



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0096335
(43) 공개일자 2010년09월02일

(51) Int. Cl.

G06Q 50/00 (2006.01) H04W 88/18 (2009.01)
H04W 4/14 (2009.01)

(21) 출원번호 10-2009-0015164

(22) 출원일자 2009년02월24일

심사청구일자 2009년02월24일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

임해창

서울특별시 서초구 서초1동 1642-34 래미안아파트
서초7차 102-1002

(74) 대리인

특허법인화우

전체 청구항 수 : 총 29 항

(54) 한국어 단문 메시지 스팸 필터링 시스템 및 방법

(57) 요약

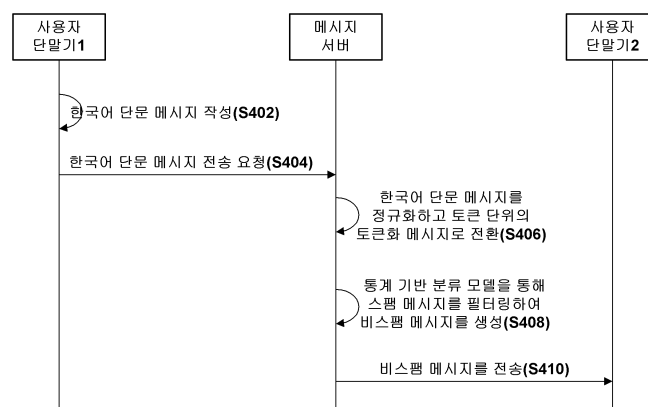
본 발명은 한국어 단문 메시지에서 형태소를 추출해 형태소 단위의 토큰들로 표현하여 토큰화하고, 이렇게 형성된 형태소 단위의 토큰들에 대해, 토큰화 과정과 정보이득 기반 자질 선택 과정으로 학습된 통계 기반 분류 모델을 이용해 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링하도록 하는, 한국어 단문 메시지 스팸 필터링 시스템 및 방법에 관한 것이다.

본 발명에 따른 한국어 단문 메시지 스팸 필터링 방법은, (a) 외부로부터 입력된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및 (b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 스팸과 비스팸으로 분류하는 단계를 포함한다.

여기서, 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델이다.

본 발명에 의하면, 의미를 나타내는 최소 단위인 형태소를 사용함으로써 스팸 필터링에 효과적인 어휘들을 정확히 추출할 수 있는 장점이 있다. 또한, 스팸 필터링에 효과적인 어휘들을 추출함에 따라 스팸과 비스팸 메시지를 구분할 시 정확률을 높일 수 있다. 또한, 실질형태소와 기능형태소가 조합된 어절에 대해서도 형태소 정보를 이용함으로써 스팸 분류 시 생길 수 있는 미등록어(Out of Vocabulary)를 줄일 수 있다. 그리고, 사용자가 직접 스팸 필터링을 위한 특정 단어를 직접 등록할 필요 없이 스팸 학습 말뭉치와 비스팸 학습 말뭉치를 이용하여 스팸 필터링에 도움이 되는 어휘들을 대량으로 추출함으로써 사용자의 편의성을 대폭 향상시킬 수 있다.

대표도 - 도4



특허청구의 범위

청구항 1

(a) 외부로부터 입력된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및
(b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 스팸과비스팸으로 분류하는 단계;
를 포함하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 2

제 1 항에 있어서,

상기 (b) 단계에서 상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 정규화된 메시지를 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델인 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 3

제 2 항에 있어서,

상기 정보이득 값은, 어떤 토큰이 스팸 또는 비스팸에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 4

제 1 항에 있어서,

상기 통계 기반 분류 모델은, 나이브 베이지언(Naive Bayesian) 분류기, 지지 벡터 머신(Support Vector Machine), 최대 엔트로피 모델(Maximum Entropy Model)중 하나 이상의 모델로 구성된 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 5

제 1 항에 있어서,

상기 (b) 단계는, 상기 형태소 단위의 토큰에 의해 미등록어(Out of Vocabulary)를 줄여주는 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 6

제 1 항에 있어서,

상기 (a) 단계 이전에 상기 한국어 단문 메시지에 대해 철자 오류나 띄어쓰기 오류를 교정하여 정규화하는 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 방법.

청구항 7

외부로부터 입력된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여토큰화하고, 상기 형태소 단위의 토큰들에 대해 스팸 필터링 모델을 이용하여 스팸 메시지를 필터링하는 메시지 서버;

를 포함하는 한국어 단문 메시지 스팸 필터링 시스템.

청구항 8

제 7 항에 있어서,

상기 스팸 필터링 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 시스템.

청구항 9

제 8 항에 있어서,

상기 정보이득 값은, 어떤 토큰이 스팸 또는 비스팸에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 시스템.

청구항 10

통신망을 통해 사용자 단말기들과 통신하기 위한 통신부;

스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 학습된 통계 기반 분류 모델이 저장되어 있는 분류모델 저장부;

상기 사용자 단말기로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하고, 상기 형태소 단위의 토큰들에 대해 상기 통계 기반 분류 모델을 통해 스팸과 비스팸으로 구분하는 스팸 필터링부;

상기 스팸 필터링부를 통해 구분된 비스팸 메시지를 상기 사용자 단말기에게 전송해 주는 메시지 전송부; 및

상기 사용자 단말기로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 비스팸 메시지가 상기 사용자 단말기에게 전송되도록 제어하는 제어부;

를 포함하는 한국어 단문 메시지 스팸 필터링 서버.

청구항 11

제 10 항에 있어서,

상기 스팸 필터링부는, 상기 한국어 단문 메시지를 토큰화하기 이전에 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하는 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 서버.

청구항 12

제 10 항에 있어서,

상기 통계 기반 분류 모델의 정보이득 기반 자질 선택 과정은, 상기 형태소 단위의 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 상기 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택하는 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 서버.

청구항 13

제 10 항에 있어서,

상기 통계 기반 분류 모델은, 나이브 베이시언(Naive Bayesian) 분류기, 지지 벡터 머신(Support Vector Machine), 최대 엔트로피 모델(Maximum Entropy Model) 중 하나 이상의 모델로 구성된 것을 특징으로 하는 한국어 단문 메시지 스팸 필터링 서버.

청구항 14

발신측 단말기로부터 한국어 단문 메시지를 수신하여 착신측 단말기로 전송해 주는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법으로서,

(a) 상기 발신측 단말기로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및

(b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 스팸 메시지와 비스팸 메시지로 구분하는 단계;

를 포함하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 15

제 14 항에 있어서,

(c) 상기 비스팸 메시지를 상기 착신측 단말기로 전송하는 단계;

를 더 포함하는 것을 특징으로 하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 16

제 14 항에 있어서,

상기 (a) 단계는, 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하는 정규화 과정을 포함하는 것을 특징으로 하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 17

제 14 항에 있어서,

상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 상기 정규화된 메시지를 형태소 단위의 토큰들로 형성하여 토큰화한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델인 것을 특징으로 하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 18

제 14 항에 있어서,

상기 정보이득 값은, 상기 토큰들 중에 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것을 특징으로 하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 19

제 16 항에 있어서,

상기 통계 기반 분류 모델은, 나이브 베이지언(Naive Bayesian) 분류기, 지지 벡터 머신(Support Vector Machine), 최대 엔트로피 모델(Maximum Entropy Model) 중 하나 이상의 모델인 것을 특징으로 하는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법.

