

명세서

발명의 명칭: 딥러닝 기반의 자연어 처리를 통한 메시지 내 이체 정보 판단 방법 및 장치

기술분야

- [1] 본 발명은 딥러닝을 통한 메시지 내 이체 정보의 포함 여부를 판단하고, 이체 정보를 분류하는 이체 정보 판단 방법 및 장치에 관한 것이다.

배경기술

- [2] 스마트폰과 모바일 통신 기술의 발달에 힘입어 사용자들은 과거에 오프라인을 통해 수행하던 기능을 언제 어디서나 간편하고 편리하게 이용할 수 있게 되었고, 특히 스마트폰의 금융 어플리케이션을 통해 대부분의 금융 업무를 비대면을 처리할 수 있게 되면서 일상생활에서의 금융 활동에 편의성들이 향상되었다.
- [3] 종래에도 폰 뱅킹이나 인터넷 뱅킹과 같은 서비스들은 존재하였으나, 사용자가 자금의 이체를 위해서는 이체 대상의 계좌번호를 저장하거나 암기하는 것 외에 온라인 상에서 사용자와 주고받은 메시지에 포함된 계좌번호, 은행명 등과 같은 이체 정보를 직접 보고 입력해야만 하였다. 따라서 이러한 번거로움을 피하고자 최근에는 메시지에 포함된 이체 정보나 복사된 이체 정보에 금융 어플리케이션의 접근이 허용된 경우 메시지의 내용으로부터 직접 이체 정보를 추출하여 금융 어플리케이션에 입력되는 기능들이 제공되고 있다.
- [4] 하지만, 기존의 이체 정보 판단 기능은 메시지 내 포함된 문자들 고유의 패턴을 이용하고 있으나, 문자 패턴의 경우 사용자의 특성에 따라 다양한 케이스들이 존재하게 됨으로 이체 정보를 정확히 인식하기 어렵다는 단점이 있다.
- [5] 또한, 이러한 단점을 개선하기 위해 개인화된 특징들을 반영한 다양한 패턴들을 테이블로 정의하는 경우 테이블 자체가 대용량화 되어 스마트폰과 같은 모바일 장치에서는 사용하기 어려워지게 되는 문제가 발생한다.
- [6] 따라서, 메시지 내에서 이체 정보를 정확하게 판별하면서 적은 리소스를 통해 스마트폰에 이용가능한 이체 정보를 판단하는 방법이 요구되고 있다.

발명의 상세한 설명

기술적 과제

- [7] 본 발명은 저용량의 판단 모델을 이용하면서 메시지 내 이체 정보를 보다 정확하게 판단하기 위한 방법을 제안하는 것을 목적으로 한다.
- [8] 보다 구체적으로, 본 발명은 딥러닝 기반의 자연어 처리와, 정규식 기반의 판단 방법을 조합하여 보다 효율적인 이체 정보의 판단 방법을 제안하는 것을 목적으로 한다.
- [9] 또한, 본 발명은 개인화된 메시지의 특징을 반영하고, 개인정보의 이용을 최소화한 이체 정보 판단 모델의 학습 방법을 제안하는 것을 목적으로 한다.

과제 해결 수단

- [10] 상기 기술적 과제를 해결하기 위한 본 발명에 따른 딥러닝 기반의 자연어 처리를 통한 메시지 내 이체 정보를 판단하는 방법은 사용자 단말기 내 획득된 메시지를 기준 단어에 따라 전처리하는 단계; 상기 전처리된 메시지에서 분절된 텍스트 별로 대응되는 임베딩 벡터를 추출하고, 상기 추출된 임베딩 벡터의 가중 연산을 통해 이체 정보의 포함 여부를 판단하는 단계; 및 이체 정보를 포함하는 것으로 판단된 상기 전처리된 메시지 내 이체 정보를 분류하는 단계;를 포함한다.
- [11] 상기 전처리하는 단계는 상기 메시지에 포함된 숫자 또는 특정 단어를 기준으로, 상기 숫자 또는 특정 단어의 위치 전후에 공백을 추가하여 상기 메시지를 분절하는 것이 바람직하다.
- [12] 상기 판단하는 단계는 상기 획득된 메시지 내 이체 정보의 포함 여부를 상기 사용자 단말기 상에서 동작하는 로컬 모델을 통해 판단하되, 상기 로컬 모델은 상기 이체 정보의 포함 여부에 대한 판단 결과를 이용하여 내부 가중 연산을 수행하는 레이어의 학습 정보를 생성하는 것이 바람직하다.
- [13] 상기 로컬 모델은 복수의 사용자 단말기로부터 생성된 학습 정보로 연합 학습하여 배포되는 마스터 모델을 이용하여 갱신(update)되는 것이 바람직하다.
- [14] 상기 판단하는 단계는, 상기 획득된 메시지에 대하여 상기 로컬 모델에서 산출된 이체 정보의 제1 예상 포함 확률 및, 상기 마스터 모델에서 산출된 이체 정보의 제2 예상 포함 확률을 이용하여 이체 정보의 포함 여부를 판단하는 것이 바람직하다.
- [15] 상기 이체 정보를 분류하는 단계는 분류 별 텍스트의 패턴을 문법으로 정의한 정규식(regular expression)으로 상기 메시지 내 이체 정보를 분류하는 것이 바람직하다.
- [16] 상기 이체 정보를 분류하는 단계는 이체 정보를 분류하도록 학습된 신경망 모델에 상기 전처리된 메시지를 입력함으로써 상기 메시지 내 이체 정보를 분류하는 것이 바람직하다.
- [17] 상기 기술적 과제를 해결하기 위한 본 발명에 따른 메시지 내 이체 정보를 판단하는 사용자 단말기 상에서 동작하는 어플리케이션은 사용자 단말기 내 획득된 메시지를 기준 단어에 따라 전처리하는 전처리부; 상기 전처리된 메시지에서 분절된 텍스트 별로 대응되는 임베딩 벡터를 추출하고, 상기 추출된 임베딩 벡터의 가중 연산을 통해 이체 정보의 포함 여부를 판단하는 이체 정보 판단부; 및 이체 정보를 포함하는 것으로 판단된 상기 전처리된 메시지 내 이체 정보를 분류하는 이체 정보 분류부;를 포함한다.
- [18] 상기 전처리부는 상기 메시지에 포함된 숫자 또는 특정 단어를 기준으로, 상기 숫자 또는 특정 단어의 위치 전후에 공백을 추가하여 상기 메시지를 분절하는 것이 바람직하다.
- [19] 상기 이체정보 판단부는 상기 획득된 메시지 내 이체 정보의 포함 여부를 상기 사용자 단말기 상에서 동작하는 로컬 모델을 통해 판단하되, 상기 로컬 모델은

