



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0036097
(43) 공개일자 2022년03월22일

(51) 국제특허분류(Int. Cl.)
G06F 40/58 (2020.01) G06F 40/284 (2020.01)
G06N 3/04 (2006.01) G06N 3/08 (2006.01)
(52) CPC특허분류
G06F 40/58 (2020.01)
G06F 40/284 (2020.01)
(21) 출원번호 10-2020-0118268
(22) 출원일자 2020년09월15일
심사청구일자 없음

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
신종훈
대전광역시 유성구 배울1로 13, 201동 2101호(관평동, 대덕테크노밸리2단지아파트)
(74) 대리인
특허법인지명

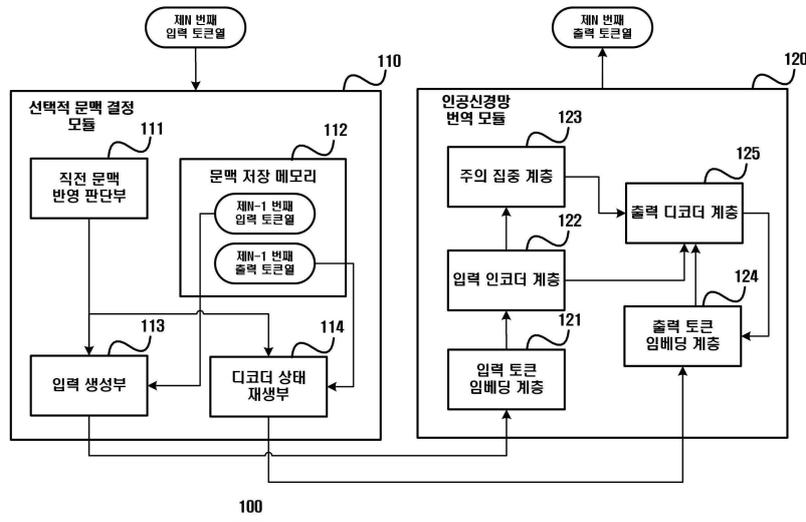
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 방법

(57) 요약

분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동 통역 방법이 제공된다. 상기 방법은 신규의 제N 번째(N은 자연수) 입력 토큰열을 입력받는 단계; 상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계; 및 상기 인공신경망 번역 모듈의 디코더를 통해 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하는 단계를 포함하되, 상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계는, 상기 제N번째 입력 토큰열과 함께, 메모리 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역 모듈로 전달 여부를 결정하는 단계; 및 상기 결정에 따라 상기 제N-1 번째 입력 토큰열과 상기 제N 번째 입력 토큰열을 순차적으로 결합하여 상기 인공신경망 번역 모듈로 전달하는 단계를 포함한다.

대표도 - 도2



- (52) CPC특허분류
G06N 3/04 (2013.01)
G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103258
과제번호	R7119-16-100
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발사업
연구과제명	지식증강형 실시간 동시통역 원천기술 개발
기 여 율	1/1
과제수행기관명	한국전자통신연구원
연구기간	2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

컴퓨터에 의해 수행되는 방법에 있어서,

신규의 제N 번째(N은 자연수) 입력 토큰열을 입력받는 단계;

상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계; 및

상기 인공신경망 번역 모듈의 디코더를 통해 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하는 단계를 포함하되,

상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계는,

상기 제N번째 입력 토큰열과 함께, 메모리 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역 모듈로 전달 여부를 결정하는 단계; 및

상기 결정에 따라 상기 제N-1 번째 입력 토큰열과 상기 제N 번째 입력 토큰열을 순차적으로 결합하여 상기 인공신경망 번역 모듈로 전달하는 단계를 포함하는,

분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동 통역 방법.

발명의 설명

기술분야

[0001] 본 발명은 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 방법에 관한 것이다.

배경기술

[0002] 번역 또는 통역을 위한 시스템은 사용자의 제1 언어를 인식하여 제2 언어로 번역 또는 통역을 수행한다. 이러한 번역 또는 통역을 위한 기술은 그 대상인 발화 신호 구간을 결정하는 과정이 완전히 이루어져야 온전한 결과가 제공될 수 있다.

[0003] 하지만, 실시간성이 요구되는 화자의 발화를 연속으로 자동통역시 단락된 문장을 대상으로 자동통역을 할 경우, 그 의미가 잘못 번역되는 문제가 발생하기도 한다.

[0004] 이러한 문제를 해소하기 위해 이미 번역했던 원문을 다시 입력하여 재번역하는 것을 고려할 수 있으나, 이는 결국 유사한 내용을 반복해서 전달하는 문제가 발생하게 되어, 제2 언어 사용자의 혼란을 불러 일으킨다는 문제가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 공개특허공보 제10-2019-0015081호(2019.02.13)

발명의 내용

해결하려는 과제

[0006] 본 발명이 해결하고자 하는 과제는 현재 입력 토큰열에 대한 번역을 수행함에 있어, 이전 번역 결과를 선택적으로 적용하여 보다 정확한 번역 결과를 생성할 수 있는, 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 방법을 제공하는 것이다.

[0007] 다만, 본 발명이 해결하고자 하는 과제는 상기된 바와 같은 과제로 한정되지 않으며, 또다른 과제들이 존재할 수 있다.

과제의 해결 수단

[0008] 상술한 과제를 해결하기 위한 본 발명의 일면에 따른 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동 통역 방법은 신규의 제N 번째(N은 자연수) 입력 토큰열을 입력받는 단계; 상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계; 및 상기 인공신경망 번역 모듈의 디코더를 통해 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하는 단계를 포함한다. 이때, 상기 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하는 단계는, 상기 제N번째 입력 토큰열과 함께, 메모리 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역 모듈로 전달 여부를 결정하는 단계; 및 상기 결정에 따라 상기 제N-1 번째 입력 토큰열과 상기 제N 번째 입력 토큰열을 순차적으로 결합하여 상기 인공신경망 번역 모듈로 전달하는 단계를 포함한다.