청구항 20

통신망을 통해 메시지 서버와 통신하기 위한 통신부;

스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 학습된 통계 기반 분류 모델이 저장되어 있는 분류모델 저장부;

상기 메시지 서버로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하고, 상기 형태소 단위의 토큰들에 대해 상기 통계 기반 분류 모델을 통해 스팸 메시지와 비스팸 메시지로 분류하는 메시지 필터링부;

상기 비스팸 메시지를 디스플레이하기 위한 표시부; 및

상기 메시지 서버로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 얻은 비스팸 메시지만을 저장하거나 표시하도록 제어하는 제어부;

를 포함하는 사용자 단말기.

청구항 21

제 20 항에 있어서,

상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 정규화된 메시지를 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델인 것을 특징으로 하는 사용자 단말기.

청구항 22

제 20 항에 있어서,

상기 메시지 필터링부를 통해 분류된 비스팸 메시지를 저장하기 위한 저장부;

를 더 포함하는 것을 특징으로 하는 사용자 단말기.

청구항 23

메시지 서버로부터 수신한 한국어 단문 메시지를 표시하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법으로서,

(a) 상기 메시지 서버로부터 수신한 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및

(b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 이용해 스팸 메시지와 비스팸 메시지로 분류하는 단계;

를 포함하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법.

청구항 24

제 23 항에 있어서,

상기 (a) 단계는, 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하는 정규화 과정을 포함하는 것을 특징으로 하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법.

청구항 25

제 23 항에 있어서,

(c) 상기 비스팸 메시지를 표시하는 단계;

를 더 포함하는 것을 특징으로 하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법.

청구항 26

제 23 항에 있어서,

상기 (b) 단계는, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 통계 기반 분류 모델을 이용하는 것을 특징으로 하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법.

청구항 27

제 26 항에 있어서,

상기 정보이득 값은, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것을 특징으로 하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법.

청구항 28

제 14 항 내지 제 19 항 중 어느 한 항의 방법을 프로그램으로 기록한 컴퓨터로 읽을 수 있는 매체.

청구항 29

제 20 항 내지 제 27 항 중 어느 한 항의 방법을 프로그램으로 기록한 컴퓨터로 읽을 수 있는 매체.

명세서

발명의 상세한 설명

기술 분야

[0001]

본 발명은 이동 단말기에 수신되는 한국어 SMS 메시지에 대한 스팸 메시지를 필터링하는 시스템 및 방법에 관한 것으로, 더욱 상세하게는 자연어 처리 분야에서 문서 분류와 형태소/품사에 근거한 정보 여과에 속하고, 자연어 처리를 인공지능으로 수행하는 것으로, 스팸 메시지와 비스팸 메시지의 말뭉치를 이용해 형태소 단위에서 스팸 메시지 구분에 도움을 주는 자질을 추출하여 필터링에 이용하는, 한국어 단문 메시지 스팸 필터링 시스템 및 방

법에 관한 것이다.

배경 기술

- [0002] 종래에 이동통신 단말기는 스팸 SMS(Short Message Service) 메시지가 수신되면, 수신된 메시지의 내용과 상관 없이 무조건 수신하여 사용자에게 알람 기능을 통하여 알리도록 되어 있다.
- [0003] 이동통신 단말기로 수신되는 스팸 SMS 메시지 중에는 각종 음성 메시지와 특정 문자열을 갖는 메시지, 예를 들어 각종 음란성 또는 사행성 내용을 담고 있는 광고성 메시지 등으로 사용자에게 필요없는 선정적인 문구가 대부분인 메시지가 무차별적으로 수신되고 있다.
- [0004] 이러한 문제점을 해결하기 위하여 종래의 기술은 이동통신 단말기에 수신을 거부하고자 하는 특정 단어를 등록해서 선택적으로 스팸 SMS 메시지를 필터링하고 있다.
- [0005] 그러나 이러한 방법은 사용자 개개인이 스팸 메시지에 자주 나타나는 단어를 직접 등록해야 한다는 번거로운 점이 있다.
- [0006] 특히, 교착어인 한국어의 특성 상 매우 다양한 어절들이 사용되며, 다양한 통신어 표현들이 새로 생겨나고 있기 때문에 이를 수작업으로 일일이 등록하는 데에 한계가 있는 문제점이 있다.
- [0007] 한편, 내용 기반 SMS 스팸 필터링은 정보 여과 분야에서 비교적 최근에 연구가 시작되었다. 하지만, 이들 연구는 영어와 같은 외래어로 작성된 SMS 메시지를 대상으로 한 것으로써, 한국어 SMS 스팸 필터링에 그대로 적용하는데 어려움이 있다.
- [0008] 한국어 어절은 실질형태소와 기능형태소가 결합되어 생성된다. 예를 들어, "연락해라", "연락해 주세요", "연락줘"는 "연락"이라는 기능형태소에 실질형태소들이 결합되어 생성된 어절들이다. 이와 같은 어휘결합의 다양성은 SMS 메시지의 길이 제한과 더불어 한국어 SMS 스팸 필터링 시 어휘정보 부족 문제를 더욱 심화시키고 있다. 또한 기존 연구에서 영어의 단어보다 더 작은 단위에서 의미를 고려하기 위해 사용된 문자-bigram(두 개의 연속된 문자열)과 문자-trigram(세 개의 연속된 알파벳)은 한국어 단어의 의미 부분을 정확히 추출해 낼 수 없는 한계의 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- [0009] 전술한 문제점을 해결하기 위한 본 발명의 목적은, 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 자동으로 교정하여 정규화하고, 한국어 메시지에서 형태소를 추출해 형태소 단위의 토큰들로 표현하여 토큰화하고, 토큰화 과정과 정보이득 기반 자질 선택 과정으로 학습된 통계 기반 분류 모델을 이용해 형태소 단위의 토큰들을 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링형태소를 추출하고 의미적으로 더 작은 단위의 토큰으로 표현하여 토큰화하며, 임의의 토큰이 스팸 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 대한 정보 이득에 기반하여 정보 이득 값이 높은 상위 토큰들만을 자질로써 사용하여 스팸 메시지와 비스팸 메시지로 분류하도록 하는, 한국어 단문 메시지 스팸 필터링 시스템 및 방법, 한국어 메시지 스팸 필터링 메시지 서버 및 그의 한국어 단문 메시지 스팸 필터링 방법, 한국어 메시지 스팸 필터링 사용자 단말기 및 그의 한국어 단문 메시지 스팸 필터링 방법을 제공함에 있다.

과제 해결수단

- [0010] 전술한 목적을 달성하기 위한 본 발명에 따른 한국어 단문 메시지 스팸 필터링 방법은, (a) 외부로부터 입력된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및 (b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 스팸과 비스팸으로 분류하는 단계를 포함한다.
- [0011] 이때, 상기 (b) 단계에서 상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 정규화된 메시지를 형태소 단위의 토큰들

로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델이다.