상기 이체 정보의 포함 여부에 대한 판단 결과를 이용하여 내부 가중 연산을 수행하는 레이어의 학습 정보를 생성하는 것이 바람직하다.

- [20] 상기 로컬 모델은 복수의 사용자 단말기로부터 생성된 학습 정보로 연합 학습하여 배포되는 마스터 모델을 이용하여 갱신(update)되는 것이 바람직하다.
- [21] 상기 이체 정보 판단부는, 상기 획득된 메시지에 대하여 상기 로컬 모델에서 산출된 이체 정보의 제1 예상 포함 확률 및, 상기 마스터 모델에서 산출된 이체 정보의 제2 예상 포함 확률을 이용하여 이체 정보의 포함 여부를 판단하는 것이 바람직하다.
- [22] 상기 이체 정보를 분류부는 분류 별 텍스트의 패턴을 문법으로 정의한 정규식(regular expression)으로 상기 메시지 내 이체 정보를 분류하는 것이 바람직하다.
- [23] 상기 이체 정보를 분류부는, 이체 정보를 분류하도록 학습된 신경망 모델에 상기 전처리된 메시지를 입력함으로써 상기 메시지 내 이체 정보를 분류하는 것이 바람직하다.
- [24] 한편, 상술한 목적을 달성하기 위한 본 발명의 일 실시 예에 따르면, 컴퓨터가 판독 가능한 기록 매체에 저장된 이체 정보 판단 방법을 실행하기 위한 프로그램 코드를 포함할 수 있다.

발명의 효과

- [25] 본 발명에 따르면, 저용량의 임베딩 기반의 판단 모델을 이용하여 이체 정보의 포함 여부를 우선적으로 판단함으로써, 메시지 내 이체 정보의 분류 효율을 높일 수 있다.
- [26] 또한, 사용자의 메시지를 기초로 학습된 로컬 기반의 모델을 이용함으로써, 개인화된 특징이 반영된 메시지 내 이체 정보를 보다 정확하게 판단 및 분류할 수 있다.
- [27] 또한, 다양한 사용자의 메시지를 기초로 공통적인 특징을 학습한 모델과 로컬 기반의 모델을 독립적으로 운영함으로써 사용자의 특성에 따른 이체 정보에 대한 적응성을 높임과 동시에 학습을 위한 개인정보의 이용을 최소화할 수 있다.

도면의 간단한 설명

- [28] 도 1은 본 발명의 일 실시예에 따른 이체 정보 판단 방법을 나타낸 흐름도이다.
- [29] 도 2는 본 발명의 일 실시예에 따른 이체 정보의 포함 여부를 판단하는 판단 모델의 구조를 나타내는 예시도이다.
- [30] 도 3은 본 발명의 일 실시예에 따른 로컬 기반의 판단 모델과 마스터 모델을 통한 연합 학습 과정을 나타낸 예시도이다.
- [31] 도 4는 본 발명의 일 실시예에 따른 이체 정보 판단 방법의 이체 정보 분류를 위한 정규식 기반의 분류 모델을 나타낸 예시도이다.
- [32] 도 5는 본 발명의 일 실시예에 따른 이체 정보 판단 방법의 이체 정보 분류를 위한 신경망 모델을 나타낸 예시도이다.

[33] 도 6은 본 발명의 다른 실시예에 따른 이체 정보 판단 방법의 이체 정보 분류를 위한 신경망 모델을 나타낸 예시도이다.

[34] 도 7은 본 발명의 일 실시예에 따른 메시지 내 이체 정보를 판단하는 사용자 단말기 상에서 동작하는 어플리케이션의 구성을 나타낸 블록도이다.