[0009] 본 발명의 일부 실시예에서, 상기 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역 모듈로 전달할지 여부를 결정하는 단계는, 상기 제N 번째 입력 토큰열 및 상기 제N-1 번째 입력 토큰열 중 미리 설정된 규칙을 만족하는 토큰이 존재하는지 여부에 기초하여 결정될 수 있다.

[0010] 본 발명의 일부 실시예에서, 상기 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역 모듈로 전달할지 여부를 결정하는 단계는, 상기 제N 번째 입력 토큰열 및 상기 제N-1 번째 입력 토큰열 중 미리 설정된 규칙으로, 상기 제N 번째 입력 토큰열이 미리 설정된 대명사로 시작되는 경우, 상기 제N-1 번째 입력 토큰열이 미리 설정된 부사격 조사 또는 연결 어미로 끝나는 경우 중 적어도 하나를 만족하는지 여부에 기초하여 결정될 수 있다.

[0011] 본 발명의 일부 실시예에서, 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하는 단계는, 상기 제N 번째 입력 토큰열과 상기 제N-1 번째 입력 토큰열이 결합되어 상기 인공신경망 번역 모듈로 전달되는 것으로 결정됨에 따라, 상기 제N-1 번째 출력 토큰열에 기반하여 상기 디코더의 상태를 재생성할 수 있다.

[0012] 본 발명의 일부 실시예에서, 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하는 단계는, 상기 제N-1 번째 출력 토큰열을 구성하는 n개의 각 토큰을 순차적으로 상기 디코더에 입력시켜, 상기 n개의 각 토큰 중 제n 번째 토큰이 상기 디코더에 입력됨에 따른 연산 결과에 기반하여 상기 디코더의 상태를 재생성할 수 있다.

[0013] 본 발명의 일부 실시예에서, 상기 재생성된 디코더는 상기 제N번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성하기 위한 최초 상태로 설정될 수 있다.

[0014] 본 발명의 일부 실시예에서, 상기 제N 번째 및 제N-1 번째 입력 토큰열이 결합되어 상기 인공신경망 번역 모듈로 전달됨에 따라 출력되는 전체 출력 토큰열은 상기 메모리 상에 저장될 수 있다.

[0015] 상술한 과제를 해결하기 위한 본 발명의 다른 면에 따른 컴퓨터 프로그램은, 하드웨어인 컴퓨터와 결합되어 상기 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 방법을 실행하며, 컴퓨터 판독가능 기록매체에 저장된다.

[0016] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

[0017] 상술한 본 발명에 의하면, 문장이 완성되는 구간 또는 음성인식 서버 시스템 이후 하위 서버 시스템의 처리 단위에 맞게 입력이 분절되었을 때 불가피하게 발생할 수 있는 문맥 단락 문제를 해소할 수 있다.

[0018] 또한, 문맥이 변화되었음에도 불구하고, 이미 컴퓨터 시스템의 모니터와 같은 화면 출력 장치나 오디오로 출력된 변화를 바꾸지 않아도 되는바 최적의 번역 결과를 이끌어낼 수 있다.

[0019] 이와 더불어, 문맥의 변화 반영과 분절에 의한 성능 하락의 문제가 감소하게 되어, 더욱 빠르게 미완성된 입력에 대한 통역 결과를 출력할 수 있으며, 이에 따라 제2 언어 사용자의 편의성을 증대시킬 수 있다.

[0020] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0021] 도 1은 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치의 블럭도이다.
- 도 2는 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치의 기능 블럭도이다.
- 도 3은 본 발명의 일 실시예에 따른 신경망 기반 자동통역 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0022] 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 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0023] 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 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.
- [0024] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 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.
- [0025] 본 발명은 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 장치 및 방법에 관한 것이다.
- [0026] 본 발명은 자동통역 기술에 관련된 것으로, 보다 구체적으로는 자동 음성인식 서버 시스템을 통해 음성 신호를 같은 언어 형식의 텍스트로 변환하는 과정과 관련된 기술을 바탕으로 한다. 또는, 컴퓨터의 자판(keyboard)를 사용하여 텍스트를 직접 입력하는 방식으로 제1 언어(시작 언어)의 문장을 입력받고, 텍스트를 자동번역 서버 시스템을 통해 제2 언어(도착 언어)로 구성된 텍스트로 번역하여 사용자에게 그 결과를 출력하거나, 제2 언어의 텍스트를 오디오 신호로 변환하여 출력하는 TTS(Text-to-Speech; 음성 합성기)를 통해 오디오로 결과를 출력하는 기술을 바탕으로 한다.
- [0027] 종래 기술에 따른 자동통역 시스템은 기본적으로 음성의 발화 시작과 끝을 명시적으로 지정하는 사용자 인터페이스(물리적 버튼 또는 화면 내 지시 버튼 등)가 존재하며, 통역을 원하는 발화 내용을 말하기 전에 시작 버튼을 누르거나, 종료 여부를 명시적으로 전달하기 위해 발화 내용이 끝나는 대로 종료 버튼을 누르는 경우(또는 반대로, 누르고 있던 버튼을 손에서 떼는 경우)를 많이 사용한다. 이러한 인터페이스를 푸시-투-토크(Push-to-Talk; PTT)라 칭한다.
- [0028] 만약, 이와 같은 인터페이스에서 벗어나 발화의 시작과 종료를 인식하는데 필요한 명시적 행동(예를 들어, 버튼 누름, 버튼 뽐)을 생략하기 위해, 마이크와 같은 수음(Audio input) 장치를 통해 입력된 신호를 기준으로, 그 신호가 상대적으로 약해진 구간의 길이를 검출(음성 활동 검출, Voice Activity Detection; VAD)하는 행동을 바탕으로 하는 음성 끝점 검출(End-point Detection; EPD) 기술을 사용하기도 한다.
- [0029] 이러한 종래 기술 기반의 자동통역 시스템은, 발화의 시작과 끝을 직간접적으로 인지하여 통역하고자 하는 발화 신호의 구간을 결정하는 과정이 완전히 이루어져야 인식 결과가 완성되며, 그 결과 텍스트를 자동번역 서버 시스템에 전달하게 되는 계단식(cascade) 구성을 갖는다.
- [0030] 한편, 음성 입력이 아닌 경우도 존재하는데, 실시간으로 대화 내용을 기록하는 속기 환경과 같이 번역되어야 할 텍스트를 순차적으로 입력하는 예도 음성을 계속하여 입력받는 것과 같은 환경으로 이해할 수 있다.
- [0031] 위 두 가지 환경에서는, 통역 대상이 되는 전체 입력이 "완료"되기 전에 그보다 더 빨리 결과를 화면이나 음성

