

명세서

발명의 명칭: 주어진 문서가 독자에게 보다 높은 신뢰를 받을 수 있도록 하는 문서 신뢰도 증강 방법 및 그 시스템

기술분야

- [1] 본 발명은 문서의 신뢰도를 증강시키는 기술에 관한 것으로, 보다 구체적으로는 주어진 문서에 대해 가능한 검색 후보군에 대한 자동 신뢰도 분포 예측을 수행하고, 원본 문서와 검색 후보군의 신뢰도 분포 각각이 보이는 평균과 표준편차 변화에 기반하여 신뢰도 증강을 수행할 수 있는 방법 및 그 시스템에 관한 것이다.

배경기술

- [2] 자연언어처리 기술이 발전하고 의미론적/화용론적 특징에 대한 자동 분석 기술의 성능이 향상함에 따라 자동 문서 교정 시스템에 대한 연구 기술 개발과 상용화가 활발히 진행되고 있다.
- [3] 자동 문서 교정 시스템에 대한 일 실시예의 기술은 신경망 기반 기계 번역 기법(neural machine translation, NMT)을 활용한 학습 모델을 활용한 자동 문서 교정 시스템을 연구 개발하고, 해당하는 시스템이 문법 자동 교정에 대한 CoNLL-2014 Shared task test set에 대해 약 40%에 해당하는 F0.5 score의 높은 성능을 보임을 실험적으로 입증하였다. 자동 문서 교정 시스템에 대한 다른 일 실시예의 기술은 자동 문법 교정 방법 및 시스템의 평가에 있어 모국어 사용자의 언어적 판단과 가장 일치할 수 있는 평가 지표를 탐색하고, 기존에 자동 문법 교정 방법 및 시스템의 평가를 위해 활용되어온 BLEU 지표의 변형인 GLEU 지표를 제안하였다. 자동 문서 교정 시스템에 대한 또 다른 일 실시예의 기술은 자동문서 교정 시스템을 구축하기 위해 필요한 학습 데이터 및 평가 데이터를 효과적으로 구축하기 위한 주석 보조 도구인 ERRANT를 연구 개발하였으며, 자동 문서 교정 시스템에 대한 또 다른 일 실시예의 기술은 기술 문서에 대한 자동 심층 교정 시스템을 학습시키고 평가시키기 위한 대규모 학습 평가 데이터 코퍼스를 구축하였다. 자동 문서 교정 시스템에 대한 또 다른 일 실시예의 기술은 문법 자동 교정을 위한 대규모 학습 평가 데이터 코퍼스인 EF-Cambridge Open Language Database (EFCAMDAT)에 기반하여 학습시킨 뉴럴 시퀀스 대 시퀀스(neural sequence-to-sequence) 모델을 활용하여 다양한 주제의 문서에 대해 높은 성능을 보이는 자동 문법 교정 시스템을 연구 개발하였다.
- [4] 자동 문서 교정 시스템에 대한 대표적인 상용화 사례로, Grammarly는 사용자가 입력한 문서에 대한 자동 문법 교정 및 어조 교정 서비스를 제공한다. 이와 관련하여 미국 특허 US9465793B2 (granted, 2016-10-11), "Systems and methods for advanced grammar checking"는 자동화된 문법 오류 교정 및 검색 제안 및 문법 오류 교정 제안 후보에 대한 우선 순위 설정 방법을 제안하였다. 또한,

Grammarly는 최근, 단순 문법 교정 및 어조 교정뿐만 아니라, 문서의 의도(예를 들어, 새로운 정보 제공, 특정 주제에 대한 설명, 설득, 스토리텔링), 독자의 특징(예를 들어, 일반, 지식인, 전문가), 문서의 스타일(예를 들어, 격식 있는, 격식 없는), 문서의 감정 세기(예를 들어, 약함, 강함), 문서 도메인(예를 들어, 일반, 학술, 사업, 기술, 창의, 캐주얼)에 맞추어 문장 스타일에 대한 자동 첨삭 제안 기능을 추가하였다.

- [5] 다른 대표적인 상업화 사례로, Google의 G-Suite는 최근 Google Docs, Gmail 등의 서비스에 자동화된 문법 오류 교정 제안 기능을 추가하였다. 또 다른 대표적인 상업화 사례는 Google Docs 서비스에 대한 추가 기능 프로그램으로서, 문서의 설득력, 주제 전개, 일관성, 문법 오류, 단어 선택 등에 대한 피드백을 자동으로 제공하는 프로그램을 연구 개발하였다.
- [6] 자동 문서 교정을 위한 기존의 방법론은 대중적인 합의가 비교적 높은 개념들에 국한된 것으로, 신뢰도 개념과 같이 개인차가 큰 값에 대해서 적용되기 힘들다는 단점이 있다.
- [7] 예를 들어, 자동 문서 교정을 위한 기존의 방법론에서 관련 요소로서 분석하는 문서의 설득력, 일관성 등은 문서 내의 문장 배열 및 구성 요소의 특징에 대한 것으로 대중적 합의율이 높으며, 감정 세기 역시 대중적 합의율이 높다.
- [8] 주어진 문장에 대한 신뢰도는 주어진 문장이 서술하는 특정 사건이 실제로 일어났는지(factuality)와 관련하여 정의되거나, 특정 화자가 말하는 특정 주장/내용을 신뢰하는지와 관련되어 정의된다. 특정 사건이 실제로 일어났는지(factuality)와 관련하여 진행된 신뢰도 연구들은 이에 대해 비교적 대중적 합의율이 높다는 것을 확인하였다.
- [9] 그러나, 특정 화자가 말하는 특정 주장/내용에 대한 신뢰도 개념은 개인의 주관과 매우 밀접한 관련이 있기 때문에 어떤 문장/문서를 신뢰하는지에 대한 것은 합의율이 낮다는 점이 여러 연구를 통해 확인되었다.
- [10] 다르게 말해, 특정 문서가 높은/낮은 설득력을 갖는다는 것, 높은/낮은 일관성을 갖는다는 것, 높은/낮은 감정 세기를 갖는다는 것에 대한 대중적 합의율은 특정 문서가 신뢰할 수 있는/없는 내용을 담고 있다는 것에 대한 대중적 합의율보다 높다. 이에 대한 이유는 다음과 같다. 설득력/일관성은 논지의 구조에 대한 것이며, 명확한 논지 구조를 가진 문서의 경우 그 내용에 대한 반감이 있을 경우에도, 해당하는 문서가 높은 설득력/일관성을 갖는다는 평가가 가능하다. 또한, 감정의 세기 역시, 감정 유발 단어 및 감정 유발 표현이 제한적이기 때문에 해당하는 단어와 표현을 많이 사용하는 문서는 높은 감정 세기를 갖는다는 평가에 대한 합의율이 높을 수 있다.
- [11] 그러나, 신뢰도의 경우 개인의 주관적인 신념에 대한 것이기 때문에 그 내용에 대해 어떤 신념을 가지고 있는 지가 신뢰도를 결정짓는 가장 주요한 요소이며, 이는 인구 통계학적 특징에 의해 큰 영향을 받을 수 있으며, 인구 통계학적 특징으로 정의될 수 없는 개인의 주관에 의한 영향도 받을 수 있기 때문에

신뢰도 평가에 대한 대중적 합의율이 비교적 낮다.

- [12] 따라서, 자동 문서 교정을 위한 기존의 방법론이 단순히 일관성 점수, 설득력 점수, 감정 세기 점수 등의 한 숫자로 정의된 지표를 통해 교정 제안 혹은 문서 첨삭 후보군의 적절성을 평가하였다면, 신뢰도의 증강을 위한 문서 교정 및 첨삭을 위해서는 개인차를 고려한 신뢰도 분포 개념에 입각하여 교정 제안 혹은 문서 첨삭 후보군의 적절성을 평가해야 한다.