발명의 실시를 위한 형태

[35] 이하의 내용은 단지 발명의 원리를 예시한다. 그러므로 당업자는 비록 본 명세서에 명확히 설명되거나 도시 되지 않았지만 발명의 원리를 구현하고 발명의 개념과 범위에 포함된 다양한 장치를 발명할 수 있는 것이다. 또한, 본 명세서에 열거된 모든 조건부 용어 및 실시 예들은 원칙적으로, 발명의 개념이 이해되도록 하기 위한 목적으로만 명백히 의도되고, 이와같이 특별히 열거된 실시 예들 및 상태들에 제한적이지 않는 것으로 이해되어야 한다.

[36] 상술한 목적, 특징 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해질 것이며, 그에 따라 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다.

[37] 또한, 발명을 설명함에 있어서 발명과 관련된 공지 기술에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에 그 상세한 설명을 생략하기로 한다. 이하에는 첨부한 도면을 참조하여 본 발명의 바람직한 실시 예에 대해 상세하게 설명한다.

[38] 도 1은 본 발명의 일 실시예에 따른 이체 정보 판단 방법을 나타낸 흐름도이다.

[39] 도 1을 참조하면, 수신되는 메시지에서 이체 정보를 판단하기 위해 사용자 단말기(100)는 획득된 메시지에 포함된 텍스트 정보를 처리 단위에 따라 구분하고자 전처리할 수 있다(S100).

[40] 본 실시예에서 획득되는 메시지는 통신 망을 이용하여 수신되는 일반적인 문자 메시지, 전자메일의 형태 또는 스마트폰에 설치된 다양한 메신저 어플리케이션(예를 들어 카카오톡 메신저 등)을 통해 수신된 메시지를 모두 포괄한다.

[41] 또는, 외부로부터 수신 여부와 무관하게 사용자 단말기에 저장된 다양한 텍스트 정보 중 사용자의 복사 요청에 따라 선택된 전체 또는 일부의 텍스트 정보일 수 있다.

[42] 따라서, 본 실시예에서는 사용자 단말기(100) 내 수신된 다양한 형태의 데이터 중 사용자의 선택에 의해 복사 요청된 텍스트로서, 클립보드(clipboard)에 복사를 위해 임시 저장된 데이터를 전처리 할 수 있다.

[43] 이 외에, 획득된 메시지는 메시지를 스크린 캡처한 이미지 형태로 존재할 수 있으며, 이미지의 경우 이미지 내 포함된 텍스트를 인식하고 인식된 텍스트를 문자 인식 등을 통해 추출하는 것으로 이용될 수 있다.

[44] 이상 본 실시예에서 사용자 단말기(100)는 텍스트를 포함하는 다양한 형태에 따른 데이터를 기준 단어에 따라 전처리한다.

- [45] 구체적으로 메시지에 포함된 숫자 및 특정 단어를 기준으로 앞뒤에 공백을 추가하여 해당 메시지를 분절하는 것으로 획득된 메시지를 전처리할 수 있다. 전처리된 메시지 내 텍스트를 공백을 단위로 분절함으로써 메시지에 포함된 의미를 효율적으로 인식할 수 있다.
- [46] 본 실시예에서 분절의 기준이 되는 특정 단어로 계좌 이체의 목적에 따라 만, 천, 백 등과 같이 숫자의 크기를 나타내는 단어와, '원' 과 같이 금액을 나타내는 단어를 기준으로 설정할 수 있다. 또한 국가별 통화의 표기 방식에 따라 예를 들어 달러, 엔, 유로 등의 단어를 포함할 수 있다.
- [47] 따라서, 획득된 메시지가 “홍길동인데 30만원을 카카오뱅크 123-456-7890로 보내주세요”인 경우, 사용자 단말기(100)는 숫자 및 특정 단어를 기준으로 앞뒤에 공백을 추가하여 “홍길동인데 30 만 원 을 카카오뱅크 123 - 456 - 7890 로 보내주세요”로 분절하는 것으로 해당 메시지를 전처리할 수 있다.
- [48] 다음, 사용자 단말기(100)는 전처리된 메시지를 임베딩(Embedding) 기반의 판단 모델에 입력하고 이체 정보의 예상 포함 확률을 산출함으로써 이체 정보의 포함 여부를 판단할 수 있다(S200).
- [49] 이와 관련하여 도 2를 참조하여 설명한다.
- [50] 도 2는 본 발명의 일 실시예에 따른 이체정보 예상 포함 확률을 산출하는 판단 모델의 구성을 나타내는 도이다.
- [51] 도 2를 참조하면, 본 실시예에 따른 임베딩 기반의 판단 모델은 단어 룩업 테이블(word look-up table)을 통해 전처리된 메시지의 분절된 텍스트 각각에 대응되는 인덱스 값을 추출할 수 있다.
- [52] 단어 룩업 테이블은 기 획득한 사용자의 메시지들을 기초로 문장을 상술한 분절 기준에 따라 단어 텍스트 단위로 분리하는 토큰화(Tokenization)를 수행하고, 중복을 제거함으로써 빈도수를 고려하여 단어 집합(vocabulary)의 형태로 생성될 수 있다.
- [53] 단어 집합은 각 단어 별로 숫자 인덱스와 같은 ID를 부여하는 정수 인코딩(Integer encoding)을 통해 생성될 수 있다.
- [54] 다음, 사용자 단말기(100)는 임베딩 테이블을 참조하여 인덱스 값에 대응되는 임베딩 벡터(Embedding vector)들을 추출하고 추출된 임베딩 벡터의 평균값을 산출할 수 있다.
- [55] 임베딩 테이블은 문장에 포함된 단어들을 원-핫 인코딩 또는 설정한 차원에 따른 밀집 벡터(dense vector)의 형태로 표현한 것으로 워드 임베딩(Word Embedding)을 통해 생성될 수 있다. 이때 임베딩 테이블은 일반적인 신경망의 학습과 같은 방식으로 입력된 단어를 표현하는 벡터들을 학습하여 생성할 수 있다.
- [56] 즉, 본 실시예에서는 전처리를 통해 분절된 단어들에 대해 단어 룩업 테이블을 참조함으로써 인덱스 값을 추출하고, 인덱스 값에 대응하여 메시지에 포함된 정보들을 설정한 차원에 따른 특징 공간에 매핑함으로써 정규화한 임베딩