합성 시스템 또는 기기에 의해 조합된 음성으로 출력되기를 원하는 경우가 존재한다. 이는 통번역 결과를 받아들이는 제2 언어 사용자가 얼마나 결과를 적시에 실시간으로 잘 전달받을 수 있을지를 기준으로 통역 품질에 영향을 미치게 된다. 발화가 쌍방으로 이루어지는 양자 간 대화보다는 강연과 같이 단일 화자가 일방적으로 다수의 문장을 발화해야 하고, 이를 다른 언어로 출력해야 하는 경우를 가정해보면 쉽게 이해가 가능할 것이다.

- [0032] 이처럼 실시간으로 화자가 말한 내용을 연속으로 자동통역하기 위해, 상기에 소개된 음성 끝점 검출(EPD) 장치와 같은 음성 활동 검출 장치를 통한 음성인식 시스템은, 그 구간의 음성 신호 열(Audio sequence)을 입력으로 보고, 확률이 가장 높은 텍스트 열(text sequence)을 예측하게 된다.
- [0033] 문제는, 음성 끝점 검출 장치와 같이 음성 신호의 활동만 감지하는 모듈의 사용 등에 의해, 화자가 발화한 문장의 실제 구성 및 완료 상태와는 상관없이, 주저함(Hesitation) 등으로 인한 긴 휴지(Long pause)가 삽입되어 실제 문장의 완료와 상관없이 발화 종료나 문장의 완료로 인식될 수 있다.
- [0034] 이 과정에서, 실제 화자가 표현하고자 하는 문장과 달리, 의도하지 않은 단락이 발생할 수 있다. 아래에서는 이러한 단락의 예시를 기술하였다:
- [0035] (1) 2400년 전 살았던 플라톤이라는 수학자는 국가론이라는 책에서
- [0036] (2) 기게스라는 한 목동이 우연히 반지를 발견하는데
- [0037] (3) 그 반지를 손가락에 끼우면 투명인간이 된다는 전설을 이야기합니다
- [0038] 번역 결과를 받아들이는 제2 언어 사용자는 위와 같은 단락마다의 번역 결과를 수신하게 되면서 빠르게 화자가 말하는 대상을 이해할 수 있다. 하지만, 자동 번역 모듈 및 서버 시스템에서는 이전에 번역되어 있던 내용과 상관없이, 현재 주어진 단락에 의해서만 최선의 번역 결과를 제시하게 된다. 이 과정에서 번역 품질이 나빠지는 경우가 발생한다.
- [0039] 개별 단락을 번역하면 다음과 같은 예시 결과를 얻을 수 있다:
- [0040] 단락별 번역 결과 예시)
- [0041] (1) Plato, a mathmatician who lived 2400 years ago, in book of called 'The republic'
- [0042] (2) A shepherd named Giges finds the ring by accident
- [0043] (3) The legend tells us that when you put the ring on your finger, you become an invisible person.
- [0044] 한편, 단락의 구분 없이 전체를 한번에 번역하는 경우 다음의 결과를 얻을 수 있다:
- [0045] 전체 번역 결과 예시) A mathematician named Plato, who lived 2,400 years ago, tells the legend that a shepherd named Giges inadvertently finds a ring in a book called 'The republic', and when he puts the ring on his finger, he becomes an invisible man.
- [0046] 이때, 단락 (3)의 번역 결과와 같이, 개별 문장의 문맥이 사라짐으로 인해 반지를 끼우는 대상이 달라짐을 알 수 있다. 즉, 전체 문장 번역에서는 단락 (2) 에서 나타나는 문맥, 즉, 목동인 'he' 를 주어로 보고 번역하였으나, 분리된 경우에는 주어를 당신 'you' 로 판단하여 번역 결과가 제시됨을 알 수 있다. 이처럼, 분절로 인한 단락이 번역 결과에 악영향을 미치게 된다.
- [0047] 이를 개선하는 방법으로, 이미 번역했던 원문을 다시 입력하여 재번역하는 경우를 살펴보기로 한다.
- [0048] (1) 단락 번역 예시 결과) Plato, a mathmatician who lived 2400 years ago, in book of called 'The republic'
- [0049] (1+2) 단락 번역 예시 결과) Plato, a mathematician who lived 2400 years ago, happened to find a ring by a shepherd named Giges in a book called 'The republic',
- [0050] (1+2+3)을 통합 번역한 경우는 전체 번역 결과 예시와 같다. 문맥이 추가됨으로 인해, 이전 전체 번역 예시와 다름을 확인할 수 있다. 어순의 위치가 달라지거나, 부족한 의미가 채워지면서 바뀌는 경우도 존재한다.
- [0051] 이처럼 이미 출력했던 문장에 변경이 발생하게 되는데, 이를 실시간으로 통역을 수행하는 환경에 적용할 경우, 번역된 결과물은 음성 합성 시스템을 통해 이미 소리로 제2 언어 사용자에게 전달된 바 있으나, 다음 통역 결과물을 전달해야 하는데 있어 어떤 부분이 변경사항이고, 새로운 내용 중 어떤 부분을 버릴지 결정할 수 없어, 결

국 유사한 내용을 반복해서 다르게 전달하는 문제가 발생하게 된다. 이는 제2 언어 사용자의 혼란을 불러일으킨다는 점에서, 해소되어야 할 문제점으로 인식될 수 있다.