- [0012] 여기서, 상기 정보이득 값은, 어떤 토큰이 스팸 또는 비스팸에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것이다.
- [0013] 또한, 상기 통계 기반 분류 모델은, 나이브 베이지언 분류기, 지지 벡터 머신, 최대 엔트로피 모델 중 하나 이상의 모델로 구성된다.
- [0014] 또한, 상기 (b) 단계는, 상기 형태소 단위의 토큰에 의해 미등록어를 줄여준다.
- [0015] 그리고, 상기 (a) 단계 이전에 상기 한국어 단문 메시지에 대해 철자 오류나 띄어쓰기 오류를 교정하여 정규화한다.
- [0016] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 한국어 단문 메시지 스팸 필터링 시스템은, 외부로부터 입력된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여토큰화하고, 상기 형태소 단위의 토큰들에 대해 스팸 필터링 모델을 이용하여 스팸 메시지를 필터링하는 메시지 서버를 포함한다.
- [0017] 여기서, 상기 스팸 필터링 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 것이다. 그리고, 상기 정보이득 값은, 어떤 토큰이 스팸 또는 비스팸에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것이다.
- [0018] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 한국어 단문 메시지 스팸 필터링 서버는, 통신망을 통해 사용자 단말기들과 통신하기 위한 통신부; 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 학습된 통계 기반 분류 모델이 저장되어 있는 분류 모델 저장부; 상기 사용자 단말기로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하고, 상기 형태소 단위의 토큰들에 대해 상기 통계 기반 분류 모델을 통해 스팸과 비스팸으로 구분하는 스팸 필터링부; 상기 스팸 필터링부를 통해 구분된 비스팸 메시지를 상기 사용자 단말기에게 전송해 주는 메시지 전송부; 및 상기 사용자 단말기로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 비스팸 메시지가 상기 사용자 단말기에게 전송되도록 제어하는 제어부를 포함한다.
- [0019] 또한, 상기 스팸 필터링부는, 상기 한국어 단문 메시지를 토큰화하기 이전에 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화한다.
- [0020] 또한, 상기 통계 기반 분류 모델의 정보이득 기반 자질 선택 과정은, 상기 형태소 단위의 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 상기 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택한다.
- [0021] 그리고, 상기 통계 기반 분류 모델은, 나이브 베이지언 분류기, 지지 벡터 머신, 최대 엔트로피 모델 중 하나 이상의 모델로 구성될 수 있다.
- [0022] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법은, 발신측 단말기로부터 한국어 단문 메시지를 수신하여 착신측 단말기로 전송해 주는 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법으로서, (a) 상기 발신측 단말기로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및 (b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 스팸 메시지와 비스팸 메시지로 구분하는 단계를 포함한다.
- [0023] 또한, (c) 상기 비스팸 메시지를 상기 착신측 단말기로 전송하는 단계를 더 포함한다.
- [0024] 또한, 상기 (a) 단계는, 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하는 정규화 과정을 포함한다.
- [0025] 또한, 상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 상기 정규화된 메시지를 형태소 단위의 토큰들로 형성하여 토큰화

큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델이다.

[0026] 여기서, 상기 정보이득 값은, 상기 토큰들 중에 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것이다.

[0027] 그리고, 상기 통계 기반 분류 모델은, 나이브 베이저언 분류기, 지지 벡터 머신, 최대 엔트로피 모델 중 하나 이상의 모델일 수 있다.

[0028] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 사용자 단말기는, 통신망을 통해 메시지 서버와 통신하기 위한 통신부; 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 학습된 통계 기반 분류 모델이 저장되어 있는 분류모델 저장부; 상기 메시지 서버로부터 수신된 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하고, 상기 형태소 단위의 토큰들에 대해 상기 통계 기반 분류 모델을 통해 스팸 메시지와 비스팸 메시지로 분류하는 메시지 필터링부; 상기 비스팸 메시지를 디스플레이하기 위한 표시부; 및 상기 메시지 서버로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 얻은 비스팸 메시지만을 저장하거나 표시하도록 제어하는 제어부를 포함한다.

[0029] 여기서, 상기 통계 기반 분류 모델은, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 정규화된 메시지를 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 모델이다.

[0030] 그리고, 상기 메시지 필터링부를 통해 분류된 비스팸 메시지를 저장하기 위한 저장부를 더 포함할 수 있다.

[0031] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법은, 메시지 서버로부터 수신한 한국어 단문 메시지를 표시하는 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법으로서, (a) 상기 메시지 서버로부터 수신한 한국어 단문 메시지를 형태소 단위의 토큰들로 형성하여 토큰화하는 단계; 및 (b) 상기 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 이용해 스팸 메시지와 비스팸 메시지로 분류하는 단계를 포함한다.

[0032] 또한, 상기 (a) 단계는, 상기 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정하는 정규화 과정을 포함한다.

[0033] 또한, (c) 상기 비스팸 메시지를 표시하는 단계를 더 포함한다.

[0034] 그리고, 상기 (b) 단계는, 스팸과 비스팸으로 분류된 학습 말뭉치를 이용하여 메시지에서 나타나는 철자 오류나 띄어쓰기 오류를 교정하여 정규화하고, 형태소 단위의 토큰들로 형성하여 토큰화 한 후, 정보이득 기반으로 정보이득 값이 높은 토큰들만을 선택하는 동작이 학습된 통계 기반 분류 모델을 이용한다. 이때, 상기 정보이득 값은, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지를 수치로 나타낸 것이다.

[0035] 한편, 전술한 목적을 달성하기 위한 본 발명에 따른 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법을 컴퓨터로 읽을 수 있는 매체에 프로그램으로 기록할 수 있으며, 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법도 컴퓨터로 읽을 수 있는 매체에 프로그램으로 기록할 수 있다.

효 과

[0036] 본 발명에 의하면, 의미를 나타내는 최소 단위인 형태소를 사용함으로써 스팸 필터링에 효과적인 어휘들을 정확히 추출할 수 있는 장점이 있다.

[0037] 또한, 스팸 필터링에 효과적인 어휘들을 추출함에 따라 스팸과 비스팸 메시지를 구분할 시 정확률을 높일 수 있다.

[0038] 또한, 실질형태소와 기능형태소가 조합된 어절에 대해서도 형태소 정보를 이용함으로써 스팸 분류 시 생길 수 있는 미등록어(Out of Vocabulary)를 줄일 수 있다.

[0039] 그리고, 사용자가 직접 스팸 필터링을 위한 특정 단어를 직접 등록할 필요 없이 스팸 학습 말뭉치와 비스팸 학습 말뭉치를 이용하여 스팸 필터링에 도움이 되는 어휘들을 대량으로 추출함으로써 사용자의 편의성을 대폭 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

[0040] 본 발명의 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하 상세한 설명에 의해 보다 명확하게 이해될 것이다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 실시예를 상세하게 설명한다.

[0041] 본 발명은 SMS 스팸 필터링 과정 중에서도 SMS 메시지를 의미적으로 더 작은 단위로 표현하는 토큰화(예를 들어 문장을 단어로 나누는 과정을 들 수 있음) 기술을 적용하고, 형태소 단위의 토큰(토큰화 과정에서 작은 단위로 나뉘어진 어휘)들을 스팸 필터링을 위한 통계 기반 분류 모델을 학습시키는데 사용한다.

[0042] 이에, 본 발명은 SMS 스팸 필터링을 통계 기반 문서 분류 관점에서 접근하여 SMS 스팸 필터링 모델의 학습 단계와 실제 SMS 스팸 필터링 단계를 수행한다.

[0043] 이를 위해, 학습 단계에서는 스팸 SMS 메시지와 비스팸 SMS 메시지로 분류된 학습 말뭉치가 필요하다. 또한, 필터링 단계에서는 학습 단계에서 구축된 스팸 필터링 모델에 의해 최종적으로 스팸 SMS 메시지인지 아닌지 여부를 판가름 할 수 있다.