벡터를 추출할 수 있다.

- [57] 다음, 추출된 임베딩 벡터의 평균을 산출하고, 산출된 평균을 통해서 메시지에 포함된 의미가 이체정보를 포함하는지 여부를 판단한다.
- [58] 구체적으로, 사용자 단말기(100)는 임베딩 벡터의 평균값을 선형 레이어에 입력하여 예상 포함 확률을 산출한다. 선형 연산을 위한 선형 레이어는 임베딩 벡터와 차원이 같은 미리 정의된 가중치 W 와 바이어스(bias) b 를 커널로 이용하여 평균 벡터를 가중합(weighted sum) 한다.
- [59] 예를 들어, 임베딩 벡터의 평균값(x)은 $\{0.13, 0.52, 0.333, 0.446, 0.356, 0.335\}$, 가중치(w)는 $\{0.123, -0.44; 0.67, 0.34; 0.33, 0.82; -0.34, 0.66; -0.33, -0.45; 0.55, 0.89\}$, 바이어스(b)는 $\{0.5, -0.52\}$ 인 경우, 사용자 단말기(100)는 예상포함확률($c = W \cdot X + b$)을 $\{10.3, -0.412\}$ 로 산출할 수 있으며, 해당 메시지에는 이체 정보가 포함되었을 확률이 10.3으로 불포함 확률보다 크므로 이체 정보를 포함한 것으로 판단할 수 있다.
- [60] 이상의 과정을 통해 본 실시예에서는 사용자 단말기(100) 상의 임베딩 기반의 판단 모델을 통하여 입력된 메시지 상에 이체 정보가 포함되었는지 여부를 로컬 차원에서 우선적으로 판단할 수 있다.
- [61] 이때, 로컬 모델은 임베딩 기반의 판단 모델의 결과를 이용하여 직접 학습을 수행하는 것도 가능하다. 즉, 학습을 통해 가중합 연산에 이용되는 가중치 W 와 바이어스 b 를 갱신함으로써 사용자 고유의 특징에 따른 이체 정보를 더욱 잘 판단할 수 있다
- [62] 나아가, 본 실시예에 따른 판단 모델은 로컬 차원의 학습 결과를 수집하고, 수집된 결과를 이용하여 전체 판단 모델의 성능을 향상시키는 것도 가능하다.
- [63] 다만, 판단 모델의 학습을 위해서 이용되는 메시지는 개인의 계좌 정보를 포함한 개인 정보가 포함될 수 있으므로, 메시지와 판단 결과를 수집하는 대신 실제 판단에 이용되는 로컬 차원의 판단 모델로부터 생성된 학습 결과를 중앙의 서버가 수집할 수 있다.
- [64] 서버는 수집된 학습 결과를 이용하여 마스터 모델을 갱신함으로써 개인정보가 수집 및 학습에 직접 이용되는 것을 방지한다.
- [65] 따라서, 본 실시예에 따른 임베딩 기반의 판단 모델은 로컬 기반의 판단 모델(이하, 로컬 모델이라 함)과 중앙의 서버에서 학습을 위해 관리되는 마스터 모델로 구분하여 운영될 수 있다.
- [66] 본 실시예에 따른 로컬 모델 및 마스터 모델은 모두 상술한 도 2에 따른 구조와 같이 인덱스와 임베딩 벡터(Embedding vector)들이 매칭된 단어 집합(Vocabulary)과 인덱스에 따른 임베딩 벡터 값들이 정의된 임베딩 테이블로 구성될 수 있으며, 임베딩 테이블을 통해 생성된 임베딩 벡터들은 선형 레이어를 통해 가중 연산되어 확률값을 출력한다.
- [67] 따라서, 로컬 기반의 판단 모델과 중앙의 마스터 모델은 대응되는 구조를 통하여 로컬에서 생성된 학습 정보를 이용하여 중앙의 마스터 모델을 연함

학습한다.