- [0052] 이와 달리, 본 발명의 일 실시예는 이전 번역 결과를 유지한 상태에서 문맥의 변화를 선택적으로 적용함으로써, 음성, 텍스트 단위의 분절 과정이 사용되는 실시간 자동통역 시스템에서 분절에 의한 성능 하락을 개선할 수 있다.
- [0053] 본 발명의 일 실시예는 명시적으로 발화의 끝임을 알려주지 않거나 알려주기 어려운 연속 발화 환경(예를 들어, 강의 및 회의)에서, 사용자가 별도의 인터페이스를 통해 발화의 끝임을 명시하기 전에 실시간으로 자동통역을 수행하는 시스템에 적용될 수 있다.
- [0054] 이하, 첨부된 도면을 참조하여 본 발명의 실시예에 따른 분절에 의한 문맥 소실을 방지하는 신경망 기반 실시간 자동통역 장치 및 방법(이하, 신경망 기반 실시간 자동통역 장치 및 방법)을 상세하게 설명한다.
- [0055] 도 1은 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치(100)의 블록도이다.
- [0056] 도 1을 참조하면, 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치(100)는 통신모듈(110), 메모리(120) 및 프로세서(130)를 포함한다.
- [0057] 통신모듈(110)은 사용자 단말 등과 송수신하여 입력 토큰열을 수신하고, 프로세서(130)에 의해 처리된 출력 토큰열을 사용자 단말로 제공한다.
- [0058] 메모리(120)에는 제1 언어의 음성을 인식하여 제2 언어의 음성으로 자동통역을 수행하며, 이전 입력 및 출력 토큰열의 적용 여부를 결정하여 현재 입력 토큰열에 대한 출력 토큰열을 생성하기 위한 프로그램이 저장된다.
- [0059] 프로세서(130)는 메모리(120)에 저장된 프로그램을 실행시킴에 따라, 신규의 제N 번째 입력 토큰열을 입력받으면, 제N 번째 입력 토큰열을 인공신경망 번역 모듈로 전달하여, 디코더를 통해 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성한다. 이때, 프로세서(130)는 제N번째 입력 토큰열과 함께, 메모리(120) 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공신경망 번역모듈로 전달 여부를 결정하고, 결정 여부에 따라 제N-1 번째 입력 토큰열과 제N 번째 입력 토큰열을 순차적으로 결합하여 인공신경망 번역 모듈로 전달한다.
- [0060] 한편, 인공신경망 번역 모듈은 일 실시예로 장단기 메모리(Long-short Term Memory 또는 LSTM) 및 GRU(Gated Recurrent Unit)을 사용하는 순환 신경망 구조가 적용될 수 있으며, 또는 컨볼루션 신경망(CNN) 또는 Transformer 기반의 모듈이 적용될 수 있다.
- [0061] 도 2는 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치(100)의 기능 블록도이다.
- [0062] 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 장치(100)는 기능적으로 선택적 문맥 결정 모듈(110)과, 인코더(120a) 및 디코더(120b)로 구성되는 인공신경망 번역 모듈(120)로 구성될 수 있다.
- [0063] 선택적 문맥 결정 모듈(110)은 신규의 제N 번째 입력 토큰열을 입력받았을 때, 제N 번째 입력 토큰열과 함께, 문맥 저장 메모리(112) 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1 번째 출력 토큰열을 인공신경망 번역 모듈(120)로 전달할지 여부를 결정한다. 만약 전달하기로 결정한 경우, 제N-1 번째 입력 토큰열과 제N 번째 입력 토큰열을 순차적으로 결합하여 인공신경망 번역 모듈(120)로 전달하는 역할을 수행한다.
- [0064] 통역이 개시되었을 때, 선택적 문맥 결정 모듈(110)에 최초로 입력되는 제1 언어에 대한 입력 토큰열은 이전 저장된 토큰열이 없으므로, 입력 생성부(113)를 통해 추가적인 입력에 대한 변화 없이 번역을 위해 인공신경망 번역 모듈(120)로 곧바로 전달된다.
- [0065] 최초 입력된 입력 토큰열은 인공신경망 번역 모듈(120)의 인코더(120a)로 입력되며, 인코더(120a)의 입력 토큰 임베딩 계층(121)을 통해 벡터 형태로 치환되고, 치환된 출력은 입력 인코더 계층(122)과 주의 집중 계층(123)을 거쳐 추상화된 벡터 형태로 출력된다.
- [0066] 인공신경망 번역 모듈(120)은 제2 언어로 구성되는 출력 토큰열을 획득하기 위하여, 디코더(120b)의 구성인 출력 디코더 계층(125)을 통해 출력된 개별 토큰을 다시 출력 토큰 임베딩 계층(124)으로 입력시키는 재귀적인 구동을 수행한다. 이를 자동 재귀(Auto-regressive) 디코딩이라 한다.
- [0067] 이와 같은 구조를 통해 디코더(120b)는 입력 인코더 계층(122)과 주의 집중 계층(123)뿐만 아니라, 출력 디코더 계층(125)의 상태, 정확하게는 이전에 출력했던 어휘 토큰을 생성하면서 그 과정에서 산출된 상태 벡터를 유지

또는 생성해야 한다. 또한, 디코더(120b)는 최초 출력 토큰을 생성하기 위해 기 정의된(pre-defined) 예약 토큰(예를 들어, <start>)을 입력하여 출력 토큰 임베딩 계층(121)을 거쳐 출력 디코더 계층(125)을 통해 실제 출력되는 첫 번째 토큰을 생성하게 된다.

