 메시지 필터링 시스템을 설명하기 위한 도면이다.
- [0100] <도면의 주요 번호에 대한 설명>
- [0101] 10a : 제1 이동 통신 단말기 10b : 제2 이동 통신 단말기
- [0102] 30 : SMS 중계 서버 50 : 데이터베이스
- [0103] 70 : 이동 통신망

도면

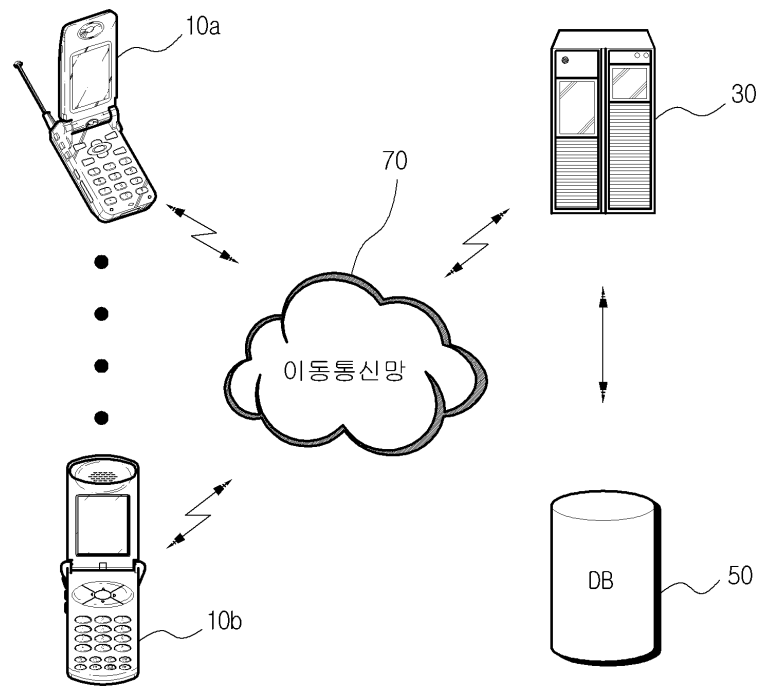
도면1



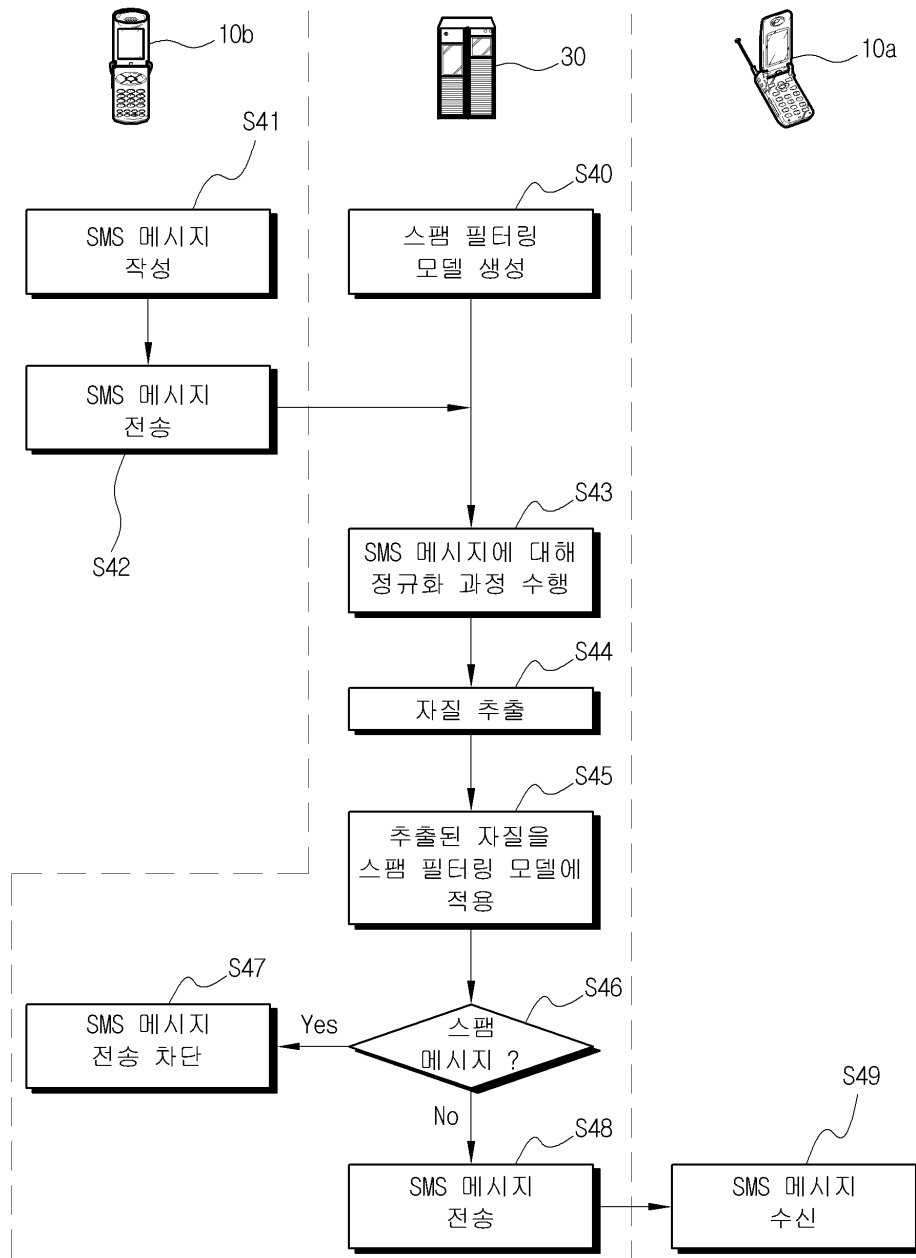
