2001년도 대한전기학회 하계학술대회 논문집 2001.7.18-20

통신환경에서 음성인식 인터페이스

°한태근*, 김종근**, 이동욱*
동국대학교 *전기공학과 **전자계산원

Speech Recognition Interface in the Communication Environment

°Tai-Kun Han*, Jong-Keun Kim**, Dong-Wook Lee*
*Dep. of Electrical Engineering, Dongguk University, **DUCSI

Abstract - This study examines the recognition of the user's sound command based on speech recognition and natural language processing, and develops the natural language interface agent which can analyze the recognized command. The natural language interface agent consists of speech recognizer and semantic interpreter. Speech recognizer understands speech command and transforms the command into character strings. Semantic interpreter analyzes the character strings and creates the commands and questions to be transferred into the application program. We also consider the problems, related to the speech recognizer and the semantic interpreter, such as the ambiguity of natural language and the ambiguity and the errors from speech recognizer. This kind of natural language interface agent can be applied to the telephony environment involving all kind of communication media such as telephone, fax, e-mail, and so on.

1. 서 론

최근 들어서 사용자와 기계의 인터페이스는 키보드나 마우스등 단순한 기계적 인터페이스의 한계를 넘어 점차 음성, 문장, 제스쳐, 표정 등 다양한 형태로 발전하고 있다. 이러한 현상은 발전한 컴퓨터 기술에 기인한다고 할 수 있다. 특히, 반도체 기술의 발달로 인한 컴퓨터의 처리 속도 향상 및 메모리의 저장 용량 확대로 개인용 컴퓨터(PC)의 계산력이 크게 향상되어 정보수집, 개인 통신 등까지 그 역할이 다양해 지고 있다. 또한, PC는 그 크기조차 점차 소형화되어 노트북에 이어 손바닥만한 PDA (Personal Data Assitant)까지 나와, 이제는 때와 장소에 관계없이 꼭 필요한 현대인의 필수품이 되어 가고 있다. 따라서 이제는 기존의 사용자 인터페이스의 고정 관념에서 벗어난 새로운 형태의 인터페이스 개발의 필요성이 널리 인식되고 있다.

한편, 인공 지능 분야에서는 지식과 추론 능력을 가지고 사용자를 대신하여 주어진 작업을 수행하는 독립적인 프로그램에 대한 연구가 활발히 진행되고 있는데 이러한 프로그램을 에이전트라고 부른다. 따라서, 에이전트는 지능을 가진 사용자 인터페이스(Intelligent User Interface)라 할 수 있다. 에이전트는 기능별로 인터페이스 에이전트, 회의 및 일정관리 에이전트, 전자우편 처리 및 뉴스 선별 에이전트, 엔터테인먼트 선별 에이전트 등으로 구분되며, 이러한 에이전트 중 사용자와의 상호작용을 원활히 하는 것을 목적으로, 음성, 문장, 제스쳐, 표정 등의 이해나 표현을 담당하는 에이전트를 인터페이스 에이전트라고 한다.

자연 언어 인터페이스 에이전트는 사람의 음성을 문자열로 인식, 변환하는 음성인식기와 그 문자열로부터 사용자 명령의 의미를 분석하는 의미 해석기로 구성되어 있다. 본 논문에서는 음성인식기와 의미해석기의 결합시 발생되는 문제점을 해결함으로써, 본래의 자연언어에 포함된 애매성 뿐만 아니라 음성 인식기로부터 부가된 애매성 및 오류를 모두 제거하여 사용자의 음성 명령으로부터 올바른 의미를 파악하는 의미해석기를 설계하는데 그 목적을 두고있다.