- [0068] 두 번째 토큰의 생성 지점에서는 첫 번째 토큰을 생성하는 과정에서 변경된 상태 벡터를 통해 다음 출력 토큰을 다시 생성하게 된다. 만약, 출력 토큰열이 총 10개의 토큰으로 구성된 경우, 디코더(120b)는 '토큰수 +1'회만큼 동작하여 출력이 완성되었다는 기 정의된 토큰(예를 들어, <end>)을 발견하면 그 동작을 완료해야 한다.
- [0069] 이와 같은 방법에 따라 최초 입력 토큰열에 대한 출력 토큰열이 생성되면, 입력 토큰열과 이에 상응하는 출력 토큰열은 문맥 저장 메모리(112) 상에 저장된다. 이후, 추가적인 N번째(N은 자연수) 입력 토큰열이 주어지는 경우, 기 저장된 제N-1 번째 입력 토큰열과 이에 상응하는 출력 토큰열을 직전 문맥 반영 판단부(111)의 결정을 기준으로 반영하게 된다.
- [0070] 직전 문맥 반영 판단부(111)는 언어와 환경에 따라 그 세부 구현이 달라질 수 있으나, 기본적으로는 신규로 입력된 제N 번째 입력 토큰열과 현재 문맥 저장 메모리(112) 상에 저장된 제N-1 번째 입력 토큰열을 조건 변수로 입력받고, 이전 문맥의 반영을 '수행한다/수행하지 않는다'를 구분하는 이진 분류기(binary classifier) 형태의 출력을 수행하는 것을 특징으로 한다.
- [0071] 일 실시예로 직전 문맥 반영 판단부(111)는 규칙 기반 분류기에 기초하여, 제N 번째 입력 토큰열 및 제N-1 번째 입력 토큰열 중 미리 설정된 규칙을 만족하는 토큰이 존재하는지 여부에 기초하여 제N-1 번째 출력 토큰열을 인공신경망 번역 모듈(120)로 전달할지 여부를 결정할 수 있다.
- [0072] 예를 들어, 직전 문맥 반영 판단부(111)는 제N 번째 입력 토큰열 및 제N-1 번째 입력 토큰열 중 미리 설정된 규칙으로, 제N 번째 입력 토큰열이 미리 설정된 대명사로 시작되는 경우, 제N-1 번째 입력 토큰열이 미리 설정된 부사격 조사 또는 연결 어미로 끝나는 경우 중 적어도 하나를 만족하는지 여부에 기초하여 제N-1 번째 출력 토큰열을 인공신경망 번역 모듈(120)로 전달할지 여부를 결정할 수 있다.
- [0073] 즉, 직전 문맥 반영 판단부(111)는 신규로 입력된 제N 번째 입력 토큰열 또는 제N-1 번째 입력 토큰열이 하기의 조건 중 하나 이상을 만족하는 경우 제N-1 번째 출력 토큰열을 인공신경망 번역 모듈(120)로 전달할 수 있다.
- [0074] [규칙 1] 직전 토큰열이 ~에서, ~에게서, ~에, ~에게, ~한테와 같은 부사격 조사로 끝나는 경우
- [0075] [규칙 2] 직전 토큰열이 ~은데와 같은 연결어미로 끝나는 경우
- [0076] [규칙 3] 현재 토큰열이 '이것', '그것', '저것' 과 같이 대명사로 시작되는 경우
- [0077] 위 예시에서 기록되어 있는 규칙의 표현은 정규 표현식(Regular Expression) 처리 모듈을 사용하여 구현하거나, 형태소 분석 또는 부-어휘 단위 토큰화(Sub-word tokenization)을 통해서 자주 나타나는 고빈도의 N-문자를 분리함으로써 조사나 어미, 대명사의 사용을 감지하는 방식으로 구현된다.
- [0078] 뿐만 아니라, 직전 문맥 반영 판단부(111)는 데이터 기반의 지도학습(supervised learning) 방법을 사용하는 이진 분류 모델을 통해서 제N-1 번째 출력 토큰열을 인공신경망 번역 모듈(120)로 전달할지 여부를 결정할 수 있다. 즉, 규칙 기반 대신 통계 기반 또는 기계학습 기반 모델을 사용하는 경우, 제N 및 제N-1 번째 입력 토큰열 전체 또는 두 입력 토큰열로부터 추출한 자질 집합(feature set)과 함께, 참/거짓이나 0과 1 같은 이진 표현 형태로 데이터 집합을 구축하여, 이진 분류 모델을 학습할 수 있다.
- [0079] 또한, 직전 문맥 반영 판단부(111)는 전술한 규칙 기반 방법과 기계학습 기반 모델을 혼합하여 앙상블(Ensemble)하여 목적하는 바를 달성할 수도 있다.
- [0080] 위 규칙을 통해 직전 문맥 반영 판단부(111)는 아래의 입력들을 다음과 같이 처리하게 된다.
- [0081] <예시 1>
- [0082] (직전 입력) 2400년 전 살았던 플라톤이라는 수학자는 국가론이라는 책에서
- [0083] (현재 입력) 기계스라는 한 목동이 우연히 반지를 발견하는데
- [0084] 위와 같은 입력이 들어오게 되면, 직전 문맥 반영 판단부(111)는 [규칙 1]에 의해 문맥이 이어져야 한다는 결정을 내리게 된다. 한편, 아래와 같은 예시가 입력되어 있는 경우,
- [0085] <예시 2>