발명의 상세한 설명

기술적 과제

- [13] 본 발명의 실시예들은, 주어진 문서에 대해 가능한 첨삭 후보군에 대한 자동 신뢰도 분포 예측을 수행하고, 원본 문서와 첨삭 후보군의 신뢰도 분포 각각이 보이는 평균과 표준편차 변화에 기반하여 신뢰도 증강을 수행할 수 있는 방법 및 그 시스템을 제공한다.

과제 해결 수단

- [14] 본 발명의 일 실시예에 따른 문서 신뢰도 증강 방법은 사용자로부터 문서가 입력되면 상기 입력된 문서의 각 문장에 대한 첨삭 후보군을 생성하는 단계; 상기 입력된 문서와 상기 생성된 첨삭 후보군 각각에 대한 신뢰도 분포를 예측하는 단계; 및 상기 예측된 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식을 통해 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 단계를 포함한다.
- [15] 나아가, 본 발명의 일 실시예에 따른 문서 신뢰도 증강 방법은 상기 입력된 문서에 대해 신뢰도 증강된 신뢰도 증강 문서 및 상기 신뢰도 증강의 전과 후에 대한 신뢰도 분포 변화 정보를 상기 사용자에게 제공하는 단계를 더 포함할 수 있다.
- [16] 상기 사용자에게 제공하는 단계는 상기 사용자와의 상호작용을 통해 상기 사용자가 희망하는 신뢰도 분포의 양태에 가장 유사한 문서 첨삭 결과를 상기 사용자에게 더 제공할 수 있다.
- [17] 상기 복수 증강 방식은 상기 신뢰도 분포의 미리 설정된 기준 평균 이상의 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 상기 신뢰도 분포의 미리 설정된 기준 표준편차 이하의 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 상기 신뢰도 분포의 미리 설정된 기준 {평균/표준편차} 이상의 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함할 수 있다.
- [18] 상기 생성하는 단계는 상기 사용자로부터 상기 복수 증강 방식에 대한 제어 정보를 입력 받는 단계; 상기 입력된 문서의 각 문장에 대한 전처리를 수행하는 단계; 및 상기 입력된 문서의 각 문장에 대한 복수의 첨삭 후보 문장 집합을 생성하는 단계를 포함하고, 상기 예측하는 단계는 상기 입력된 문서의 각 문장과 상기 첨삭 후보 문장 집합 내의 각 문장을 통해 조합 가능한 복수의 문서 집합의

각 문장에 대한 신뢰도 분포를 예측하며, 상기 수행하는 단계는 상기 예측된 복수의 문서 집합의 각 문장에 대한 신뢰도 분포와 상기 제어 정보에 기초하여 신뢰도 증강 문서를 선별함으로써, 상기 입력된 문서에 대한 신뢰도 증강을 수행할 수 있다.

[19]

[20] 본 발명의 일 실시예에 따른 문서 신뢰도 증강 시스템은 사용자로부터 문서가 입력되면 상기 입력된 문서의 각 문장에 대한 검색 후보군을 생성하는 생성부; 상기 입력된 문서와 상기 생성된 검색 후보군 각각에 대한 신뢰도 분포를 예측하는 예측부; 및 상기 예측된 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식을 통해 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 증강부를 포함한다.

[21] 나아가, 본 발명의 일 실시예에 따른 문서 신뢰도 증강 시스템은 상기 입력된 문서에 대해 신뢰도 증강된 신뢰도 증강 문서 및 상기 신뢰도 증강의 전과 후에 대한 신뢰도 분포 변화 정보를 상기 사용자에게 제공하는 출력부를 더 포함할 수 있다.

[22] 상기 출력부는 상기 사용자와의 상호작용을 통해 상기 사용자가 희망하는 신뢰도 분포의 양태에 가장 유사한 문서 검색 결과를 상기 사용자에게 더 제공할 수 있다.

[23] 상기 복수 증강 방식은 상기 신뢰도 분포의 미리 설정된 기준 평균 이상의 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 상기 신뢰도 분포의 미리 설정된 기준 표준편차 이하의 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 상기 신뢰도 분포의 미리 설정된 기준 {평균/표준편차} 이상의 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함할 수 있다.

[24] 상기 생성부는 상기 사용자로부터 상기 복수 증강 방식에 대한 제어 정보를 입력 받고, 상기 입력된 문서의 각 문장에 대한 전처리를 수행하며, 상기 입력된 문서의 각 문장에 대한 복수의 검색 후보 문장 집합을 생성하고, 상기 예측부는 상기 입력된 문서의 각 문장과 상기 검색 후보 문장 집합 내의 각 문장을 통해 조합 가능한 복수의 문서 집합의 각 문장에 대한 신뢰도 분포를 예측하며, 상기 증강부는 상기 예측된 복수의 문서 집합의 각 문장에 대한 신뢰도 분포와 상기 제어 정보에 기초하여 신뢰도 증강 문서를 선별함으로써, 상기 입력된 문서에 대한 신뢰도 증강을 수행할 수 있다.

발명의 효과

[25] 본 발명의 실시예들에 따르면, 주어진 문서에 대해 가능한 검색 후보군에 대한 자동 신뢰도 분포 예측을 수행하고, 원본 문서와 검색 후보군의 신뢰도 분포 각각이 보이는 평균과 표준편차 변화에 기반하여 신뢰도 증강을 수행할 수 있다.

[26] 본 발명의 실시예들에 따르면, 분포 개념의 도입이라는 필요조건이 만족되기

때문에 큰 개인차를 포함하는 신뢰도 개념에 대한 문서 탐색을 가능하게 한다. 즉, 본 발명은 (1) 신뢰도 분포의 높은 평균을 지향하는 방식의 신뢰도 증강을 진행하여 보다 많은 사람들이 신뢰도 증강의 대상이 되는 문서를 신뢰할 수 있도록 문서를 탐색하는 것, (2) 신뢰도 분포의 낮은 표준편차를 지향하는 방식의 신뢰도 증강을 진행하여 신뢰도 증강의 대상이 되는 문서에 대한 신뢰 여부가 보다 논란의 여지가 없도록 탐색하는 것, (3) 신뢰도 분포의 높은 {평균/표준편차}를 지향하는 방식의 신뢰도 증강을 진행하여 상기 두 방식의 결합된 방식으로 신뢰도 증강의 대상이 되는 문서를 탐색하는 것을 가능하게 한다.

[27] 본 발명의 실시예들에 따르면, 교정 대상이 되는 문서 내의 각 문장에 대해 사용자가 드래그 앤 드롭 방식을 통해 희망하는 신뢰도 분포의 양태를 입력하고 해당하는 신뢰도 분포의 양태와 가장 유사한 신뢰도 분포 양태를 보이는 문장 탐색 후보를 사용자에게 제공할 수 있다. 따라서 본 발명은 개인의 주관에 대해 큰 영향을 받기 때문에 자동화된 교정 시스템에 적용되기 어려웠던 신뢰도 증강에 대한 문서 교정을 분포 개념의 도입을 통해 가능하도록 하며, 이를 통해 불특정 타인 집단으로부터 보다 신뢰받기 위해 문서가 어떻게 변형되어야 하는지에 대한 상세 정보를 다양한 형태로 제공받을 수 있다.

[28] 이러한 본 발명은 독자의 신뢰도가 중요한 공문서, 뉴스, 기술 문서의 자동 교정을 수행할 수도 있고, 개인의 주관에 기반하여 신뢰성이 높은 문서를 생산할 수도 있다.