[0044] 따라서, 본 발명은 토큰화 과정에서 SMS 메시지를 형태소 단위로 표현하고, 이를 SMS 스팸 필터링을 위한 자질로써 사용하는 것이 중요하다고 할 수 있다.

[0045] 도 1은 본 발명의 실시예에 따른 한국어 단문 메시지 스팸 필터링 시스템의 한 예를 나타낸 구성도이다.

[0046] 도 1을 참조하면, 본 발명의 하나의 실시예에 따른 한국어 단문 메시지 스팸 필터링 시스템(100)은, 사용자 단말기1(110), 사용자 단말기2(120), 통신망(130), 메시지 서버(140) 및 메시지 데이터베이스(150)를 포함한다.

[0047] 사용자 단말기1(110)은 메시지 발신 단말기(110)로써 사용자에게 의해 SMS 메시지를 작성하여 사용자 단말기2(120)로 전송을 요청하는 단말기이다.

[0048] 여기서, 사용자 단말기1(110) 및 사용자 단말기2(120)는 SMS 메시지를 송신 또는 수신할 수 있는 핸드폰이나, PDA 등의 이동 단말기일 수 있으며, SMS 메시지를 송수신할 수 있는 컴퓨터 등의 유선 단말기일 수 있다.

[0049] 사용자 단말기2(120)는 메시지 수신 단말기(120)로써 메시지 서버(140)에 접속하여, 자신에게 수신되는 한국어 단문 메시지에 대해 스팸 메시지를 필터링한 비스팸 메시지만을 전송해 주도록 설정할 수 있다.

[0050] 따라서, 메시지 수신 단말기(120)는 메시지 서버(140)에 스팸 메시지 필터링 서비스를 등록 설정한 이후, 메시지 서버(140)를 통해 자신에게 전달되는 한국어 단문 메시지 중 스팸 메시지가 필터링된 비스팸 메시지만을 수신할 수 있다.

[0051] 통신망(130)은 사용자 단말기1(110) 및 사용자 단말기2(120)가 무선 단말기일 경우에 이동 통신망이나 근거리 통신망 등의 무선 통신망일 수 있고, 유선 단말기일 경우에 인터넷 등의 유선 통신망일 수 있다.

[0052] 메시지 서버(140)는 사용자 단말기1(110) 또는 사용자 단말기2(120)가 작성한 SMS 메시지를 수신처로 전송할 때 전송할 SMS 메시지의 내용에 대해 정규화하고 형태소 단위로 토큰화하여 스팸 필터링을 수행한 후 비스팸 메시지만을 사용자 단말기2(120) 또는 사용자 단말기1(110)로 전송해 준다.

[0053] 이때, 메시지 서버(140)는 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 자질들을 추출하고, 추출한 자질들에 근거해 통계 기반 분류 모델을 학습시킨다. 또한, 메시지 서버(140)는, 토큰화 과정에서 추출한 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택하는 정보이득 기반 자질 선택 과정을 통해 통계 기반 분류 모델을

학습시킨다.

- [0054] 또한, 메시지 서버(140)는 사용자 단말기1(110) 또는 사용자 단말기2(120) 등 외부로부터 수신된 한국어 단문 메시지에 대해 정규화 과정과 토큰화 과정을 통해 형태소 단위의 토큰화 메시지로 전환하고, 토큰화 메시지로부터 통계 기반 분류 모델을 통해 스팸 메시지를 필터링한다.
- [0055] 즉, 메시지 서버(140)는, 외부로부터 수신된 한국어 단문 메시지에 대해스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여, 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하여 정규화하고, 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로 형태소 단위의 토큰들로 구성된 토큰화 메시지로 전환한다. 이어, 메시지 서버(140)는 학습된 통계 기반 분류 모델을 이용해 형태소 단위의 토큰들을 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링한다.
- [0056] 메시지 데이터베이스(150)는 사용자 단말기1(110) 또는 사용자 단말기2(120)로부터 수신한 메시지 데이터를 저장하거나, 수신처로 전송할 SMS 메시지에 대해 스팸 메시지를 필터링하여 얻은 비스팸 메시지를 저장할 수 있다.
- [0057] 즉, 전술한 도 1의 구성에 따른 한국어 단문 메시지 스팸 필터링 시스템(100)은, 사용자 단말기1(110) 또는 사용자 단말기2(120)가 스팸 필터링 요청한 한국어 단문 메시지에 대해 메시지 서버(140)가 통계 기반 문서 분류 관점으로 스팸 필터링 모델의 학습 단계와 실제 스팸 필터링 단계를 통해 스팸 필터링을 수행하는 것을 나타낸다.
- [0058] 도 2는 본 발명의 실시예에 따라 한국어 단문 메시지 스팸 필터링을 수행하는 메시지 서버의 블록 구성을 개략적으로 나타낸 구성도이다.
- [0059] 도 2를 참조하면, 본 발명에 따른 메시지 서버(140)는, 통신부(210), 필터링 학습분류모델 저장부(220), 스팸 필터링부(230), 메시지 생성부(240), 메시지 전송부(250/240), 데이터베이스 처리부(260/250) 및 제어부(270/260)를 포함한다.
- [0060] 통신부(210)는 통신망(130)을 통해 사용자 단말기1(110) 또는 사용자 단말기2(120)와 통신한다.
- [0061] 필터링 학습분류모델 저장부(220)는 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해 자질들을 추출하고, 추출한 자질들에 근거해학습된 통계 기반 분류 모델이 저장되어 있 생성한다.
- [0062] 스팸 필터링부(230)는 사용자 단말기1(110) 또는 사용자 단말기2(120) 등 외부로부터 수신된 한국어 단문 메시지에 대해 정규화 과정과 토큰화 과정을 통해 형태소 단위의 토큰화 메시지로 전환하고, 토큰화 메시지로부터 통계 기반 분류 모델을 통해 스팸 메시지를 필터링한다.
- [0063] 여기서, 정규화 과정은, 학습 말뭉치 또는 한국어 단문 메시지에서 자주 나타나는 철자 오류나 띄어쓰기 오류를 자동으로 교정한다. 이때, 정규화 과정에서 철자 오류의 교정은 규칙 기반 방식으로 동작하는 철자오류 교정 모델을 이용하며, 띄어쓰기 오류의 교정은 히든 마르코브(Hidden Markov) 모델을 이용한다.
- [0064] 또한, 토큰화 과정은, 학습 말뭉치 또는 한국어 단문 메시지를 구성하는 텍스트에서 형태소를 추출하여 토큰으로 설정하여, 의미적으로인 작은 단위인 토큰 단위형태소 단위의 토큰들로 표현한다.
- [0065] 또한, 정보이득 기반 자질 선택 과정은, 토큰화 과정에서 추출한 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택한다.
- [0066] 그리고, 통계 기반 분류 모델은, 나이브 베이지언(Naive Bayesian) 분류기, 지지 벡터 머신(Support Vector Machine), 최대 엔트로피 모델(Maximum Entropy Model) 중 하나 이상의 모델로 구성될 수 있을 포함한다.
- [0067] 메시지 생성부(240)는 정규화 과정과 토큰화 과정을 거친 한국어 메시지를 토큰화 메시지로 전환하여 생성하거나, 토큰화 메시지로부터 통계 기반 분류 모델을 통해 스팸 메시지가 필터링된 비스팸 메시지를 생성한다.
- [0068] 메시지 전송부(250/240)는 메시지 생성부(240)를 통해 생성된 비스팸 메시지를 사용자 단말기1(110) 또는 사용자 단말기2(120)에게 전송해 준다.