- [0086] (직전 입력) 우리는 잘못된 결론을 이끌어내었습니다.
- [0087] (현재 입력) 부분적으로는 아무런 이상이 없더라도, 조화가 맞지 않으면 안됩니다.
- [0088] 직전 문맥 반영 판단부(111)는 위 규칙에서 두 문장이 독립적인 것임을 판단하고, 현재 입력만 사용하도록 입력 생성부(113)와 디코더 상태 재생부(114)에게 현재 문맥만 전송하고, 번역 결과 생성을 위한 첫 토큰을 지정하도록 명령을 전달하게 된다.
- [0089] 이후, 입력 생성부(113)는 직전 문맥 반영 판단부(111)의 결정에 따라 이후 절차를 수행하게 된다.
- [0090] 일 실시예로, 직전 문맥 반영 판단부(111)가 제N-1 번째 입력 토큰열과 관련된 문맥을 반영하기로 결정한 경우, 입력 생성부(113)는 제N 번째 입력 토큰열과 제N-1 번째 입력 토큰열을 결합(concatenate)한다. 이때, 입력 생성부(113)는 직전-현재의 순과 같이 제N-1 번째 입력 토큰열과 제N 번째 입력 토큰열을 순차적으로 결합하여 인공신경망 번역 모듈(120)로 전달할 입력 토큰열을 완성한다.
- [0091] 이와 달리, 직전 문맥 반영 판단부(111)가 직전 문맥을 반영하지 않는 것으로 결정한 경우, 입력 생성부(113)는 현재 입력 토큰열인 제N 번째 입력 토큰열만 인공신경망 번역 모듈(120)로 전달한다.
- [0092] 일 예로, <예시 1>의 경우에는 입력 생성부(113)가 제N-1 번째 및 제N 번째 입력 토큰열을 결합한 결과, "[2400년 전 살았던 플라톤이라는 수학자는 국가론이라는 책에서 기게스라는 한 목동이 우연히 반지를 발견하는데]"의 입력 토큰열이 생성되며, 이는 인공신경망 번역 모듈(120)로 전달된다.
- [0093] <예시 2>의 경우 입력 생성부(113)는 현재 입력인 "[부분적으로는 아무런 이상이 없더라도, 조화가 맞지 않으면 안됩니다.]" 제N 번째 입력 토큰열만 인공신경망 번역 모듈(120)로 전달한다.
- [0094] 다음으로, 디코더 상태 재생부(114)는 이미 출력된 문장을 다시 바꾸지 않고 새로운 문맥을 반영한 결과를 획득하는데 필요하다. 즉, 디코더 상태 재생부(114)는 인공신경망 번역 모듈(120)의 구성 중 디코더(120b) 부분에 해당하는 출력 토큰 임베딩 계층(124)과 출력 디코딩 계층(125)에 데이터를 전달하여, 직전 번역 결과인 제N-1 번째 출력 토큰열을 이용하여 디코더(120b)의 새로운 상태가 재생성되도록 한다.
- [0095] 구체적으로, 문맥 저장 메모리(112)에 저장된 제N-1 번째 출력 토큰열의 길이가 n개($1 \leq m \leq n$)라고 하면, 디코더 상태 재생부(114)는 첫 번째 토큰($m=1$)부터 마지막 토큰($m=n$)까지 순차적으로 디코더(120b)에 전달하며, 디코더(120b)에서는 이를 수신하여 순차적으로 연산을 수행하게 된다. 이후, 마지막 토큰인 제n 번째 토큰이 디코더(120b)에 입력됨에 따른 연산 결과에 기초하여 디코더(120b)의 상태가 재생성된다. 즉, 프로세서(130)가 n번의 디코더(120b) 연산을 수행한 결과가, 새로 들어올 제N 번째 입력 토큰열을 위한 번역을 수행하기 위한 최초 상태가 된다.
- [0096] 이는 종래 기술에서 새로 들어온 입력 토큰열의 번역을 위한 디코더(120b)의 첫 상태 생성을 위해 기 지정된 시작 토큰(예를 들어, <start>)을 입력하여 생성하는 것과는 다른 구성을 갖는다. 또한, m-1번째 토큰을 직접 디코더(120b)에 전달 및 실행하여 그 결과로 추출된 m번째 토큰 결과는 무시하고, 직전 번역 결과인 m번째 토큰을 사용하여 강제로 상태를 생성한다는 점에서 종래 기술과는 다른 구성을 갖는다. 이러한 과정을 통해 본 발명의 일 실시예는 늘어난 문맥에 따라 번역 결과가 상이해지는 것을 방지할 수 있다.
- [0097] 만약, 직전 문맥 반영 판단부(111)가 직전 문맥을 반영하지 않는 것으로 결정한 경우에는, 종래 기술에서 디코더(120b)의 첫 상태 생성을 처리하는 방법과 동일하게 기 지정된 시작 토큰을 입력하여 생성하게 된다.
- [0098] 즉, <예시 2>의 경우는 번역을 생성하기 위해 디코더(120b) 상태를 초기화할 때, 디코더(120b)의 시작을 알리는 예약 토큰(예를 들어, <start>)을 입력하여 디코더(120b)의 상태를 생성하고, 그 예측 결과를 재귀적으로 입력하여 다른 토큰을 구해 최종적으로 입력문에 대한 번역 결과를 얻게 된다.
- [0099] <예시 2>에서의 현재 입력을 번역한 결과로, "[Even if there is no abnormality in part, harmony must be achieved.]"가 출력되었다면, 출력 토큰열은 입력에 입력 토큰열과 함께 문맥 저장 메모리(112)에 저장된다. 즉, 제N 번째 입력 및 출력 토큰열은 제N-1 번째 입력 및 출력 토큰열이 저장되어 있던 위치에 저장되어, 다음 새 문장인 제N+1 번째 입력 토큰열이 입력되었을 때 사용될 수 있다.
- [0100] 이와 달리, 문맥을 반영해야 하는 <예시 1>의 경우에는, 문맥 저장 메모리(112)에 저장된 제N-1 번째 입력 토큰열로 "[2400년 전 살았던 플라톤이라는 수학자는 국가론이라는 책에서]"가 저장되어 있고, 그 출력 토큰열로 "[Plato, a mathematician who lived 2400 years ago, in book of called 'The republic']"가 저장된

상태이다.