도면의 간단한 설명

[29] 도 1은 본 발명의 일 실시예에 따른 문서 신뢰도 증강 시스템에 대한 구성을 나타낸 것이다.

[30] 도 2는 도 1에 도시된 신뢰도 증강부의 일 실시예 구성을 나타낸 것이다.

[31] 도 3은 도 1에 도시된 출력부의 일 실시예 구성을 나타낸 것이다.

[32] 도 4는 본 발명의 문서 신뢰도 증강 시스템의 출력 결과를 설명하기 위한 일 예시도를 나타낸 것이다.

[33] 도 5는 문서 신뢰도 증강 시스템에 포함되는 문장 탐색 후보 코퍼스의 예시를 나타낸 것이다.

[34] 도 6은 문서 신뢰도 증강 시스템에 포함되는 신뢰도 분포 코퍼스의 예시를 나타낸 것이다.

[35] 도 7은 도 1에 도시된 신뢰도 증강부에 의한 일 실시예의 동작 흐름도를 나타낸 것이다.

[36] 도 8은 도 1에 도시된 출력부에 의한 일 실시예의 동작 흐름도를 나타낸 것이다.

발명의 실시를 위한 최선의 형태

[37] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본

발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

- [38] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며, 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 구성요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다.
- [39] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [40] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예들을 보다 상세하게 설명하고자 한다. 도면 상의 동일한 구성요소에 대해서는 동일한 참조 부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [41]
- [42] 본 발명의 실시예들은, 주어진 문서에 대해 가능한 침삭 후보군에 대한 자동 신뢰도 분포 예측을 수행하고, 원본 문서와 침삭 후보군의 신뢰도 분포 각각이 보이는 평균과 표준편차 변화에 기반하여 신뢰도를 증강시키는 것을 그 요지로 한다.
- [43] 여기서, 본 발명은 신뢰도 분포의 높은 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 신뢰도 분포의 낮은 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 신뢰도 분포의 높은 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함하는 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 따라 신뢰도 증강을 수행할 수 있다.
- [44] 나아가, 본 발명은 사용자가 입력한 문서에 대한 신뢰도 증강을 진행한 뒤 사용자에게 증강된 문서와 관련된 통계 지표를 그래프 형태로 제공하며, 사용자와의 상호작용을 통해 불특정 타인 집단이 사용자가 입력한 문서의 미세 조정에 따라 어떤 신뢰도 변화를 보일 지에 대하여 사전 모니터링을 수행할 수 있게 한다.
- [45]
- [46] 도 1은 본 발명의 일 실시예에 따른 문서 신뢰도 증강 시스템에 대한 구성을 나타낸 것이다.

- [47] 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 문서 신뢰도 증강 시스템(100)은 신뢰도 증강부(110), 문장 첨삭 후보 코퍼스(120), 신뢰도 분포 코퍼스(130), 출력부(140) 및 제어부(150)를 포함한다.
- [48] 여기서, 본 발명의 문서 신뢰도 증강 시스템(100)은 도 4에 도시된 바와 같이, 사용자로부터 입력 받은 문서(입력 텍스트)에 대해 각 문장 별로 신뢰도 증강을 수행하며 신뢰도가 증강된 문서를 사용자에게 제공함과 동시에 신뢰도 증강 전/후에 해당하는 독자 신뢰도 분포를 각 문장에 대해 사용자에게 출력 제공할 수 있다.
- [49] 신뢰도 증강부(110)는 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 따라 입력 받은 문서에 대한 신뢰도 증강을 수행한다.
- [50] 이 때, 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식은 신뢰도 분포의 높은 평균 예를 들어, 미리 설정된 기준 평균 이상의 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 신뢰도 분포의 낮은 표준편차 예를 들어, 미리 설정된 기준 표준편차 이하의 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 신뢰도 분포의 높은 {평균/표준편차} 예를 들어, 미리 설정된 기준 {평균/표준편차} 이상의 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함할 수 있다.
- [51] 구체적으로, 높은 신뢰도 평균을 지향하는 방식의 신뢰도 증강은 보다 많은 사람들이 신뢰도 증강의 대상이 되는 문서를 신뢰할 수 있도록 문서를 첨삭하는 것으로 해석할 수 있으며, 낮은 신뢰도 표준편차를 지향하는 방식의 신뢰도 증강은 신뢰도 증강의 대상이 되는 문서에 대한 신뢰 여부가 보다 논란의 여지가 없도록 첨삭하는 것으로 해석할 수 있고, 높은 {평균/표준편차} 값을 지향하는 방식의 신뢰도 증강은 상기 두 방식의 결합으로 해석될 수 있다.
- [52] 즉, 신뢰도 증강부(110)는 사용자로부터 신뢰도 증강의 대상이 되는 문서를 입력으로 받고, 선택적으로 사용자로부터 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 중 어떤 방식에 해당하는 정보를 출력 제공할 것인지 대한 정보(이하, "제어 정보"라 칭함)를 입력 받고, 입력 받은 제어 정보를 제어부(130)로 전달하며, 주어진 문서 내의 각 문장에 대해 첨삭 후보군을 생성하고, 제어 정보를 다시 제어부(130)로부터 전달받아 첨삭 후보군을 조합하여 생성 가능한 문서 중 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 각각 해당되는 문서 중 상기 제어 정보와 일치하는 문서를 선별한다.
- [53] 여기서, 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 각각 해당되는 문서는 (1) 문서 전반에 대해 예측된 신뢰도 분포의 평균이 가장 높은 첨삭 후 문서와, (2) 문서 전반에 대해 예측된 신뢰도 분포의 표준편차가 가장 낮은 첨삭 후 문서와, (3) 문서 전반에 대해 예측된 신뢰도 분포의 {평균/표준편차} 값이 가장 높은 첨삭 후 문서를 포함할 수 있다.
- [54] 나아가, 신뢰도 증강부(110)는 제어 정보가 사용자로부터 입력되지 않았을

경우 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 모두에 해당하는 정보를 제어 정보로 출력 제공할 수 있다.