- [68]
- [69] 이에 대해서는 도 3을 참조하여 본 실시예에 따른 학습 과정에 대하여 상세히 설명한다.
- [70] 도 3은 본 발명의 일 실시예에 따른 로컬 모델과 마스터 모델을 통한 연합 학습 과정을 나타낸 예시도이다.
- [71] 도 3을 참조하면, 마스터 모델을 관리하는 관리 서버(200)는 사용자의 메시지들을 기초로 생성된 로컬 모델의 학습 결과를 이용하여 마스터 모델(20)을 학습시킨다.
- [72] 사용자 단말기들(100-1, 100-2, ... 100-N)은 관리 서버(200)로부터 미리 학습하여 배포된 마스터 모델(20)을 수신하고 내부의 프로세서를 통해 동작 가능한 형태로 설치함으로써 사용자 단말기 자체에서 동작하는 로컬 모델의 형태로 업무를 수행할 수 있다.
- [73] 즉 각 사용자 단말기(100)에 설치된 로컬 모델들(10-1, 10-2, ... 10-N)은 외부 서버와의 통신 없이 수신된 메시지의 이체 정보 여부를 직접 판단할 수 있으며, 판단 결과를 이용하여 재 학습될 수 있다.
- [74] 구체적으로 학습은 로컬 모델의 예측 값과 실제 값(Ground Truth) 간의 오차를 줄이는 방향으로 수행되며, 학습을 위해서 이용되는 레이블링 값으로 사용자가 복사한 메시지와 메시지를 복사한 직후 금융 어플리케이션의 실행 여부를 이용할 수 있다.
- [75] 로컬 모델 내 신경망은 실제 값(예를 들어 금융 어플리케이션의 실행의 경우 1, 그렇지 않은 경우 0)과 예측 값 간의 차이를 오차로 정의하고 오차가 최소가 되도록 내부의 레이어 들의 가중치들의 갱신 값을 산출할 수 있다.
- [76] 이때, 각 사용자 단말기(100)에 배포되는 마스터 모델은 로컬 단에서 일어나는 학습 과정을 반영함으로써 다양한 환경에 따른 이체 정보의 특징 들을 학습할 수 있으며 개인화된 메시지 작성 방식에 대해서도 적응적인 성능을 가질 수 있다.
- [77] 구체적으로, 관리 서버(200)는 각 사용자 단말기(100)에서 수신된 메시지와 메시지의 이체 정보 포함 여부를 정의하는 레이블링 값들을 수집하는 대신, 각 로컬 모델에서의 갱신 값을 수신하여 학습에 이용한다.
- [78] 따라서, 관리 서버(200)는 복수의 사용자 단말기들(100-1, 100-2, ... 100-N)에서 학습 결과로써 로컬 모델 각각에서 생성된 갱신된 가중치들($W_1, W_2, \dots, W_N$)을 수신하고, 수신된 복수의 가중치를 취합하여 내부 레이어의 가중치들을 업데이트함으로써 마스터 모델의 성능을 향상시킬 수 있다.
- [79] 취합의 예로, 관리 서버(200)는 복수의 사용자 단말기들(100-1, 100-2, ... 100-N)로부터 수신된 복수의 가중치($W_1, W_2, \dots, W_N$)의 평균을 산출한 후, 산출된 평균값을 마스터 모델의 갱신을 위한 가중치로 이용할 수 있다. 이상의 과정을 통해 학습된 마스터 모델은 버전으로 관리될 수 있으며 신규 버전으로 업데이트된 마스터 모델은 각 사용자 단말기들(100-1, 100-2, ... 100-N)로 재 배포

후 다시 설치 됨으로써 로컬 모델로 동작할 수 있다.

- [80] 또한, 취합에 있어서 사용자 정보를 더욱 이용하여 사용자의 연령, 성별, 이체 금액 등을 이용하여 연령별 이체 금액의 표현 방식, 연령별 은행 명의 표현 방식을 구분하여 학습시킬 수 있다. 따라서 마스터 모델은 세부적으로 사용자의 특성에 맞추어 구분하여 관리 및 학습되어 보다 정확한 예측이 가능하도록 한다.
- [81]
- [82] 나아가, 본 실시예에서 사용자 단말기(100)는 로컬 모델과 마스터 모델을 독립적으로 이용하고 각각의 예측 결과를 이용하여 보다 정확한 이체 정보의 포함 여부를 판단할 수 있다.
- [83] 즉, 가중치를 제공하는 로컬 모델은 로컬 차원에서 학습된 결과를 이용하여 학습함으로써 로컬 사용자의 개인화된 특성에 보다 좋은 성능을 가질 수 있다. 반면 마스터 모델은 절대 다수의 공통된 이체 정보들의 특징들을 반영하도록 학습하므로 새로운 형태의 메시지에 대해서는 보다 좋은 성능을 가질 수 있다.
- [84] 따라서, 연합 학습 후 배포된 이전의 마스터 모델을 기초로 사용자의 메시지에 포함되는 특수한 패턴을 학습한 로컬 모델의 예측 결과와 새로운 버전의 마스터 모델을 모두 이용하여 이체 정보의 포함 여부를 판단할 수 있다.
- [85] 예를 들어, 획득된 메시지에 대하여 로컬 모델에서 산출된 이체 정보의 제1 예상 포함 확률과 마스터 모델에서 산출된 이체 정보의 제2 예상 포함 확률의 평균을 이용하여 이체 정보의 포함 여부를 판단하는 것도 가능하다.
- [86] 이를 통해 다른 사용자들의 특징을 반영한 결과와 사용자 자신의 특징을 반영한 결과를 모두 이용하여 이체 정보가 포함되었는 지를 정확히 판단할 수 있다.
- [87]
- [88] 다음, 이상의 과정(S100, S200)을 통해 학습된 판단 모델을 이용하여 사용자 단말기(100)는 메시지가 이체 정보를 포함하는 지를 판단하고, 이체 정보가 포함된 경우 사용자 단말기(100)는 전처리된 메시지에서 이체 정보를 분류할 수 있다(S300).
- [89] 먼저 도 4를 참고하면, 사용자 단말기(100)는 기 설정된 정규식(regular expression)(50)을 기초로 패턴 매칭을 통해 메시지 내에 포함된 이체 정보를 종류별로 분류(51)할 수 있다.
- [90] 구체적으로 정규식에는 금액, 은행, 계좌의 고유 패턴을 문법에 따라 정의함으로써 이체 정보가 포함되었다고 판단된 메시지 내에서 해당 정보들을 추출할 수 있다.
- [91] 예를 들어, 계좌의 패턴에는 10자리 이상 연속하는 숫자, 숫자와 하이픈(-)의 조합, 또는 은행 별로 결정된 고유의 숫자열이 반복되고 10자리 이상 연속하는 경우를 패턴으로 정의할 수 있다.
- [92] 또한, 금액에는 소수점을 포함하는 산용 숫자와 한글로 된 금액 단위의 조합(예를 들어, 5천원, 5.2만원 등), 한글만으로 된 숫자와 금액 단위 조합(예를