- [0101] 인공지능망 번역 모듈(120)의 인코더 부분은 입력 생성부(113)에 의해 제N-1 번째 입력 토큰열과 제N 번째 입력 토큰열이 결합된 문장이 전달됨에 따라 추상화된 벡터 형태를 출력하게 되며, 디코더 상태 재생부(114)는 제N-1 번째 출력 토큰열을 이용하여 디코더(120b)의 상태를 재생성한다. 제N-1 번째 출력 토큰열의 길이에 맞게 디코더(120b)의 동작이 완료되면, 디코더(120b)가 예측하는 출력 토큰은 무시하지 않고 다시 전체 번역 결과에 붙여서 탐색 작업을 시행한다.
- [0102] 최종적으로 문장 생성의 종료를 알리는 기 정의된 토큰(예를 들어, <end>)가 나타나면 디코딩(120b)을 완료하고, 그 번역 결과인 출력 토큰열을 컴퓨터의 모니터나 음성 합성 장치(Text-to-Speech)를 통해 오디오, 스피커로 출력한다.
- [0103] 그리고 제N 번째 및 제N-1 번째 입력 토큰열이 결합되어 출력되는 전체 출력 토큰열은 다시 문맥 저장 메모리(112) 상에 저장되어 활용된다.
- [0104] 즉, 제N 번째 및 제N-1 번째 입력 토큰열이 결합되어 출력되는 출력 토큰열은 "Plato, a mathematician who lived 2400 years ago, in book of called 'The republic', a shepherd named Ghigess accidentally discovered a ring,"과 같이 'a shepherd' 이전까지 기존 결과와 동일함을 알 수 있으며, 이 결과는 문맥 저장 메모리(112)의 이전 출력 토큰열로 저장된다. 여기에서 볼드체는 문맥 저장 메모리(112)에 저장된 직전의 출력 토큰열을 나타낸다.
- [0105] 그리고 이때의 입력인 제N 번째 및 제N-1 번째 입력 토큰열이 결합된 입력 토큰열 "[2400년 전 살았던 플라톤이라는 수학자는 국가론이라는 책에서 기게스라는 한 목동이 우연히 반지를 발견하는데]" 역시 문맥 저장 메모리(112)의 이전 입력 토큰열로 저장된다.
- [0106] 이후, 다음 입력 토큰열인 제N+1 번째 입력 토큰열로 "[그 반지를 손가락에 끼우면 투명인간이 된다는 전설을 이야기합니다]"가 입력되면, 직전 문맥 반영 판단부(111)는 [규칙 2]에 의해 직전 문맥을 반영하도록 결정하고, 전술한 과정에 따라 입력 생성부(113)는 문맥 저장 메모리(112)에 저장된 입력 토큰열과 결합하고, 디코더 상태 재생부(114)는 문맥 저장 메모리(112)에 저장된 출력 토큰열을 그대로 사용하여 디코더(120b)의 상태를 재생성하는 방법으로 다음과 같은 최종 번역 결과를 생성하게 된다. 여기에서 볼드체는 문맥 저장 메모리(112)에 저장된 직전의 출력 토큰열을 나타낸다.
- [0107] Plato, a mathematician who lived 2400 years ago, in book of called 'The republic', a shepherd named Giges accidentally discovered a ring, and he became a transparent human when he put it on his finger.]
- [0108] 한편, 화면이나 음성 합성 장치를 통한 오디오 출력시에는, 중복되는 부분을 확인한 후, 완전 일치(exact matching)하는 영역을 삭제한 후 출력할 수 있다.
- [0109] 이하에서는 도 3을 참조하여 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 방법에 대하여 설명하도록 한다.
- [0110] 도 3은 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동통역 방법의 순서도이다.
- [0111] 한편, 도 3에 도시된 단계들은 신경망 기반 실시간 자동통역 장치(100)를 구성하는 서버(이하, 서버)에 의해 수행되는 것으로 이해될 수 있지만, 이에 제한되는 것은 아니다.
- [0112] 먼저, 서버는 신규의 제N 번째(N은 자연수) 입력 토큰열을 입력받으면(S110), 제N번째 입력 토큰열과 함께, 메모리 상에 저장된 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 인공지능망 번역 모듈(120)로 전달 여부를 결정한다(S120).
- [0113] 이때, 제N-1 번째 입력 토큰열 및 이에 상응하는 제N-1번째 출력 토큰열을 함께 전달하기로 결정한 경우(S120-Y), 제N-1 번째 입력 토큰열과 제N 번째 입력 토큰열을 순차적으로 결합하여 인공지능망 번역 모듈(120)로 전달하여(S130), 결합된 입력 토큰열에 상응하는 출력 토큰열을 생성한다(S140).
- [0114] 이와 달리, 제N 번째 입력 토큰열만 전달하기로 결정한 경우(S120-N), 제N 번째 입력 토큰열만을 인공 신경망 번역 모듈(120)로 전달하여(S150), 인공지능망 번역 모듈(120)의 디코더를 통해 상기 제N 번째 입력 토큰열에 상응하는 제N 번째 출력 토큰열을 생성한다(S160).
- [0115] 한편, 상술한 설명에서, 단계 S110 내지 S160은 본 발명의 구현예에 따라서, 추가적인 단계들로 더 분할되거나,

더 적은 단계들로 조합될 수 있다. 또한, 일부 단계는 필요에 따라 생략될 수도 있고, 단계 간의 순서가 변경될 수도 있다. 아울러, 기타 생략된 내용이라 하더라도 도 1 내지 도 2의 신경망 기반 실시간 자동 통역 장치(100)의 내용은 도 3의 내용에도 적용될 수 있다.