- [55] 문장 첨삭 후보 코퍼스(120)는 언어 전문가에 의해 수동적으로 문장의 의미를 유지한 채로 문장 구조를 변형하는 방식의 첨삭을 진행한 문장(이하, "첨삭 후 문장"이라 칭함), 원본 문장, 언어 전문가에 의해 0점(즉, 의미 차이 없음)에서 5점(즉, 큰 의미 차이) 사이의 범위 내 점수로 판단된 원본 문장과 첨삭 후 문장의 의미 차이를 저장한다.
- [56] 여기서, 문장 첨삭 후보 코퍼스(120)는 신뢰도 증강부(110)가 주어진 문서 내의 각 문장에 대한 첨삭 후보군을 생성하는 모델을 활용할 때 해당하는 모델을 지도 학습(supervised learning) 시키기 위한 학습 기준으로 활용될 수 있다.
- [57] 도 5는 문서 신뢰도 증강 시스템에 포함되는 문장 첨삭 후보 코퍼스의 예시를 나타낸 것으로, 도 5에 도시된 바와 같이, 문장 첨삭 후보 코퍼스는 복수 예시 문장 집합과 각 예시 문장에 대해서 복수 첨삭된 문장들을 저장하며, 첨삭된 문장마다 첨삭 전 원본 문장과의 유사도를 저장한다.
- [58] 신뢰도 분포 코퍼스(130)는 직접적인 신뢰도 설문을 통해 수집한 각 문장에 대한 독자들의 신뢰도 분포를 저장한다.
- [59] 여기서, 신뢰도 분포 코퍼스(130)는 신뢰도 증강부(110)가 첨삭 후 문서의 신뢰도 분포를 예측함에 있어 신뢰도 분포 예측 모델을 포함할 때 해당하는 모델을 지도 학습(supervised learning) 시키기 위한 학습 기준으로 활용될 수 있다.
- [60] 도 6은 문서 신뢰도 증강 시스템에 포함되는 신뢰도 분포 코퍼스의 예시를 나타낸 것으로, 도 6에 도시된 바와 같이, 신뢰도 분포 코퍼스는 복수 문서 집합과 각 문서에 대해 독자 그룹들이 각각 실제로 설문을 통해 보인 신뢰도 분포를 저장한다.
- [61] 나아가, 신뢰도 분포 코퍼스(130)가 저장하는 복수 문서 집합은 일상생활에서 접할 수 있는 다양한 주제들에 대한 다양한 종류의 문서들을 포함할 수 있어야 한다. 상세한 설명을 위한 예시로, 문서의 주제는 생활, 건강, 정치, 정책, 경제, 환경을 포함할 수 있으며, 문서의 종류는 SNS 게시글, 블로그 게시글, 온라인 뉴스, 온라인 포럼 게시글, 연구 논문, 도서를 포함할 수 있다. 바람직하게, 코퍼스 전반에 걸쳐 직접적인 설문 대상이 되는 문서들 각각이 갖는 주제와 종류는 획일화되지 않아야 한다. 이는 본 발명의 일 실시예에 따른 시스템에서 설문을 통해 수집한 신뢰도 분포 코퍼스(130)가 신뢰도 분포를 자동 예측하기 위한 학습 기준으로 활용되기 때문이다. 신뢰도 분포 코퍼스(130)가 생활/건강에 대한 문서들과 이 문서들에 대해 독자들이 보이는 신뢰도 설문 결과만을 포함할 경우 해당하는 예측 기준 코퍼스를 통해 학습된 예측 모델은 정치/정책에 대한 문서의 신뢰도 분포를 예측하기에 부적절할 것이며 예측 결과가 실제 독자들이 보일 신뢰도 분포와는 상이할 것으로 예상할 수 있다.
- [62] 출력부(140)는 신뢰도 증강부(110)에 의해 선별된 신뢰도 증강된 문서를 사용자에게 출력 제공한다. 동시에, 출력부(140)는 사용자에게 의해 입력된 원본

문서 내 각 문장에 대해 예측된 신뢰도 분포와 신뢰도 증강부(110)에 의해 증강된 문서 내 각 문장에 대해 예측된 신뢰도 분포를 그래프 형태로 병렬 대조 비교하여 사용자에게 출력 제공한다. 예컨대, 출력부(140)는 도 4에 도시된 출력 결과와 같이 원본 문서 내 각 문장에 대해 예측된 신뢰도 분포와 증강된 문서 내 각 문장에 대해 예측된 신뢰도 분포를 그래프 형태로 병렬 대조 비교하여 사용자에게 출력 제공할 수 있다.

[63] 이 때, 출력부(140)는 신뢰도 증강 전/후에 해당되는 독자 신뢰도 분포를 각 문장에 대해 사용자에게 출력 제공하며, 사용자 상호작용을 통해 신뢰도 증강에 대한 자세한 정보를 선택적, 단계별로 제공함으로써, 사용자는 불특정 타인 집단이 사용자가 입력한 문서의 미세 변형에 따라 어떤 신뢰도 반응 변화를 보일지에 대한 사전 모니터링을 수행할 수 있다.

[64] 또한, 출력부(140)는 사용자와 출력된 그래프에 대한 드래그 앤 드랍 방식으로 상호작용할 수 있으며, 사용자가 드래그 앤 드랍 방식을 통해 희망하는 신뢰도 분포의 양태를 입력하였을 때 입력 받은 신뢰도 분포와 가장 가까운 신뢰도 분포를 갖는 칩삭 후보를 선별하여 신뢰도 분포의 대상이 되는 문장을 해당하는 칩삭 후보로 변경 재출력 제공하며, 해당하는 칩삭 후보의 신뢰도 분포로 기존에 사용자가 드래그 앤 드랍 방식으로 상호작용했던 그래프를 변경 재출력하여 제공할 수 있다.

[65] 제어부(150)는 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 중 어떤 종류의 신뢰도 증강을 수행할 것인지를 결정한다.

[66] 이 때, 제어부(150)는 신뢰도 증강부(110)로 수신된 제어 정보에 기초하여 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 중 어떤 종류의 신뢰도 증강을 수행할 것인지를 결정할 수 있다.

[67]

[68] 도 2는 도 1에 도시된 신뢰도 증강부의 일 실시예 구성을 나타낸 것이다.

[69] 도 2에 도시된 바와 같이, 신뢰도 증강부(110)는 입력부(111), 전처리부(112), 칩삭 후보 생성부(113), 신뢰도 분포 예측부(114) 및 증강 문서 선별부(115)를 포함한다.

[70] 입력부(111)는 사용자로부터 신뢰도 증강의 대상이 되는 문서를 입력으로 받고, 선택적으로 사용자로부터 제어 정보를 입력 받아 제어부(130)로 전달한다.

[71] 이 때, 입력부(111)는 제어 정보가 사용자로부터 입력되지 않았을 경우 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 모두에 해당하는 정보를 제어 정보로 제공할 수 있다.

[72] 바람직하게, 상기 제어 정보가 사용자로부터 입력되지 않았을 경우, 상기 제어 정보는 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식 모두에 해당하는 정보를 제어 정보로 출력 제공할 수 있다.

[73] 전처리부(112)는 사용자로부터 입력 받은 문서를 입력부(111)로부터 전달받고, 이에 대해 의미역 분석(semantic role labeling), 의존성 분석(dependency parsing),

담화 구조 분석(discourse parsing)을 수행한다.