들어, 이만원, 오천원 등) 등과 같은 패턴을 가진 문자열을 정규식으로 정의할 수 있다.

- [93] 그 외, 은행의 경우는 OO은행 또는 00BANK 등의 은행을 지칭하는 용어로 끝나는 문자열 또는, 시중의 은행명을 직접 패턴으로 정의하여 저장할 수 있다.
- [94] 따라서, 사용자 단말기(100)는 정규식의 패턴 매칭을 통해 은행명, 계좌번호 및 이름에 대응되는 텍스트를 전처리된 메시지 내에서 분류할 수 있다.
- [95]
- [96] 또한, 사용자 단말기(100)는 기 학습된 신경망 모델을 이용하여 전처리된 메시지로부터 예측 결과를 산출하고, 산출된 예측 결과를 기초로 메시지 내 이체 정보를 분류할 수 있다. 여기서, 신경망 모델은 전처리된 메시지의 분절된 텍스트 별로 예측 결과를 산출하도록 학습될 수 있다.
- [97] 이와 관련하여 도 5 및 도 6을 참조하여 설명한다.
- [98] 먼저 도 5를 참조하면, 본 실시예에 따른 신경망 모델(40)은 복수의 합성곱 연산을 수행하는 레이어로 구성된 CNN(Convolution Neural Network) 모델로 구성될 수도 있다.
- [99] 사용자 단말기(100)는 전처리된 메시지의 분절된 텍스트, 바람직하게는 텍스트에 대한 임베딩 벡터를 신경망 모델(40)에 입력함으로써 해당 텍스트의 종류를 예측하도록 할 수 있다. 신경망 모델(40)에 입력된 텍스트 내부의 특징 값들은 각각의 합성곱 레이어(Convolution layer)를 거치면서 특징 맵의 형태로 생성되며 반복 연산을 통해 생성된 특징 맵은 완전연결 레이어(Fully-connected layer)에 따라 분류 예측 결과(41)로 산출될 수 있다. 사용자 단말기(100)는 이상의 산출된 예측 결과를 기초로 분절된 텍스트를 분류한다.
- [100]
- [101] 또는 도 6을 참조하면, 본 실시예에 따른 신경망 모델(30)은 연속된 데이터에서 패턴을 인식하는 RNN(Recurrent Neural Network) 모델로 구성될 수 있다.
- [102] 사용자 단말기(100)는 전처리된 메시지 내 분절된 텍스트 또는 텍스트의 인덱스 값에 따른 임베딩 벡터를 각각 추출하고, 추출된 임베딩 벡터를 신경망 모델(30)에 순차적으로 입력할 수 있다.
- [103] 신경망 모델(30)은 입력 레이어에 순차적으로 입력된 임베딩 벡터를 RNN 모델의 히든 레이어에 입력되고, RNN 노드는 이전의 레이어의 출력과 함께, 주변 RNN 노드의 출력을 이용하여 연산을 수행할 수 있다.
- [104] 따라서, 신경망 모델(30)은 분절된 텍스트의 순차적인 상관 관계에 기초한 분류 결과를 예측 값으로 산출될 수 있다.
- [105] 예를 들어, 신경망 모델(30)을 이용하여 순차적으로 입력된 단위 텍스트(X1, X2, X3, X4)에 대한 복수의 예측 결과(31, 32, 33, 34)가 산출된 경우, 사용자 단말기(100)는 제1 예측 결과(31)에 따라 텍스트 X1은 이체 정보 중 은행명으로 분류하고, 제2 예측 결과(32)에 따라 텍스트 X2는 이체 정보 중 이름으로 분류하고, 제3 예측 결과(33)에 따라 텍스트 X3은 이체 정보와 무관한 텍스트로

분류하고 제4 예측 결과(34)에 따라 텍스트 X4는 이체 정보 중 계좌번호로 분류할 수 있다.