- [0069] 데이터베이스 처리부(260250)는 메시지 데이터베이스(150)의 데이터 입출력을 처리한다.
- [0070] 제어부(270260)는 사용자 단말기1(110) 또는 사용자 단말기2(120)로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 비스팸 메시지가 수신측 단말기로 전송되도록 필터링 학습부(220), 스팸 필터링부(230), 메시지 생성부(240), 메시지 전송부(250) 및 데이터베이스 처리부(260)의 동작을 제어한다.
- [0071] 도 3은 본 발명의 다른 실시예에 따른 한국어 단문 메시지 스팸 필터링을 수행하는 사용자 단말기의 블록 구성을 나타낸 구성도이다.
- [0072] 도 3을 참조하면, 본 발명의 다른 실시예에 따른 사용자 단말기(110)는, 스팸 메시지가 필터링되지 않은 한국어 단문 메시지를 수신하여 스팸 메시지를 필터링하는 단말기로서, 통신부(310), 통계기반 분류모델 생성저장부(320), 메시지 필터링부(330), 저장부(340), 표시부(350) 및 제어부(360)를 포함한다.
- [0073] 통신부(310)는 통신망(130)을 통해 다른 단말기나 메시지 서버(140)와 통신한다.
- [0074] 통계기반 분류모델 생성저장부(320)는 학습 말뭉치에 대하여를 이용해 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정을 통해으로 자질들을 추출하고, 추출한 자질들에 근거해 학습된 통계기반 분류모델을 생성한 저장하고 있다. 이때, 생성된 통계기반 분류모델은 메시지 서버(140)로부터 수신된 한국어 단문 메시지로부터 스팸 메시지를 필터링하는데 이용된다.
- [0075] 또한여기서, 통계기반 분류모델은 생성부(320)는 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여, 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하여는 정규화 과정과하고, 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로 형태소 단위인 토큰 단위의 토큰화 메시지로 전환들로 표현하는 토큰화 과정을 수행한다.
- [0076] 또한, 통계기반 분류모델 생성부(320)은는 토큰화 메시지로 전환형태소 단위의 토큰들로 표현하는 과정에서 추출한 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택하는 정보이득 기반 자질 선택 과정을 통해 통계 기반 분류 모델을 생성수행한다.
- [0077] 메시지 필터링부(330)는 메시지 서버(140)로부터 수신된 한국어 단문 메시지에 대해 정규화 과정과 토큰화 과정을 통해 형태소 단위의 토큰화 메시지로 전환하고, 통계 기반 분류 모델을 통해 토큰화 메시지를 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링한다.
- [0078] 저장부(340)는 메시지 필터링부(330)를 통해 스팸 메시지가 필터링된 비스팸 메시지를 저장한다.
- [0079] 표시부(350)는 한국어 단문 메시지에서 스팸 메시지가 필터링된 비스팸 메시지를 디스플레이한다.
- [0080] 제어부(360)는 메시지 서버(140)로부터 수신된 한국어 단문 메시지에서 스팸 메시지를 필터링하여 비스팸 메시지만을 저장하거나 표시하도록 통계기반 분류모델 생성부(320), 메시지 필터링부(330), 저장부(340) 및 표시부(350)의 동작을 제어한다.
- [0081] 도 4는 본 발명의 실시예에 따른 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한 흐름도이다.
- [0082] 도 4를 참조하면, 사용자 단말기1(110)은 사용자에게 의해 한국어 단문 메시지를 작성하여(S402), 메시지 서버(140)로 사용자 단말기2(120)에게 전송하도록 요청한다(S404).
- [0083] 이때, 사용자 단말기1(110) 및 사용자 단말기2(120)는 모두 스팸 필터링 서비스에 가입되어 있는 경우로 가정하여 설명한다.
- [0084] 이에 따라, 메시지 서버(140)는 사용자 단말기1(110)로부터 수신한 한국어 단문 메시지를 정규화하고, 정규화된 한국어 단문 메시지를 구성하는의 텍스트에서 형태소를 추출해 토큰으로 설정하여고 토큰형태소 단위의 토큰들로 형성하여 토큰화 메시지로 전환한다(S406).
- [0085] 이어, 메시지 서버(140)는 형태소 단위의 토큰들에 대해 통계 기반 분류 모델을 통해 토큰화 메시지를 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링한 후 비스팸 메시지를 생성한다(S408).
- [0086] 여기서, 메시지 서버(140)는 스팸 메시지와 비스팸한국어 단문 메시지로 분류된 학습 말뭉치에 대하여, 자주 나

타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하여 정규화 과정과, 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로 형태소 단위인 토큰 단위의 토큰화 메시지로 전환하는 토큰화 과정을 수행한다.

- [0087] 또한, 메시지 서버(140)는 토큰화 메시지로 전환하는 과정에서 추출한 형태소 단위의 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택하여 정보이득 기반 자질 선택 과정을 통해 통계 기반 분류 모델을 생성스팸 메시지를 분류한다.
- [0088] 메시지 서버(140)는 한국어 단문 메시지에서 스팸 메시지를 필터링한 비스팸 메시지를 수신측인 사용자 단말기 2(120)로 전송한다(S410).
- [0089] 도 5는 본 발명의 실시예에 따른 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한 동작 흐름도이다.
- [0090] 도 5를 참조하면, 본 발명에 따른 메시지 서버(140)는, 필터링 학습부(220)먼저 를 통해 학습 말뭉치에 대하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정(S502).
- [0091] 이후, 사용자 단말기1(110)로부터 통신부(210)를 통해 한국어 단문 메시지가 수신되면(S504S502), 메시지 서버(140)는 스팸 필터링부(230)를 통해 사용자 단말기1(110)로부터 수신된 한국어 단문 메시지에 대해 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하는 정규화 과정을 수행한다(S506S504).
- [0092] 이렇게 메시지 서버(140)가 정규화 과정을 수행하는 것은 다음 토큰화 과정에서 사용하는 형태소/품사 부착기의 성능을 높이기 위함이다.
- [0093] 이때, 메시지 서버(140)는, 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치를 이용해 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정으로 학습된 통계 기반 분류 모델을 분류모델 저장부(220)에 저장하고 있다.
- [0094] 또한이어, 메시지 서버(140)는 스팸 필터링부(230)를 통해, 정규화된 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로 정규화된 메시지를 형태소 단위인 토큰 단위의 토큰들로 형성화 메시지로 전환하는 토큰화 과정을 수행한다(S508S506). 이때, 스팸 필터링부(230)는 정규화된 메시지를 형태소 단위의 토큰들로 형성하는 형태소/품사 부착기의 기능을 수행하는 것이다.
- [0095] 이어, 메시지 서버(140)는 토큰화 과정을 통해 전환형성된 형태소 단위의 토큰화 메시지들에 대해, 스팸 필터링부(230)에 의해 통계 기반 분류 모델을 통이용해 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링한다(S5108).
- [0096] 메시지 서버(140)는 스팸 필터링부(230)에 의해 스팸 메시지가 필터링되면, 비 메시지 생성부(240)를 통해 한국어 메시지에서 스팸 메시지가 필터링된 비스팸 메시지를 생성한다(S512).
- [0097] 메시지 서버(140)는 생성한 비스팸 메시지를 메시지 전송부(250)를 통해 사용자 단말기2(120)에게 전송한다(S514S510).
- [0098] 따라서, 수신측 사용자 단말기2(120)는 스팸 메시지가 필터링되고 한국어로 된 비스팸 메시지를 수신할 수 있게 된다.
- [0099] 도 6은 본 발명의 다른 실시예에 따른 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한 동작 흐름도이다.
- [0100] 도 6을 참조하면, 본 발명에 따른 사용자 단말기(110)는, 먼저 통계기반 분류모델을 생성부(320)를 통해 학습 말뭉치에 대하여를 이용하여 정규화 과정과 토큰화 과정, 정보이득 기반 자질 선택 과정 등으로 학습시킨 통계 자질들을 추출하고, 추출한 자질들에 근거해 통계기반 분류모델을 생성한다(S602).
- [0101] 여기서, 통계기반 분류모델 생성부(320)는은 스팸 메시지와 비스팸 메시지로 분류된 학습 말뭉치에 대하여, 자