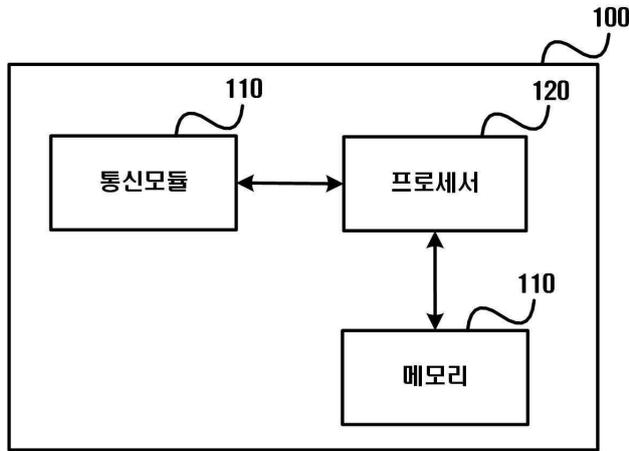
- [0116] 도 1 내지 도 2를 참조하여 설명한 신경망 기반 실시간 자동 통역 장치(100)는 상술한 서버의 구성요소로 제공될 수 있다.
- [0117] 이상에서 전술한 본 발명의 일 실시예에 따른 신경망 기반 실시간 자동 통역 방법은, 하드웨어인 컴퓨터와 결합되어 실행되기 위해 프로그램(또는 어플리케이션)으로 구현되어 매체에 저장될 수 있다.
- [0118] 상기 전술한 프로그램은, 상기 컴퓨터가 프로그램을 읽어 들여 프로그램으로 구현된 상기 방법들을 실행시키기 위하여, 상기 컴퓨터의 프로세서(CPU)가 상기 컴퓨터의 장치 인터페이스를 통해 읽힐 수 있는 C, C++, JAVA, Ruby, 기계어 등의 컴퓨터 언어로 코드화된 코드(Code)를 포함할 수 있다. 이러한 코드는 상기 방법들을 실행하는 필요한 기능들을 정의한 함수 등과 관련된 기능적인 코드(Functional Code)를 포함할 수 있고, 상기 기능들을 상기 컴퓨터의 프로세서가 소정의 절차대로 실행시키는데 필요한 실행 절차 관련 제어 코드를 포함할 수 있다. 또한, 이러한 코드는 상기 기능들을 상기 컴퓨터의 프로세서가 실행시키는데 필요한 추가 정보나 미디어가 상기 컴퓨터의 내부 또는 외부 메모리의 어느 위치(주소 번지)에서 참조되어야 하는지에 대한 메모리 참조관련 코드를 더 포함할 수 있다. 또한, 상기 컴퓨터의 프로세서가 상기 기능들을 실행시키기 위하여 원격(Remote)에 있는 어떠한 다른 컴퓨터나 서버 등과 통신이 필요한 경우, 코드는 상기 컴퓨터의 통신 모듈을 이용하여 원격에 있는 어떠한 다른 컴퓨터나 서버 등과 어떻게 통신해야 하는지, 통신 시 어떠한 정보나 미디어를 송수신해야 하는지 등에 대한 통신 관련 코드를 더 포함할 수 있다.
- [0119] 상기 저장되는 매체는, 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 관독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상기 저장되는 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있지만, 이에 제한되지 않는다. 즉, 상기 프로그램은 상기 컴퓨터가 접속할 수 있는 다양한 서버 상의 다양한 기록매체 또는 사용자의 상기 컴퓨터상의 다양한 기록매체에 저장될 수 있다. 또한, 상기 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장될 수 있다.
- [0120] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0121] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

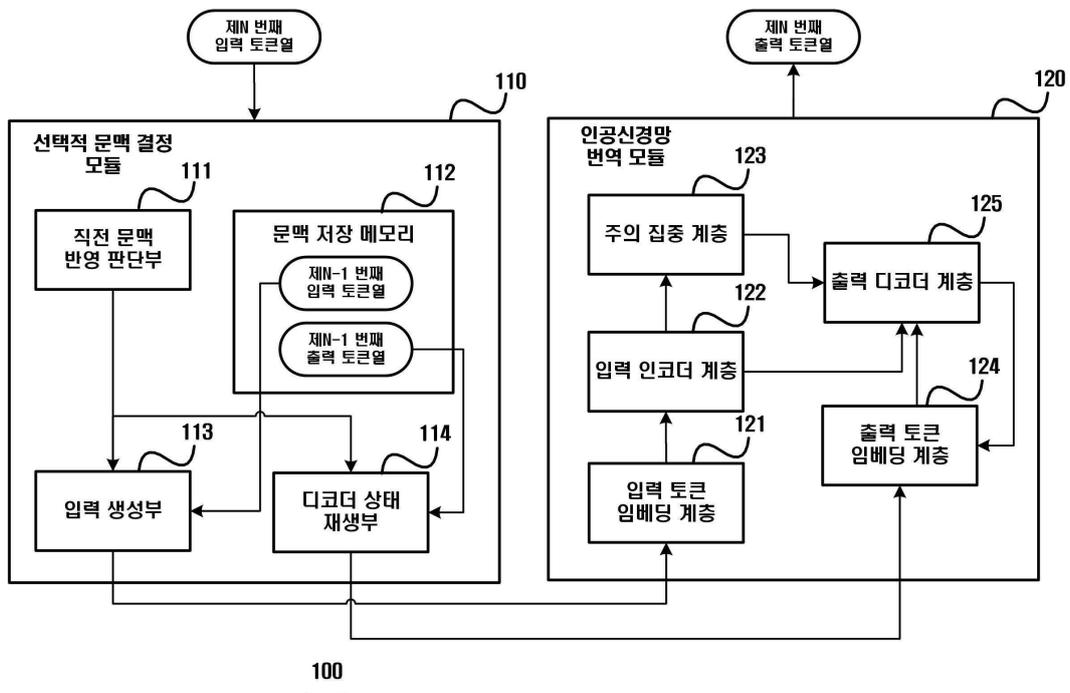
- [0122] 100 : 신경망 기반 실시간 자동통역 장치
- 110: 통신모듈
- 120 : 메모리
- 130 : 프로세서

도면

도면1



도면2



도면3