- [106] 이상, 본 실시예에 따르면 사용자 단말기(100)는 간소화된 구성의 판단 모델을 통하여 이체 정보의 포함 여부를 우선적으로 판단하고, 이체 정보를 포함하는 경우에는 정규식 또는 추가 신경망을 이용하여 이체 정보를 분류함으로써 보다 효율적인 처리가 가능하도록 한다.
- [107]
- [108] 이어서, 도 7을 참조하여 상술한 이체 정보 판단 방법을 수행하는 어플리케이션의 구성에 대해서 설명한다.
- [109] 도 7은 본 발명의 일 실시예에 따른 사용자 단말기(100)에서 동작하는 어플리케이션의 구조를 나타낸 블록도이다.
- [110] 도 7을 참조하면, 사용자 단말기(100) 상의 어플리케이션은 전처리부(110), 이체 정보 판단부(120) 및 이체 정보 분류부(130)로 구성될 수 있다.
- [111] 전처리부(110)는 메시지에서 이체 정보를 판단하기 위해 사용자 단말기(100) 내 획득된 메시지를 전처리할 수 있다.
- [112] 구체적으로, 전처리부(110)는 획득된 메시지에 포함된 숫자 및 특정 단어를 기준으로 앞뒤에 공백을 추가하여 해당 메시지를 분절하는 것으로 획득된 메시지를 전처리할 수 있다.
- [113] 이체 정보 판단부(120)는 임베딩 기반의 판단 모델을 기초로 전처리된 메시지에서 이체 정보의 예상 포함 확률을 산출함으로써 메시지 내 이체 정보의 유무를 판단할 수 있다.
- [114] 구체적으로 이체 정보 판단부(120)는 전처리된 메시지에서 분절된 텍스트 별로 대응되는 임베딩 벡터(Embedding vector)들을 추출하고, 추출된 임베딩 벡터의 평균값을 산출한 후, 로컬 모델을 기초로 가중합(weighted sum)하여 예상 포함 확률을 산출하고, 산출된 예상 포함 확률과 기준값을 비교하여 이체 정보의 포함여부를 판단할 수 있다.
- [115] 이체 정보의 포함여부를 판단함과 동시에, 이후 사용자의 이체 여부 또는 이체 정보의 수정이나 편집에 따른 결과를 실제 값으로 이용하여 로컬 모델 내부 레이어의 갱신 값을 생성할 수 있으며, 갱신 값은 학습에 이용될 수 있다.
- [116] 이체 정보 분류부(130)는 기 설정된 정규식(regular expression)을 기초로 패턴 매칭을 통해 메시지 내에 포함된 이체 정보를 종류별로 분류할 수 있다.
- [117] 또한, 이체 정보 분류부(130)는 기 학습된 신경망 모델을 이용하여 전처리된 메시지로부터 예측 결과를 산출하고, 산출된 예측 결과를 기초로 메시지 내 이체 정보를 분류할 수 있다.
- [118] 이상의 본 발명에 따르면, 저용량의 임베딩 기반의 판단 모델을 이용하여 이체 정보의 포함 여부를 우선적으로 판단함으로써, 메시지 내 이체 정보의 분류 효율을 높일 수 있다.
- [119] 또한, 사용자의 메시지를 기초로 학습된 로컬 기반의 모델을 이용함으로써,

개인화된 특징이 반영된 메시지 내 이체 정보를 보다 정확하게 판단 및 분류할 수 있다.

- [120] 또한, 다양한 사용자의 메시지를 기초로 공통적인 특징을 학습한 모델과 로컬 기반의 모델을 독립적으로 운영함으로써 사용자의 특성에 따른 이체 정보에 대한 적응성을 높임과 동시에 학습을 위한 개인정보의 이용을 최소화할 수 있다.
- [121]
- [122] 나아가, 여기에 설명되는 다양한 실시예는 예를 들어, 소프트웨어, 하드웨어 또는 이들의 조합된 것을 이용하여 컴퓨터 또는 이와 유사한 장치로 읽을 수 있는 기록매체 내에서 구현될 수 있다.
- [123] 하드웨어적인 구현에 의하면, 여기에 설명되는 실시예는 ASICs (application specific integrated circuits), DSPs (digital signal processors), DSPDs (digital signal processing devices), PLDs (programmable logic devices), FPGAs (field programmable gate arrays, 프로세서(processors), 제어기(controllers), 마이크로 컨트롤러(micro-controllers), 마이크로 프로세서(microprocessors), 기타 기능 수행을 위한 전기적인 유닛 중 적어도 하나를 이용하여 구현될 수 있다. 일부의 경우에 본 명세서에서 설명되는 실시예들이 제어 모듈 자체로 구현될 수 있다.
- [124] 소프트웨어적인 구현에 의하면, 본 명세서에서 설명되는 절차 및 기능과 같은 실시예들은 별도의 소프트웨어 모듈들로 구현될 수 있다. 상기 소프트웨어 모듈들 각각은 본 명세서에서 설명되는 하나 이상의 기능 및 작동을 수행할 수 있다. 적절한 프로그램 언어로 쓰여진 소프트웨어 어플리케이션으로 소프트웨어 코드가 구현될 수 있다. 상기 소프트웨어 코드는 메모리 모듈에 저장되고, 제어모듈에 의해 실행될 수 있다.
- [125] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다.
- [126] 따라서, 본 발명에 개시된 실시 예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