주 나타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하여 정규화하고, 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로인 형태소 단위인 토큰 단위의 토큰들로 형성하여 토큰화 메시지로 전환한다.

[0102] 예를 들어, 한국어 단문 메시지가 "대리운전 필요하면 언제든지 연락주세요"인 경우, 통계기반 분류모델 생성부(320)가 '어절' 단위로 토큰화를 수행하면, "대리운전", "필요하면", "언제든지", "연락주세요"의 토큰들을 생성하는 것이다.

[0103] 이때, 통계기반 분류모델은 생성부(320)는 형태소/품사 부착기를 이용하여 메시지를 구성하는 텍스트에서 형태소를 추출한다. 따라서, 위의 예제 문장을 '형태소' 단위로 토큰화 하면, "대리", "운전", "필요", "하면", "하", "면", "언제", "이", "든지", "연락", "주", "시", "어요"의 토큰들이 생성된다.

[0104] 이렇게 형태소를 토큰으로 사용함에 따라, 메시지의 내용을 어절 단위보다 의미적으로 더 작은 단위인 형태소 단위로 표현함으로써, 길이 제한으로 인한 어휘 정보 부족 문제를 완화할 수 있다. 또한 교착어 특징을 고려하여 스팸 메시지와 비스팸 메시지를 구별하는데 도움을 주는 어휘를 정확하게 추출해 낼 수 있다.

[0105] 다음 표 1은 어절과 형태소의 한 예를 나타낸 것이다.

표 1

어절	형태소
연락주세요	연락+주+시+어요
연락줄래	연락+주+ㄹ래
연락주실거죠	연락+주+시+ㄹ+거+이+죠
연락줘	연락+주+어

[0107] 표 1에서, "연락"이란 어휘가 스팸 메시지와 비스팸 메시지를 구별하는데 중요하다고 가정하면, 좌측 '어절' 열에서는 "연락"을 정확히 추출할 수 없는 문제점이 있지만, 우측 '형태소' 열에서는 "연락"을 정확히 추출할 수 있다.

[0108] 이때, 형태소 단위의 토큰들은 스팸 필터링 시 미등록되어(Out of Vocabulary)를 줄여주는 기능도 수행한다. 어절은 실질형태소와 기능형태소의 조합으로 수 많은 어휘 생성이 가능한 반면, 형태소는 의미를 나타내는 최소 단위이기 때문에 이러한 어휘 조합 자체가 존재하지 않는다.

[0109] 또한, 통계기반 분류모델 생성부(320)는 토큰화 메시지로 전환하는 과정에서 추출한 토큰들에 대해, 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지에 관한 정보 이득 값을 산출하여 상기 정보 이득 값이 기준치 이상으로 높은 토큰들만을 자질로 선택하는 정보이득 기반 자질 선택 과정을 통해 통계 기반 분류 모델을 생성한다.

[0110] 이때, 토큰화 과정에서 추출한 토큰 중 어떤 토큰들이 스팸 필터링에 도움이 되는지를 평가하기 위해 정보이득(Information Gain) 기반 자질 선택을 거치는 것이다. 여기서, 정보이득은 어떤 토큰이 스팸 메시지 또는 비스팸 메시지에 대한 정보를 얼마나 포함하고 있는지를 나타낸다. 토큰들 중에는 스팸 필터링에 도움이 되지 않는 것들이 다수로 존재하기 때문에, 스팸 필터링 모델 학습 시 정보이득 값이 높은 상위 토큰들만을 자질로써 사용한다. 사용하는 토큰의 수는 학습 말뭉치의 크기나 분류 작업 등에 따라 달라질 수 있다. 이렇게 추출된 자질들을 스팸 필터링을 위한 통계기반 분류모델 학습에 이용한다.

[0111] 따라서, 통계기반 분류모델은 메시지 서버(140)로부터 수신된 한국어 단문 메시지로부터 스팸 메시지를 필터링하는데 이용된다.

[0112] 이렇게 통계기반 분류모델이 학습되어 분류모델 저장부(320)에 저장되어 있는 상태에서, 이어, 사용자 단말기 1(110)은 통신부(310)를 통해 통신망(130)을 통한 메시지 서버(140)로부터 한국어 단문 메시지를 수신되면한다(S604).

[0113] 이어, 사용자 단말기 1(110)은 메시지 필터링부(330)를 통해 한국어 단문 메시지에 대해 자주 나타나는 철자 오류나 띄어쓰기 오류를 교정 모델을 통해 교정하는 정규화 과정을 수행한다(S606).

[0114] 또한, 사용자 단말기 1(140)은 메시지 필터링부(330)를 통해 한국어 단문 메시지를 구성하는 텍스트에서 형태소를 추출해 토큰으로 설정하여, 의미적으로인 형태소 단위인 토큰 단위의 토큰들로화 메시지로 전환형성하는 토

큰화 과정을 수행한다(S608).