2. 본 론

2.1 음성 인식기와 의미 해석기의 결합

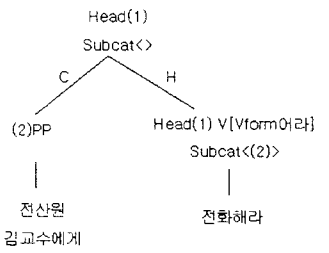
컴퓨터에 의한 음성 언어분석은 그 처리과정이 여러 단계로 구성되어 있고, 각 단계마다 자연언어 특성 상 애매성을 내포한다. 이러한 애매성은 여러 분석 단계를 거치는 동안 기하급수적으로 증가한다. 이것은 사람이 문맥이나 상황 정보를 이용하여 그 순간 순간의 애매성을 제거하는 것과는 커다란 차이가 있다. 따라서 사용자의 음성 명령으로부터 문자열로 변환하는 과정에서 여러 형태의 애매성이 포함되므로, 음성 인식기의 결과로부터의 의미분석은 실제 문장의 의미를 분석하는 것과는 아주 상이하다. 이러한 문제를 보완하기 위한 노력으로 음성 인식기에서는 가장 맞을 가능성이 높은 N개의 문자열 후보를 제공함으로써, 인식 단계에서 완전히 제거하지 못하는 문장의 애매성을 더 많은 정보가 이용가능한 의미분석 단계에서 해결하도록 한다. 이러한 음성 인식 방식을 N-best탐색이라 부른다. 자연언어 인터페이스 에이전트의 전체적인 성능은 사용자의 음성 명령에 일치한 명령실행과 올바른 질의의 생성 여부에 의해 판단 할 수 있다. 음성 인식기와 의미 해석기의 성능 모두 실행되는 도메인의 크기, 즉 도메인에서 허용하는 단어의 수에 직접적으로 영향을 받는다. 현재의 기술을 기준으로 음성 인식기는 단어 수의 증가에 대한 성능 저하가 훨씬 심하므로, 의미 해석기의 설계는 주로 문자열의 의미를 분석하는 원래의 자연언어 처리 목적 뿐만 아니라 음성 인식기의 인식오류의 처리에 촛점을 맞추어 설계한다.

2.2 의미 해석기의 구조

의미 해석기는 전처리 단계, 구문·의미 분석 단계,애매성 제거 단계, 실행 명령·질의 생성단계 등의 네 부분으로 구성된다. 첫째로, 전처리 단계에서는 음성 인식기로부터 생성된 문자열의 각 어절을 의미의 최소 단위인 형태소로 분석한다. 이러한 과정을 자연언어 처리 분야에서는 형태소 분석이라 부르는데, 언어 분류상 교착어로 구분되는 한국어는 영어와 비교할 때 분석 과정이 훨씬 복잡하다. 이외에도 전처리 단계에서는 형태소 분석 결과로 부터 다시 구문·의미 사전을 검색하여 각 단어의 구문·의미 정보를 찾아 토큰을 형성한다.

둘째로, 구문·의미 분석 단계에서는 HPSG (Head-Driven Phrase Structre Grammer)이라 불리는 문

법에 의하여 구문 구조가 분석되고, 문장의 의미가 파악된다. HPSG문법은 문맥 자유 문법(Context-Free Grammer)과 달리 문장 내의 어순이 문법에 의하여 정하여 지지 않고, 단지 구부터 시작하여 절, 문장까지 각각의 의미상 중심이 되는 머리(Head)를 기준으로 나머지 단어들의 통상적 역할과 내용적 의미를 파악하는 것이 특징이다. 따라서 이 문법은 어순이 부분적으로 자유로운 중심어(句나節의 의미상 핵심이 되는 단어)의 위치가 고정되고 나머지 문장 성분들의 위치가 자유로운(Partially Free Order Language)한국어의 의미 분석에 적합하다. 전처리 단계에서 생성된 토큰에 어휘 규칙과 구문·의미 규칙을 적용함으로써 문장내의 각 단어간의 수식관계가 트리 구조로 표시되는데, 이 트리를 파스 트리(Parse Tree)라 부른다. 단어 간의 수식 관계를 시작으로 전체 문장의 의미가 분석되는데, 이러한 방식의 구문 분석을 "Bottom-up Parsing"이라 한다. 이 파싱 방식은 의미가 부분적으로만 분석된 경우도 처리할 수 있는 장점이 있다. 구문·의미 분석 단계의 결과는 문장의 가능한 모든 의미를 나타내는 트리구조의 복합체로, 이러한 다중 트리는 같은 단어가 여러 의미로 쓰일 수 있는 중의성, 자연언어 문법의 애매성에 기인한다. 이러한 단어 및 문장의 애매성을 제거하여 문장의 여러 의미로부터 사용자 명령의 의미를 결정하는 것은 애매성 해결(Disambiguation)단계에서 행하여 진다. 자연 언어의 애매성은 형태소 분석 단계에서부터 시작하여, 구문 의미분석 단계에 이르기까지 다양하게 존재하며, 사용자 명령을 실행하기 위해서는 이러한 애매성을 완전히 제거해야만 한다. 이를 위하여, 각 단계별로 적절히 애매성을 제거하며, 이 때 제거하지 않은 애매성은 여란 단계의 정보를 종합하여, 다시 적용함으로써 제거한다.