청구범위

- [청구항 1] 딥러닝 기반의 자연어 처리를 통한 메시지 내 이체 정보를 판단하는 방법에 있어서,
 사용자 단말기 내 획득된 메시지를 기준 단어에 따라 전처리하는 단계;
 상기 전처리된 메시지에서 분절된 텍스트 별로 대응되는 임베딩 벡터를 추출하고, 상기 추출된 임베딩 벡터의 가중 연산을 통해 이체 정보의 포함 여부를 판단하는 단계; 및
 이체 정보를 포함하는 것으로 판단된 상기 전처리된 메시지 내 이체 정보를 분류하는 단계;를 포함하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 2] 제 1 항에 있어서,
 상기 전처리하는 단계는 상기 메시지에 포함된 숫자 또는 특정 단어를 기준으로, 상기 숫자 또는 특정 단어의 위치 전후에 공백을 추가하여 상기 메시지를 분절하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 3] 제 2 항에 있어서,
 상기 판단하는 단계는 상기 획득된 메시지 내 이체 정보의 포함 여부를 상기 사용자 단말기 상에서 동작하는 로컬 모델을 통해 판단하되,
 상기 로컬 모델은 상기 이체 정보의 포함 여부에 대한 판단 결과를 이용하여 내부 가중 연산을 수행하는 레이어의 학습 정보를 생성하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 4] 제 3 항에 있어서,
 상기 로컬 모델은 복수의 사용자 단말기로부터 생성된 학습 정보로 연합 학습하여 배포되는 마스터 모델을 이용하여 갱신(update)되는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 5] 제 4 항에 있어서,
 상기 판단하는 단계는,
 상기 획득된 메시지에 대하여 상기 로컬 모델에서 산출된 이체 정보의 제1 예상 포함 확률 및, 상기 마스터 모델에서 산출된 이체 정보의 제2 예상 포함 확률을 이용하여 이체 정보의 포함 여부를 판단하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 6] 제 1 항에 있어서,
 상기 이체 정보를 분류하는 단계는 분류 별 텍스트의 패턴을 문법으로 정의한 정규식(regular expression)으로 상기 메시지 내 이체 정보를 분류하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 7] 제 1 항에 있어서,
 상기 이체 정보를 분류하는 단계는
 이체 정보를 분류하도록 학습된 신경망 모델에 상기 전처리된 메시지를

- 입력함으로써 상기 메시지 내 이체 정보를 분류하는 것을 특징으로 하는 이체 정보 판단 방법.
- [청구항 8] 메시지 내 이체 정보를 판단하는 사용자 단말기 상에서 동작하는 어플리케이션에 있어서,
사용자 단말기 내 획득된 메시지를 기준 단어에 따라 전처리하는 전처리부;
상기 전처리된 메시지에서 분절된 텍스트 별로 대응되는 임베딩 벡터를 추출하고, 상기 추출된 임베딩 벡터의 가중 연산을 통해 이체 정보의 포함 여부를 판단하는 이체 정보 판단부; 및
이체 정보를 포함하는 것으로 판단된 상기 전처리된 메시지 내 이체 정보를 분류하는 이체 정보 분류부;를 포함하는 것을 특징으로 하는 어플리케이션.
- [청구항 9] 제 8 항에 있어서,
상기 전처리부는 상기 메시지에 포함된 숫자 또는 특정 단어를 기준으로, 상기 숫자 또는 특정 단어의 위치 전후에 공백을 추가하여 상기 메시지를 분절하는 것을 특징으로 하는 어플리케이션.
- [청구항 10] 제 9 항에 있어서,
상기 이체 정보 판단부는 상기 획득된 메시지 내 이체 정보의 포함 여부를 상기 사용자 단말기 상에서 동작하는 로컬 모델을 통해 판단하되,
상기 로컬 모델은 상기 이체 정보의 포함 여부에 대한 판단 결과를 이용하여 내부 가중 연산을 수행하는 레이어의 학습 정보를 생성하는 것을 특징으로 하는 어플리케이션.
- [청구항 11] 제 10 항에 있어서,
상기 로컬 모델은 복수의 사용자 단말기로부터 생성된 학습 정보로 연합 학습하여 배포되는 마스터 모델을 이용하여 갱신(update)되는 것을 특징으로 하는 어플리케이션.
- [청구항 12] 제 11 항에 있어서,
상기 이체 정보 판단부는,
상기 획득된 메시지에 대하여 상기 로컬 모델에서 산출된 이체 정보의 제1 예상 포함 확률 및, 상기 마스터 모델에서 산출된 이체 정보의 제2 예상 포함 확률을 이용하여 이체 정보의 포함 여부를 판단하는 것을 특징으로 하는 어플리케이션.
- [청구항 13] 제 8 항에 있어서,
상기 이체 정보를 분류부는 분류 별 텍스트의 패턴을 문법으로 정의한 정규식(regular expression)으로 상기 메시지 내 이체 정보를 분류하는 것을 특징으로 하는 어플리케이션.
- [청구항 14] 제 8 항에 있어서,
상기 이체 정보를 분류부는,

이체 정보를 분류하도록 학습된 신경망 모델에 상기 전처리된 메시지를 입력함으로써 상기 메시지 내 이체 정보를 분류하는 것을 특징으로 하는 어플리케이션.

[청구항 15] 제 1 항 내지 제 7 항 중 어느 한 항에 따른 이체 정보 판단 방법을 수행하는 프로그램이 저장된 컴퓨터 판독 가능한 기록 매체.