- [74] 여기서, 의미역 분석은 DeepSemanticRoleLabeling 또는 PathLSTM Semantic Role Labeler와 같은 의미역 추출기를 통해 진행될 수 있고, 의존성 분석은 StanfordCoreNLP와 같은 구문 분석기를 통해 진행될 수 있으며, 담화 구조 분석은 PDTB-style discourse parser와 같은 담화 구조 분석기를 통해 진행될 수 있다.
- [75] 첨삭 후보 생성부(113)는 전처리부(112)로부터 (1) 사용자로부터 입력 받은 문서, 이에 대한 (2) 의미역 분석 결과, (3) 의존성 분석 결과, (4) 담화 구조 분석 결과를 수신하여 원본 문서 내 각 문장에 대해 복수의 첨삭 후보 문장 집합을 생성한다.
- [76] 이 때, 첨삭 후보 생성부(113)는 각 문장에 대해 생성 가능한 복수 첨삭 후보 문장 중 원본 문장과의 의미 차이가 가장 적은 L개의 문장을 선별하여 복수의 첨삭 후보 문장 집합으로 생성할 수 있다.
- [77] 바람직하게, 첨삭 후보 생성부(113)는 지도 학습(supervised learning) 기반 첨삭 문장 생성 모델을 통해 첨삭 후보 문장을 생성할 수 있으며, 해당하는 모델은 문장 첨삭 후보 코퍼스(120)를 통해 지도 학습될 수 있다. 여기서, 학습 모델은 학습 중에 첨삭 후보 코퍼스(120)로부터 복수 문장 집합 중 원본 문장들을 불러와 의미역 분석, 의존성 분석, 담화 구조 분석을 수행하며, (1) 원본 문장들과 (2) 의미역 분석 결과, (3) 의존성 분석 결과, (4) 담화 구조 분석 결과를 입력 기준으로 삼고, 각 원본 문장마다 상응하는 첨삭 후 문장들 그리고 원본 문장과 첨삭 후 문장의 의미 차이들을 불러와 출력 기준으로 삼을 수 있다. 물론, 첨삭 후보 생성부(113)에서의 의미역 분석, 의존성 분석, 담화 구조 분석 또한 전처리부(112)에서 수행하는 것과 같은 방식으로 진행될 수 있다.
- [78] 신뢰도 분포 예측부(114)는 첨삭 후보 생성부(113)로부터 (1) 사용자로부터 입력 받은 문서 내 문장들, 이에 대한 (2) 의미역 분석 결과, (3) 의존성 분석 결과, (4) 담화 구조 분석 결과, (5) 복수 첨삭 후보 문장 집합을 수신하고 또는 전달받고, 사용자로부터 입력 받은 문서 내의 복수 문장 집합과 이에 상응하는 복수 첨삭 후보 문장 집합 내의 각 문장을 통해, 조합 가능한 복수 문서 집합 내의 각 문장에 대해 독자 그룹이 보일 것으로 예상하는 신뢰도 분포를 예측한다.
- [79] 이 때, 첨삭 후보 문장 집합 내 각 문장을 통해 복수 문서 집합을 조합하는 과정은, 사용자로부터 입력 받은 문서에 존재하는 복수 문장 집합의 문장에 상응하는 복수 첨삭 후보 중 하나의 첨삭 후보를 무작위로 선정하여 대체하는 과정을 복수 문장 집합 전체에 대해 진행하는 방식으로 할 수 있다.
- [80] 바람직하게, 조합 가능한 복수 문서 집합은 상술한 무작위 선정 대체 방식을 M번 순회하여 M개의 첨삭 후보 문서를 생성하는 방식으로 할 수 있다. 여기서, M은 시스템 초기값으로 설정되는 자연수이며, 시스템의 계산 속도와 사용자 만족도를 고려하여 M이 설정될 수 있다.
- [81] 바람직하게, 신뢰도 분포 예측부(114)는 신뢰도 분포 코퍼스(130)로부터 지도

학습(supervised learning)을 통해 학습한 모델을 이용하여 각 문장에 대한 신뢰도 분포를 예측한다. 여기서, 학습 모델은 학습 중에 신뢰도 분포 코퍼스(130)에 저장된 복수 문서 집합 내 각 문장을 신뢰도 분포 예측 대상 문장으로 삼았을 때, 신뢰도 분포 예측 대상 문장의 앞에 등장한 일정 개수 예를 들어, N개 문장, 뒤에 등장한 N개 문장과 해당 문장을 모두 포함하는 $2N+1$ 개 문장 집합에 대한 의미역 분석 결과, 의존성 분석 결과, 담화 구조 분석 결과를 입력 기준으로 삼고, 신뢰도 분포 예측 대상 문장에 대해 실제 설문을 통해 측정된 복수 문장 집합에 신뢰도 분포를 신뢰도 분포 코퍼스(130)로부터 불러와 출력 기준으로 삼을 수 있다. 여기서, N은 시스템 초기값으로 설정되는 자연수이며, 주어진 문서 내에서 신뢰도 분포 예측 대상 문장의 앞 혹은 뒤에 N개보다 적은 수의 문장이 존재할 경우에는 존재하지 않는 문장을 공백 문장(null sentence)으로 대체하여 입력 기준으로 삼을 수 있다. 물론, 신뢰도 분포 예측부(114)에서의 의미역 분석, 의존성 분석, 담화 구조 분석은 전처리부(113)에서 수행하는 것과 같은 방식으로 진행될 수 있다.

- [82] 증강 문서 선별부(115)는 신뢰도 분포 예측부(114)에서 사용자로부터 입력 받은 문서 내 복수 문장 집합에 상응하는 복수 검색 후보 문장 집합 내 각 문장을 통해 조합 가능한 복수 문서 집합과 각각에 대해 독자 그룹이 보일 것으로 예상하는 신뢰도 분포를 전달받고, 제어부(130)로부터 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 대한 제어 정보를 전달받아 미리 설정된 문서 중 제어 정보와 일치하는 문서를 신뢰도 증강 문서로서 선별한다.
- [83] 여기서, 미리 설정된 문서는 (1) 각 문장에 대해 예측된 신뢰도 분포의 평균을 해당하는 문서 내 모든 문장에 대해 평균한 값이 가장 높은 검색 후 문서와, (2) 각 문장에 대해 예측된 신뢰도 분포의 표준편차를 해당하는 문서 내 모든 문장에 대해 평균한 값이 가장 낮은 검색 후 문서와, (3) 각 문장에 대해 예측된 신뢰도 분포의 {평균/표준편차} 값을 해당하는 문서 내 모든 문장에 대해 평균한 값이 가장 높은 검색 후 문서를 포함할 수 있다.
- [84]
- [85] 도 3은 도 1에 도시된 출력부의 일 실시예 구성을 나타낸 것이다.
- [86] 도 3에 도시된 바와 같이, 출력부(140)는 증강 텍스트 출력부(141), 그래프 출력부(142) 및 사용자 상호작용부(143)를 포함한다.
- [87] 증강 텍스트 출력부(141)는 증강 문서 선별부(115)에 의해 선별된 신뢰도 증강 문서를 전달받고, 이를 사용자에게 출력 제공한다.
- [88] 그래프 출력부(142)는 증강 문서 선별부(115)로부터 선별된 신뢰도 증강 문서를 전달받고, (1) 이에 대한 신뢰도 분포 예측 결과, (2) 원본 문서, (3) 원본 문서에 대한 신뢰도 분포 예측 결과를 신뢰도 예측부(114)로부터 전달받아, 도 4에 도시된 일 예와 같이, 사용자에 의해 입력된 원본 문서 내 각 문장에 대해 예측된 신뢰도 분포와 신뢰도 증강부(110)에 의해 증강된 문서 내 각 문장에 대해 예측된 신뢰도 분포를 그래프 형태로 병렬 대조 비교하여 사용자에게 출력

제공한다.