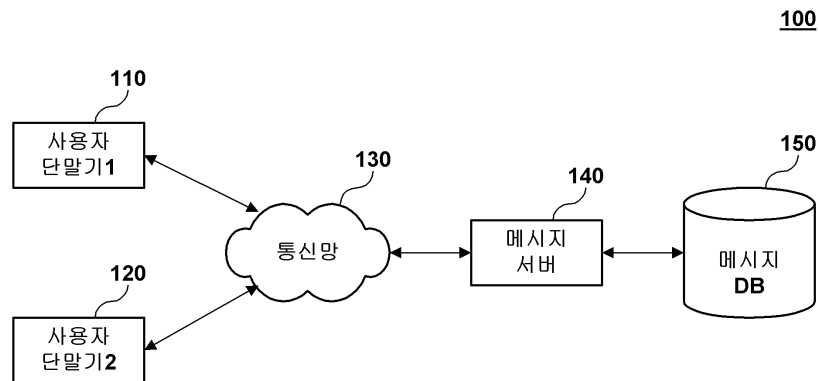
- [0115] 이어, 사용자 단말기1(110)은 형태소 단위의 토큰들토큰화 과정을 통해 전환된 토큰화 메시지에 대해, 메시지 필터링부(330)에 의해 통계 기반 분류 모델을 통해이용하여 스팸 메시지와 비스팸 메시지로 분류하여고 메시지 필터링부(330)를 통해 스팸 메시지를 필터링한다(S610).
- [0116] 사용자 단말기1(110)은 메시지 필터링부(330)에 의해 스팸 메시지가 필터링된되 면, 메시지 생성부(240)를 통해 한국어 메시지에서 스팸 메시지가 필터링된 비스팸 메시지를 생성한다(S612).
- [0117] 그리고, 사용자 단말기1(110)은 비스팸 메시지를 저장부(340)에 저장하거나, 표시부(350)를 통해 디스플레이한다(S614).
- [0118] 한편, 본 발명의 실시예에 따라 형태소 단위 어휘를 이용한 한국어 단문 메시지 스팸 필터링의 우수성을 검증하기 위해, 기존 영어에서 사용하던 SMS 스팸 필터링 방법[Content-based SMS Spam Filtering, Hidalgo, J et al, 2006]을 한국어에 맞게 적용한 Baseline과 비교 실험을 수행하였으며, 도 7에 도시된 바와 같이 토큰 수에 따른 스팸 분류 성능을 얻을 수 있었다. 도 7은 본 발명의 실시예에 따라 사용하는 토큰 수에 따른 스팸 분류 성능을 나타낸 그래프로써, (a)는 Precision을 보여주는 그래프이고, (b)는 Precision과 Recall을 1:1로 고려한 F1-measure을 보여주는 그래프이다.
- [0119] 도 7에서 (a)의 경우, 총 20,000 개의 한국어 메시지를 수집하여 이중 18,000 개를 학습 집합으로, 나머지 2,000 개를 실험 집합으로 사용하여 얻은 결과이다. 이때, 스팸 메시지와 비스팸 메시지의 비율은 9:1로 구성되어 있다. 토큰 갯수 2,000 개를 사용 시 기존 연구에 대비하여 10.6% 정도의 성능이 향상되고 있음을 알 수 있다. (b)의 경우, 토큰 갯수 2,000 개 사용 시 기존 연구Baseline에 대비 5.1% 정도의 성능이 향상되었음을 알 수 있다.
- [0120] 도 8은 본 발명의 실시예에 따라 형태소가 미등록어를 얼마나 줄이는지에 대한 평가 결과를 나타낸 그래프이다. 도 8에서 사용하는 평가 척도인 Unseen rate는 실험 집합에서 미등록어로만 구성된 메시지의 비율을 나타낸다. 따라서 토큰 갯수 2,000 개 사용 시 기존 연구에 대비 77.0% 정도 미등록어로만 구성된 메시지를 줄이는 결과가 있음을 알 수 있다.
- [0121] 이상에서 설명한 바와 같이 본 발명에 의하면, 한국어 단문 메시지에서 형태소를 추출해 형태소 단위의 토큰들로 표현하여 토큰화하고, 토큰화 과정과 정보이득 기반 자질 선택 과정으로 학습된 통계 기반 분류 모델을 이용해 형태소 단위의 토큰들을 스팸 메시지와 비스팸 메시지로 분류하여 스팸 메시지를 필터링하도록 하는, 한국어 단문 메시지 스팸 필터링 시스템 및 방법, 메시지 서버 및 그의 한국어 단문 메시지 스팸 필터링 방법, 사용자 단말기 및 그의 한국어 단문 메시지 스팸 필터링 방법을 실현할 수 있다.
- [0122] 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있으므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 등가개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.
- 도면의 간단한 설명**
- [0123] 도 1은 본 발명의 실시예에 따른 한국어 단문 메시지 스팸 필터링 시스템의 한 예를 나타낸 구성도이다.
- [0124] 도 2는 본 발명의 실시예에 따라 한국어 단문 메시지 스팸 필터링을 수행하는 메시지 서버의 블록 구성을 개략적으로 나타낸 구성도이다.
- [0125] 도 3은 본 발명의 다른 실시예에 따라 한국어 단문 메시지 스팸 필터링을 수행하는 사용자 단말기의 블록 구성을 나타낸 구성도이다.
- [0126] 도 4는 본 발명의 실시예에 따른 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한 흐름도이다.
- [0127] 도 5는 본 발명의 실시예에 따른 메시지 서버의 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한 동작 흐름도이다.
- [0128] 도 6은 본 발명의 다른 실시예에 따른 사용자 단말기의 한국어 단문 메시지 스팸 필터링 방법을 설명하기 위한

동작 흐름도이다.

- [0129] 도 7은 본 발명의 실시예에 따라 사용하는 토큰 수에 따른 스팸 분류 성능을 나타낸 그래프이다.
- [0130] 도 8은 본 발명의 실시예에 따라 형태소가 미등록어를 얼마나 줄이는지에 대한 평가 결과를 나타낸 그래프이다.
- [0131] < 도면의 주요 부분에 대한 부호의 설명 >
- [0132] 100 : 한국어 단문 메시지 스팸 필터링 시스템 110, 120 : 사용자 단말기
- [0133] 130 : 통신망 140 : 메시지 서버
- [0134] 150 : 메시지 데이터베이스 210 : 통신부
- [0135] 220 : 필터링 학습부분류모델 저장부 230 : 스팸 필터링부
- [0136] 240 : 메시지 생성부 250 : 메시지 전송부
- [0137] 250 : 데이터베이스 처리부 260 : 제어부
- [0138] 310 : 통신부 320 : 통계기반 분류모델 저장부
- [0139] 330 : 메시지 필터링부 340 : 저장부
- [0140] 350 : 표시부 360 : 제어부