〈그림 1〉 머리-보충어(H-C) 구조

마지막으로 실행 명령·질의 생성 단계에서는 구문·의미 분석 단계에서 분석된 사용자 명령의 의미로부터 응용 프로그램을 실행 시키기 위한 명령문을 생성하거나 사용자의 전화 번호나 사용기록이 저장되어 있는 데이타베이스를 검색하기 위한 질의를 생성한다. 이를 위해서 먼저, 사용자 명령의 의미가 표현되어 있는 파스 트리로부터 각 실행문이나 질의를 생성하는데 필요한 요소를 찾는다. 이러한 요소를 템플리트(Template)라 부르는데, 예를 들어 통신 환경에서는 첫째로 "Target", 전화나 팩스를 보내거나 받은 대상, 즉 사람이나 회사, 둘째로 "Domain", 전화, 팩스, 전자 우편 등, 세째로 "Time", 시간 정보 등을 포함한 템플리트로 구성되어 있다. 따라서, 각각의 실행명령이나 질의에는 필요한 정보가 정의되어 있으며, 모든 템플리트가 채워졌을 때, 그 결과가 응용 프로그램에 보내져, 사용자의 음성 명령이 실행되며, 질의의 경우, 그 데이타 베이스를 찾아 그 결과가 다시 사용자에게 보여진다.

2.3 구문·의미 분석 단계

구문·의미 분석 단계에서는 자연언어 문장의 의미를 분석하기 위해서 먼저 단어와 단어의 수식 관계가 파악되고 단어들이 모여 구성하는 구와 절의 문장 내에서의 성분이 파악된다. 자연언어 문장의 구성 성분은 주어, 목적어, 부사어, 관형어, 서술어 등으로 분류된다. 이러한 분석 과정을 자연언어 처리 분야에서는 구문 분석이라 부른다. 구문 분석을 위하여 필요한 정보로는 각 단어의 품사와 의미 정보가 있으며, 이러한 단어의 정보는 구문 사전에 보관된다. 그리고, 각각의 단어 및 구의 수식 관계를 규정하는 자연언어 문법이 필요한데, 본 시스템에서는 HPSG를 구문문법으로 사용하였다. 자연언어의 의미 문석은 문장 내에서 각 성분의 의미적 역할(의미격)을 분석하는 것으로, 의미격은 일반적으로 행위자격, 대상격, 동반자격, 장소격, 시간격, 양태격, 도구격 등 약 20여 가지로 분류된다. 예를 들어, "사장님께서 어제 미국에서 귀국하셨다"의 문장을 분석하면, 다음과 같은 의미분석 결과를 얻게된다.

사장님께서 (주어/행위자격)
어제 (부사어/시간격)
미국에서 (부사어/시작장소격)
귀국하셨다. (서술어)

분석결과를 자세히 살펴보면, 동사의 행위자, 즉 귀국한 사람은 사장님이며, 그 행위가 일어난 시점은 "어제"이고, 행위의 시작점은 "미국"이었음을 알 수 있다.