- [89] 사용자 상호작용부(143)는 그래프 출력부(142)에 의해 사용자에게 출력된 신뢰도 분포 그래프에 대한 드래그 앤 드랍 방식으로 사용자와 상호작용할 수 있도록 하며, 사용자가 드래그 앤 드랍 방식을 통해 희망하는 신뢰도 분포의 양태를 사용자 상호작용부(143)에 입력하였을 때 입력 받은 신뢰도 분포와 가장 가까운 신뢰도 분포를 갖는 검색 후보를 선별하여 신뢰도 분포의 대상이 되는 문장을 해당하는 검색 후보로 변경 재출력하며, 해당하는 검색 후보의 신뢰도 분포로 기존에 사용자가 드래그 앤 드랍 방식으로 상호작용했던 그래프를 변경 재출력하여 제공한다.
- [90] 이 때, 사용자 상호작용부(143)는 그래프 출력부(142)에 의해 출력된 신뢰도 분포 그래프에서 사용자가 드래그 앤 드랍 방식을 통해 사용자가 희망하는 신뢰도 분포의 양태를 입력할 수 있도록 할 수 있다.
- [91] 즉, 사용자 상호작용부(143)는 사용자가 드래그 앤 드랍 방식을 통해 희망하는 신뢰도 분포의 양태를 사용자 상호작용부(143)에 입력하였을 때, 해당하는 문장의 검색 후보 집합 중 각 검색 후보 및 상응하는 예측 신뢰도 분포를 신뢰도 분포 예측부(114)로부터 전달받고, 분포 유사도 지표를 통해 검색 후보 중 사용자가 희망하는 신뢰도 분포의 양태와 가장 유사도가 높은 예측 신뢰도 분포를 갖는 검색 후보를 선별한다. 여기서, 분포 유사도는 Kullback-Leibler divergence나 Levy-Prokhorov metric으로 정의될 수 있다.
- [92]
- [93] 이와 같이, 본 발명의 실시예에 따른 방법은 주어진 문서에 대해 가능한 검색 후보군에 대한 자동 신뢰도 분포 예측을 수행하고, 원본 문서와 검색 후보군의 신뢰도 분포 각각이 보이는 평균과 표준편차 변화에 기반하여 신뢰도 증강을 수행할 수 있다.
- [94] 또한, 본 발명의 실시예에 따른 방법은 분포 개념의 도입이라는 필요조건이 만족되기 때문에 큰 개인차를 포함하는 신뢰도 개념에 대한 문서 검색을 가능하게 한다. 즉, 본 발명에 따른 방법은 (1) 신뢰도 분포의 높은 평균을 지향하는 방식의 신뢰도 증강을 진행하여 보다 많은 사람들이 신뢰도 증강의 대상이 되는 문서를 신뢰할 수 있도록 문서를 검색하는 것, (2) 신뢰도 분포의 낮은 표준편차를 지향하는 방식의 신뢰도 증강을 진행하여 신뢰도 증강의 대상이 되는 문서에 대한 신뢰 여부가 보다 논란의 여지가 없도록 검색하는 것, (3) 신뢰도 분포의 높은 {평균/표준편차}를 지향하는 방식의 신뢰도 증강을 진행하여 상기 두 방식의 결합된 방식으로 신뢰도 증강의 대상이 되는 문서를 검색하는 것을 가능하게 한다.
- [95] 또한, 본 발명의 실시예에 따른 방법은 교정 대상이 되는 문서 내의 각 문장에 대해 사용자가 드래그 앤 드랍 방식을 통해 희망하는 신뢰도 분포의 양태를 입력하고 해당하는 신뢰도 분포의 양태와 가장 유사한 신뢰도 분포 양태를 보이는 문장 검색 후보를 사용자에게 제공할 수 있다. 따라서 본 발명은 개인의

주관에 대해 큰 영향을 받기 때문에 자동화된 교정 시스템에 적용되기 어려웠던 신뢰도 증강에 대한 문서 교정을 분포 개념의 도입을 통해 가능하도록 하며, 이를 통해 불특정 타인 집단으로부터 보다 신뢰받기 위해 문서가 어떻게 변형되어야 하는 지에 대한 상세 정보를 다양한 형태로 제공받을 수 있다.

[96]

[97] 도 7은 도 1에 도시된 신뢰도 증강부에 의한 일 실시예의 동작 흐름도를 나타낸 것이다.

[98] 도 7을 참조하면, 신뢰도 증강부(110)에 의한 방법은 입력 단계(S310), 전처리 단계(S320), 신뢰도 분포 예측 단계(S330), 검색 후보 생성 단계(S340), 증거 문서 선별 단계(S350)를 포함한다.

[99] 입력 단계(S310)는 사용자로부터 신뢰도 증강의 대상이 되는 문서를 입력으로 받고, 선택적으로 사용자로부터 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식에 대한 제어 정보를 입력 받는 단계이다. 전처리 단계(S320)는 사용자로부터 입력 받은 문서 내 각 문장에 대한 전처리를 진행하는 단계이다. 검색 후보 생성 단계(S330)는 원본 문서 내 각 문장에 대해 복수 검색 후보 문장 집합을 생성하는 단계이다. 신뢰도 분포 예측 단계(S340)는 사용자로부터 입력 받은 문서 내 복수 문서 집합과 이에 상응하는 복수 검색 후보 문장 집합 내 각 문장을 통해 조합 가능한 복수 문서 집합 내 각 문장에 대해 독자 그룹이 보일 것으로 예상되는 신뢰도 분포를 예측하는 단계이다. 증거 문서 선별 단계(S350)는 사용자로부터 입력 받은 문서 내 복수 문장 집합에 상응하는 복수 검색 후보 문장 집합 내 각 문장을 통해 조합 가능한 복수 문서 집합과 각각에 대해 독자 그룹이 보일 것으로 예상되는 신뢰도 분포에 대한 분석을 통해 제어 정보에 따라 신뢰도 증강 문서를 선별하는 단계이다.

[100]

[101] 도 8은 도 1에 도시된 출력부에 의한 일 실시예의 동작 흐름도를 나타낸 것이다.

[102] 도 8을 참조하면, 출력부(140)에 의한 방법은 증강 텍스트 출력 단계(S410), 그래프 출력 단계(S420) 및 사용자 상호작용 단계(S430)를 포함한다.

[103] 증강 텍스트 출력 단계(S410)는 신뢰도 증강 문서를 사용자에게 출력 제공하는 단계이다. 그래프 출력 단계(S420)는 사용자에게 의해 입력된 원본 문서 내 각 문장에 대해 예측된 신뢰도 분포와 증강된 문서 내 각 문장에 대해 예측된 신뢰도 분포를 그래프 형태로 병렬 대조 비교하여 사용자에게 출력 제공하는 단계이다. 사용자 상호작용 단계(S350)는 사용자에게 출력된 신뢰도 분포 그래프에 대한 드래그 앤 드랍 방식으로 사용자와 상호작용할 수 있도록 하며, 사용자가 드래그 앤 드랍 방식을 통해 희망하는 신뢰도 분포의 양태를 입력하였을 때 입력 받은 신뢰도 분포와 가장 가까운 신뢰도 분포를 갖는 검색 후보를 선별하여 신뢰도 분포의 대상이 되는 문장을 해당하는 검색 후보로 변경 제출하며, 해당하는 검색 후보의 신뢰도 분포로 기존에 사용자가 드래그 앤 드랍 방식으로 상호작용했던 그래프를 변경 제출하여 제공하는 단계이다.

[104]

[105] 비록, 도 7과 8의 방법에서 그 설명이 생략되었더라도, 도 7과 8을 구성하는 각 단계는 도 1 내지 도 6에서 설명한 모든 내용을 포함할 수 있으며, 이는 이 기술 분야에 종사하는 당업자에게 있어서 자명하다.

[106]

[107] 이상에서 설명된 시스템 또는 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 시스템, 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[108]

소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[109]

실시예들에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어

당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

- [110] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [111] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

- [청구항 1] 사용자로부터 문서가 입력되면 상기 입력된 문서의 각 문장에 대한 검색 후보군을 생성하는 단계;
상기 입력된 문서와 상기 생성된 검색 후보군 각각에 대한 신뢰도 분포를 예측하는 단계; 및
상기 예측된 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식을 통해 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 단계를 포함하는 문서 신뢰도 증강 방법.
- [청구항 2] 제1항에 있어서,
상기 입력된 문서에 대해 신뢰도 증강된 신뢰도 증강 문서 및 상기 신뢰도 증강의 전과 후에 대한 신뢰도 분포 변화 정보를 상기 사용자에게 제공하는 단계를 더 포함하는 것을 특징으로 하는 문서 신뢰도 증강 방법.
- [청구항 3] 제2항에 있어서,
상기 사용자에게 제공하는 단계는
상기 사용자와의 상호작용을 통해 상기 사용자가 희망하는 신뢰도 분포의 양태에 가장 유사한 문서 검색 결과를 상기 사용자에게 더 제공하는 것을 특징으로 하는 문서 신뢰도 증강 방법.
- [청구항 4] 제1항에 있어서,
상기 복수 증강 방식은
상기 신뢰도 분포의 미리 설정된 기준 평균 이상의 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 상기 신뢰도 분포의 미리 설정된 기준 표준편차 이하의 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 상기 신뢰도 분포의 미리 설정된 기준 {평균/표준편차} 이상의 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함하는 것을 특징으로 하는 문서 신뢰도 증강 방법.
- [청구항 5] 제1항에 있어서,
상기 생성하는 단계는
상기 사용자로부터 상기 복수 증강 방식에 대한 제어 정보를 입력 받는 단계;
상기 입력된 문서의 각 문장에 대한 전처리를 수행하는 단계; 및
상기 입력된 문서의 각 문장에 대한 복수의 검색 후보 문장 집합을 생성하는 단계를 포함하고,
상기 예측하는 단계는
상기 입력된 문서의 각 문장과 상기 검색 후보 문장 집합 내의 각 문장을