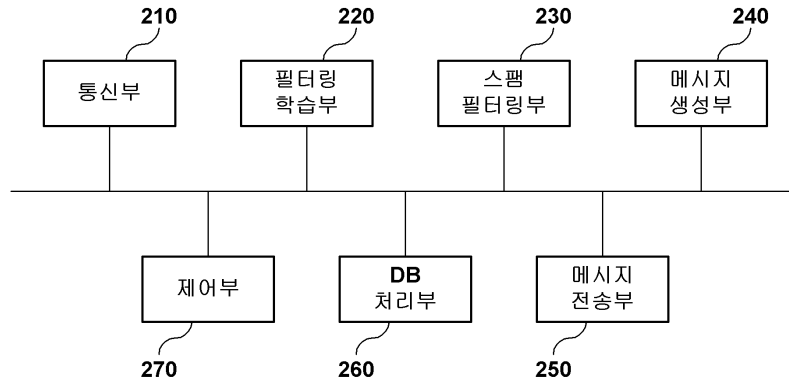
도면

도면1



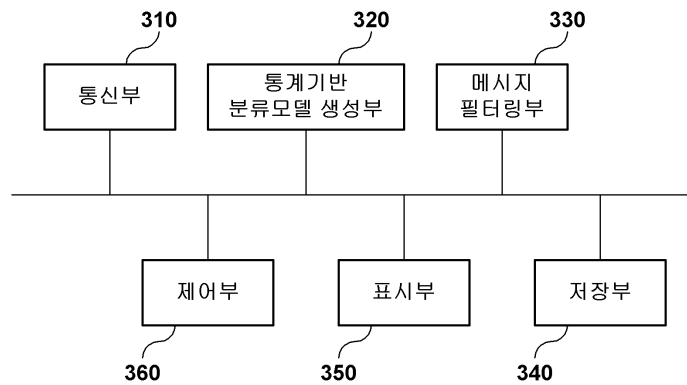
도면2

140

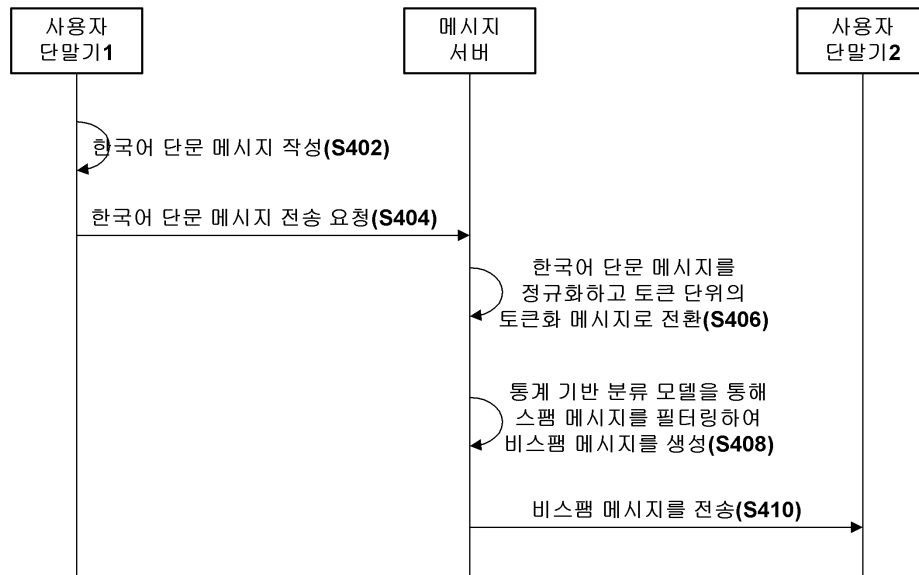


도면3

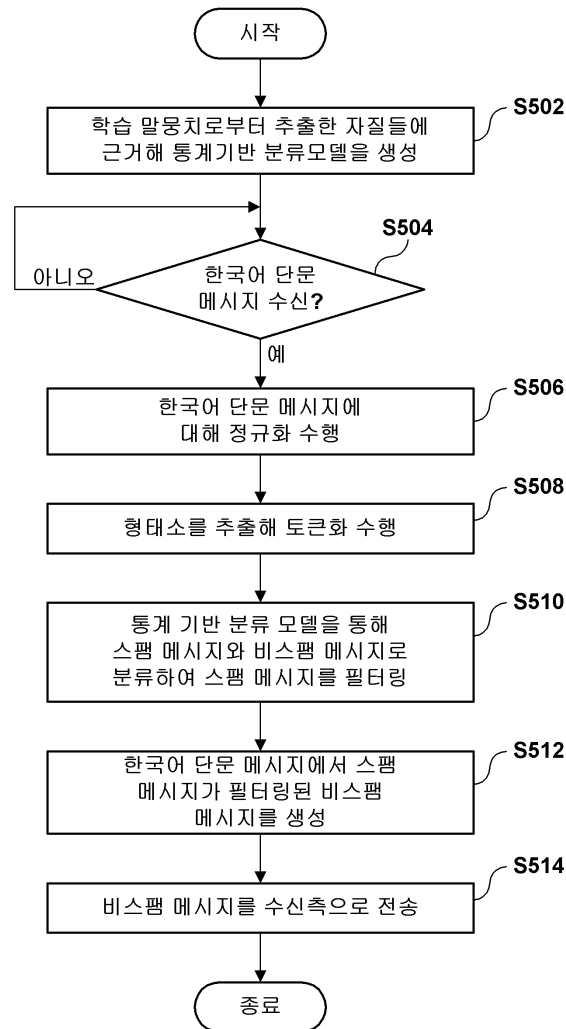
110



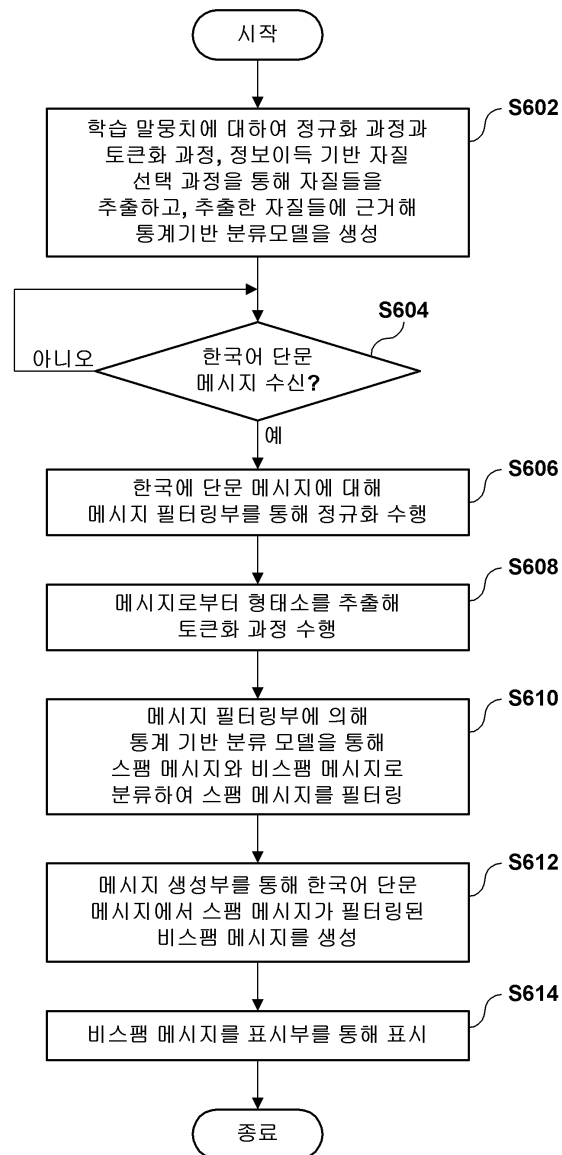
도면4



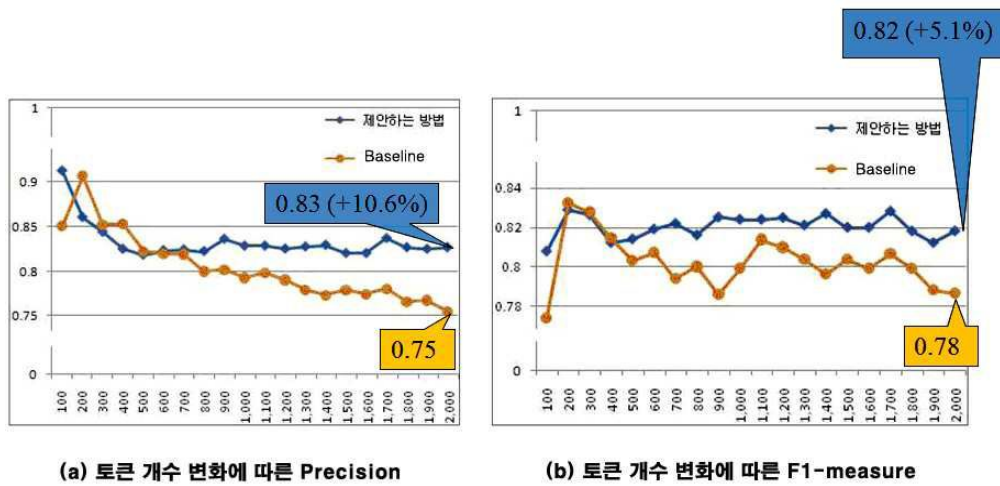
도면5



도면6



도면7



도면8