도면2



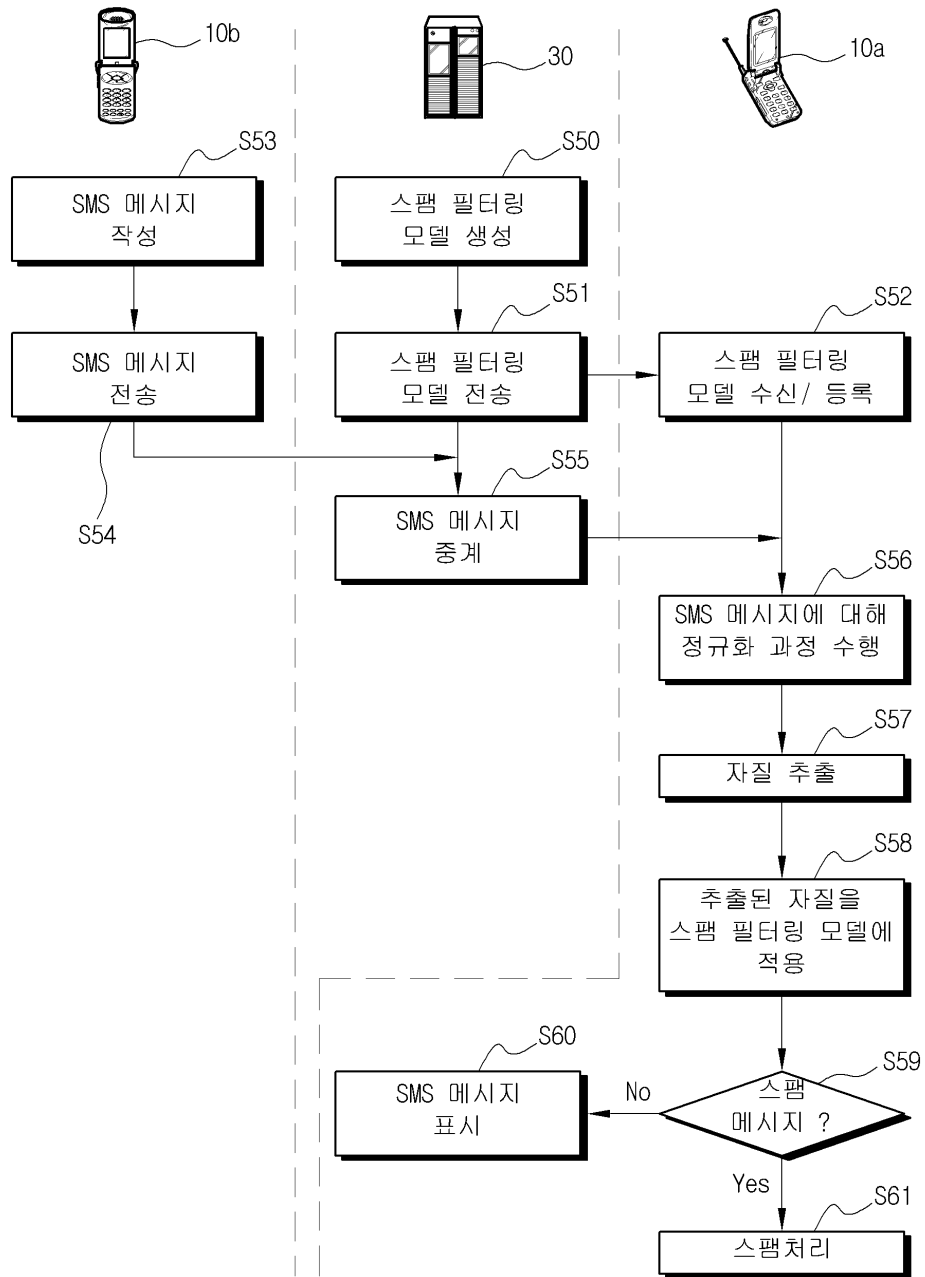
도면3



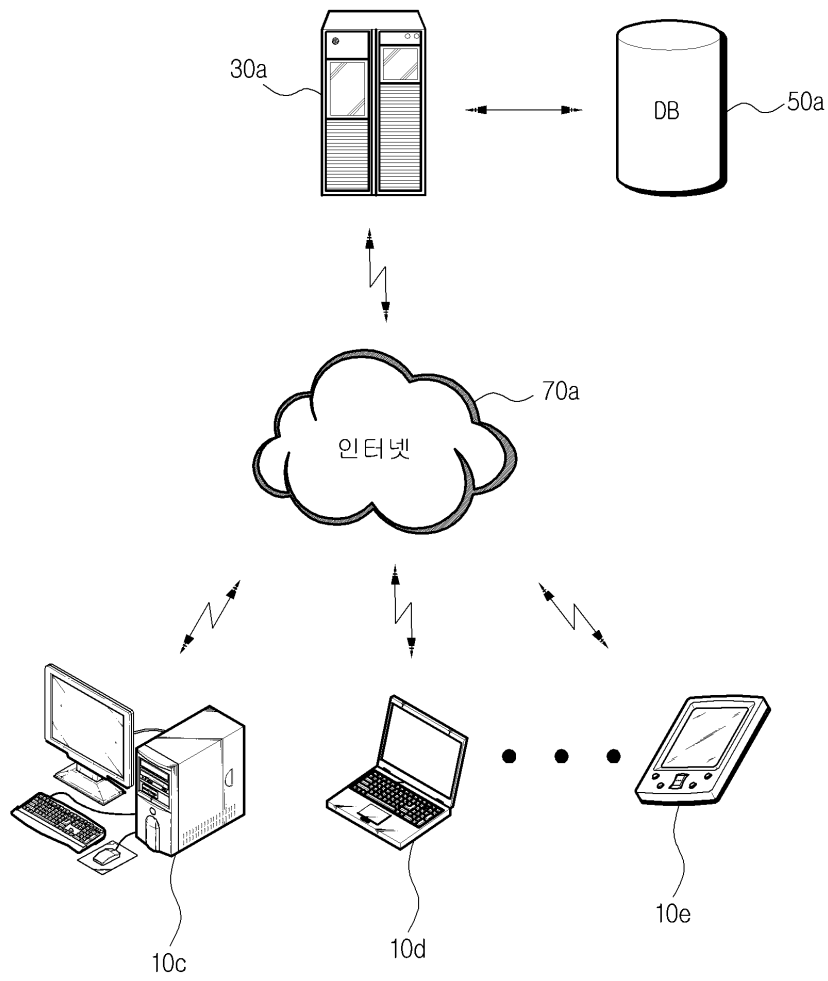
도면4



도면5



도면6



도면7