통해 조합 가능한 복수의 문서 집합의 각 문장에 대한 신뢰도 분포를 예측하며,

상기 수행하는 단계는

상기 예측된 복수의 문서 집합의 각 문장에 대한 신뢰도 분포와 상기 제어 정보에 기초하여 신뢰도 증강 문서를 선별함으로써, 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 것을 특징으로 하는 문서 신뢰도 증강 방법.
 [청구항 6] 사용자로부터 문서가 입력되면 상기 입력된 문서의 각 문장에 대한 검색 후보군을 생성하는 생성부;

상기 입력된 문서와 상기 생성된 검색 후보군 각각에 대한 신뢰도 분포를 예측하는 예측부; 및

상기 예측된 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식을 통해 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 증강부를 포함하는 문서 신뢰도 증강 시스템.

[청구항 7]

제6항에 있어서,

상기 입력된 문서에 대해 신뢰도 증강된 신뢰도 증강 문서 및 상기 신뢰도 증강의 전과 후에 대한 신뢰도 분포 변화 정보를 상기 사용자에게 제공하는 출력부

를 더 포함하는 것을 특징으로 하는 문서 신뢰도 증강 시스템.

[청구항 8]

제7항에 있어서,

상기 출력부는

상기 사용자와의 상호작용을 통해 상기 사용자가 희망하는 신뢰도 분포의 양태에 가장 유사한 문서 검색 결과를 상기 사용자에게 더 제공하는 것을 특징으로 하는 문서 신뢰도 증강 시스템.

[청구항 9]

제6항에 있어서,

상기 복수 증강 방식은

상기 신뢰도 분포의 미리 설정된 기준 평균 이상의 평균을 지향하는 방식의 신뢰도 증강에 해당하는 제1 항목, 상기 신뢰도 분포의 미리 설정된 기준 표준편차 이하의 표준편차를 지향하는 방식의 신뢰도 증강에 해당하는 제2 항목 및 상기 신뢰도 분포의 미리 설정된 기준 {평균/표준편차} 이상의 {평균/표준편차}를 지향하는 방식의 신뢰도 증강에 해당하는 제3 항목 중 적어도 하나 이상의 항목을 포함하는 것을 특징으로 하는 문서 신뢰도 증강 시스템.

[청구항 10]

제6항에 있어서,

상기 생성부는

상기 사용자로부터 상기 복수 증강 방식에 대한 제어 정보를 입력 받고, 상기 입력된 문서의 각 문장에 대한 전처리를 수행하며, 상기 입력된 문서의 각 문장에 대한 복수의 검색 후보 문장 집합을 생성하고, 상기 예측부는

상기 입력된 문서의 각 문장과 상기 검색 후보 문장 집합 내의 각 문장을 통해 조합 가능한 복수의 문서 집합의 각 문장에 대한 신뢰도 분포를 예측하며,

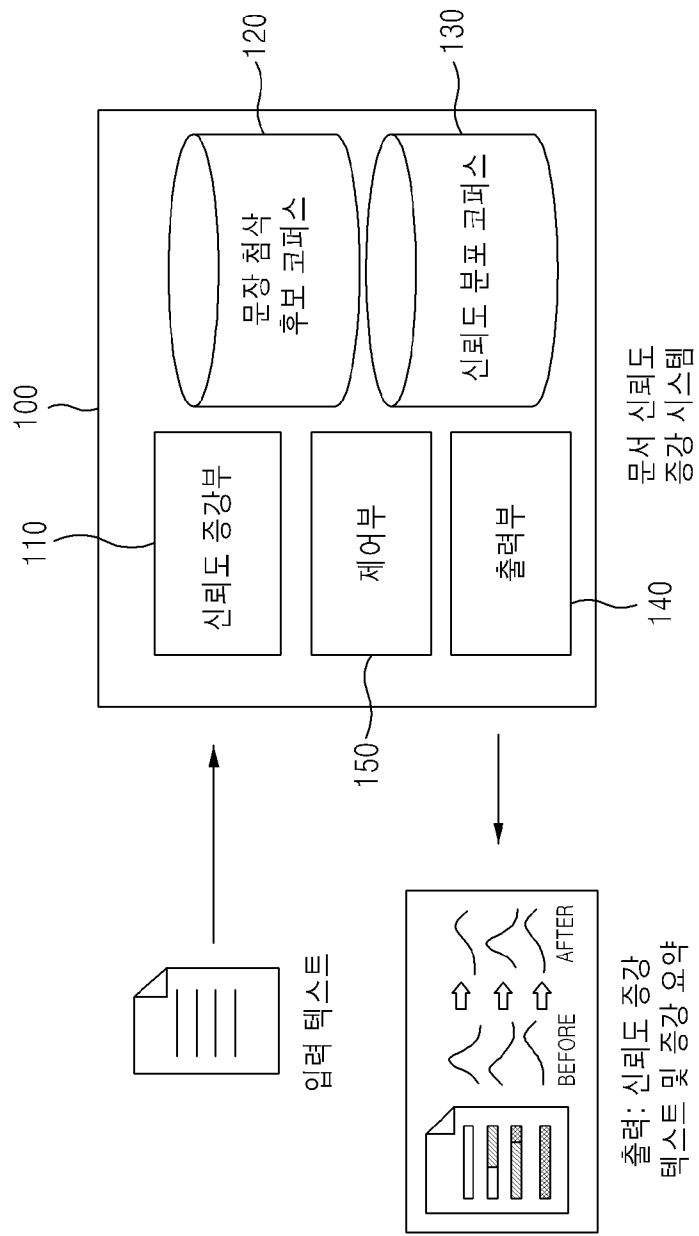
상기 증강부는

상기 예측된 복수의 문서 집합의 각 문장에 대한 신뢰도 분포와 상기 제어 정보에 기초하여 신뢰도 증강 문서를 선별함으로써, 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 것을 특징으로 하는 문서 신뢰도 증강 시스템.

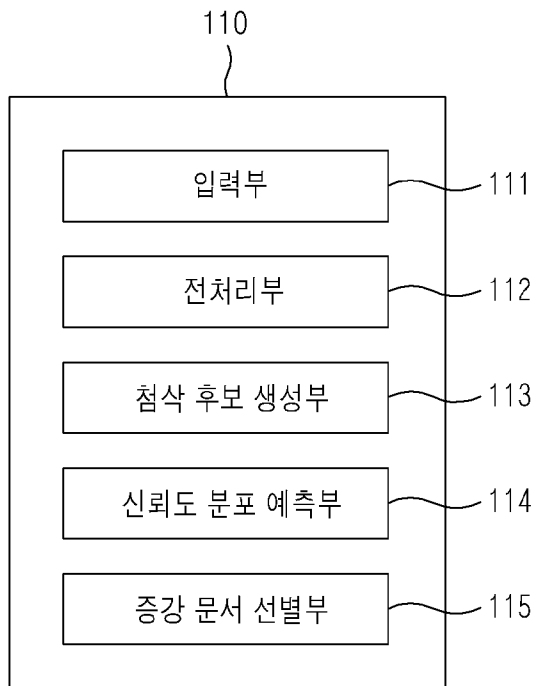
요약서

주어진 문서가 독자에게 보다 높은 신뢰를 받을 수 있도록 하는 문서 신뢰도 증강 방법 및 그 시스템이 개시된다. 본 발명의 일 실시예에 따른 문서 신뢰도 증강 방법은 사용자로부터 문서가 입력되면 상기 입력된 문서의 각 문장에 대한 침삭 후보군을 생성하는 단계; 상기 입력된 문서와 상기 생성된 침삭 후보군 각각에 대한 신뢰도 분포를 예측하는 단계; 및 상기 예측된 신뢰도 분포의 평균과 표준편차에 기반한 복수 증강 방식을 통해 상기 입력된 문서에 대한 신뢰도 증강을 수행하는 단계를 포함한다.

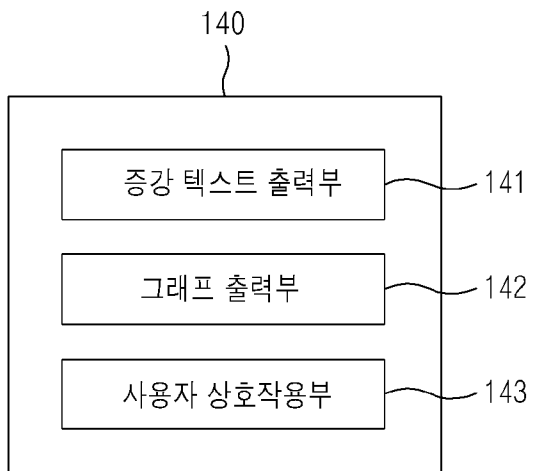
[도 1]



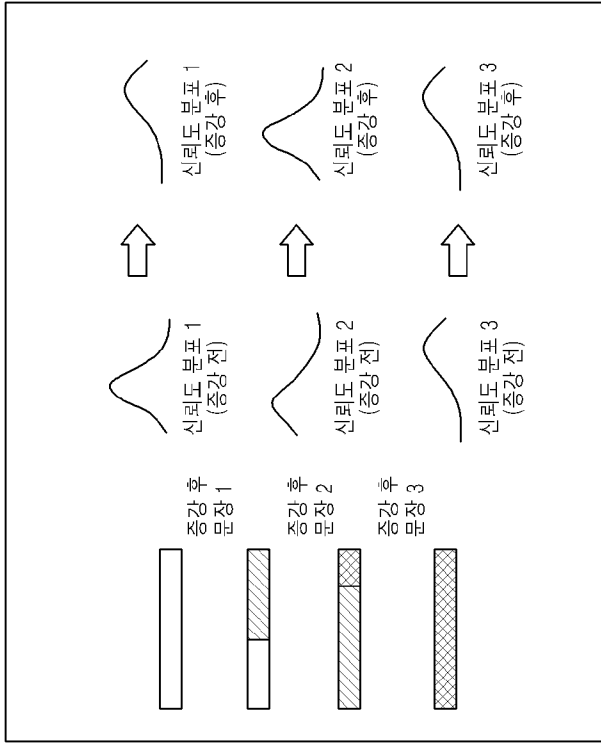
[도2]



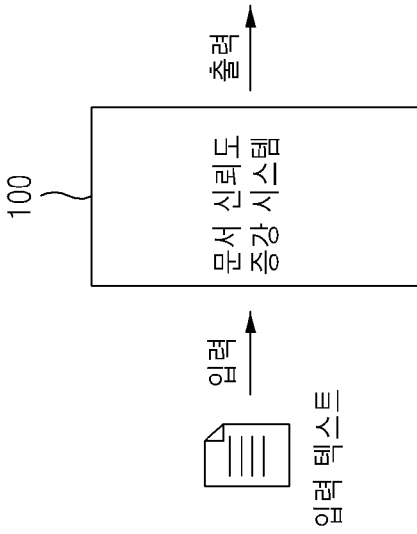
[도3]



[도4]



출력: 신뢰도 증강 테스트 및 증강 요약



[도5]

```

문장장 탐색 후보 코퍼스:{
  항목 1: {
    원본 문장: "예시 문장 1"
    탐색된 문장 1: {
      텍스트: "탐색 예시 1"
      원본 문장과의 유사도" 유사도_1
    }
    탐색된 문장 2: {
      텍스트: "탐색 예시 2"
      원본 문장과의 유사도: 유사도_2
    }
    ...
  }
  항목 2: {
    원본 문장: "예시 문장 2"
    탐색된 문장 1: {
      텍스트: "탐색 예시 3"
      원본 문장과의 유사도" 유사도_3
    }
    탐색된 문장 2: {
      텍스트: "탐색 예시 4"
      원본 문장과의 유사도: 유사도_4
    }
    ...
  }
  ...
}

```

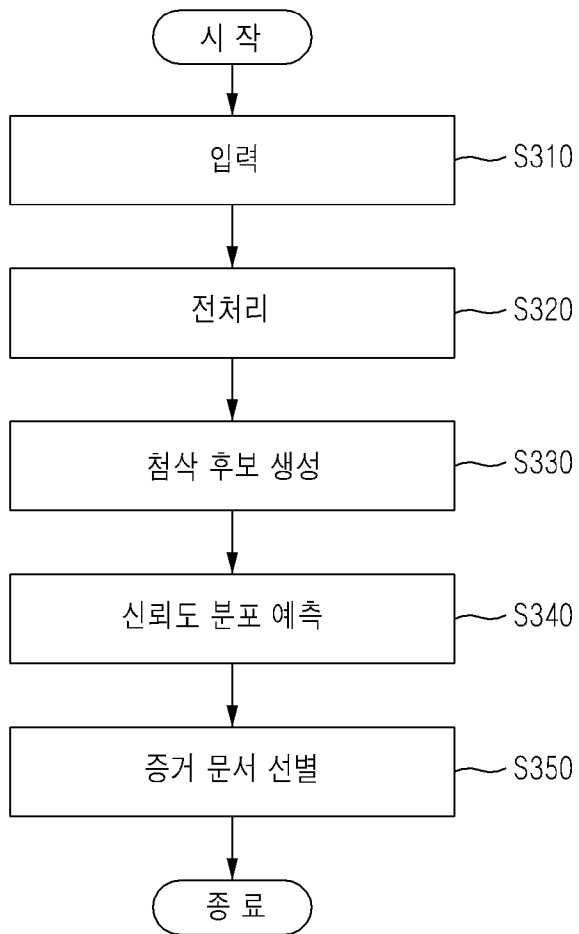
[도6]

```

신뢰도 분포 코퍼스:{
  문서 1:{
    문장 1:{
      텍스트: "예시 문장 1"
      신뢰도 분포: { -5점: g명, -4점: h명,
                    -3점: i명, ...0점: j명,
                    1점: k명, ...5점: l명
                  }
    }
  }
  문서 2:{
    텍스트: "예시 문장 2"
    신뢰도 분포: { -5점: m명, -4점: n명,
                  -3점: o명, ...0점: p명,
                  1점: q명, ...5점: r명
                }
  }
  ...
}
문서 2:{
  문장 1:{
    텍스트: "예시 문장 3"
    신뢰도 분포: { -5점: s명, -4점: t명,
                  -3점: u명, ...0점: v명,
                  1점: w명, ...5점: x명
                }
  }
}
문서 2:{
  텍스트: "예시 문장 4"
  신뢰도 분포: { -5점: y명, -4점: z명,
                -3점: a명, ...0점: b명,
                1점: c명, ...5점: d명
              }
}

```


[도7]



[도8]