2.4 애매성 해결(Disambiguation)

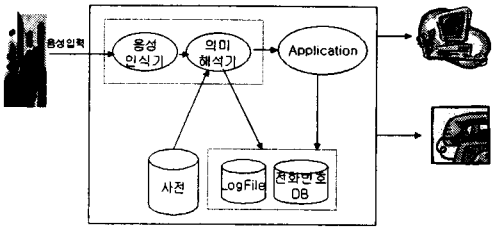
통신 환경에서 구현한 본 시스템에서는 도메인이 제한적이므로 큰 문제가 되지 않는다. 따라서 본 시스템에서는 상대적으로 구조적 애매성을 적절하게 해결하는 것이 시스템의 성능에 큰 영향을 미친다. 구조적 애매성을 대표하는 것으로 영어권에서는 전치사구 결합문제(Prepositional Phrase Attachment Problem)가 전통적으로 많이 거론되고 다음은 전치사구 "with the telescope"가 명사구 "the boy"를 수식하느냐 동사구 "saw"를 수식하느냐에 따라서 두 가지 다른 의미를 나타내는 예이다.

I saw the boy with the telescope
(나는 망원경으로 소년을 보았다)
I saw the boy with the telescope
(나는 망원경을 가진 소년을 보았다)

영어와 한국어의 구조적 차이로 인하여 전치사구 결합문제를 한국어에 직접 응용한다는 것은 무리지만, 이 문제를 해결하기 위해서 나온 많은 이론들은 한국어에서도 적용해 볼 가치가 있다. Kimball[5] 이후 발표된 여러 이론들에서, 애매성을 해소하기 위해서는 구문적인 요소와 의미적인 요소를 함께 고려해야 한다고 주장하고 있다.

2.6. 시스템 구성도

자연언어 인터페이스 에이전트가 다양한 통신 환경에서 사용되는 시스템의 구조는 〈그림 2〉에서 보여주는 것처럼 사용자의 음성 명령을 인식하여 그 의미를 분석하고, 그 결과를 다시 응용프로그램에 전달하여 전화, 팩스, 전자우편, 호출기, 자동응답기 등의 기능을 실행시킨다. 뿐만 아니라, 사용자 개개인의 사용기록 정보를 개인정보 데이터 베이스에 저장하여, 사용자가 원하면 음성 명령에 의하여 팩스수신 시각, 통화시간, 전화를 건 곳 등의 다양한 정보를 찾아 보여 준다



〈그림 2〉 시스템 구성도

2.6. 결과 및 분석

자연언어 인터페이스 에이전트를 실험한 결과, 평균 단어 10개로 구성된 명령에 대하여 Response Time은 3초 내외였으며, 의미 해석기만에 대해서는 1초미만으로 나타났다. 여러 문장들에 대해서 음성 인식결과 중, N-best 후보 중에 첫 후보가 원래 사용자가 발화한 문장과 문자 열 수준에서 완전히 맞는 문장의 수가 약66%, 첫 후보가 완전히 동일하지는 않지만, 조사나 어미가 잘못 인식되어 문법적으로는 약간 틀리더라도, 의미가 원래의 문장과 동일한 문장수가 약 16%, N-best 후보 중에 분석되는 문장이 하나도 없는 경우가 약 5%, 첫 후보가 아닌 다른 후보의 문장이 파싱에 성공한 3가지 가운데, 올바른 의미가 선택된 경우가 약 1%, 잘못된 의미가 선택된 것이 약 0.1%였다. 또 첫 후보가 파싱은 성공하였으나 그 의미가 틀린것은 전체 11% 문장이었다. 본 시스템은 N-best 방법을 이용하고 있으므로 첫 후보부터 차례로 성공할 때까지 파싱을 진행한다. 일단 파싱이 성공하고 나면, 그 후보를 맞는 의미로 결정하므로 그 다음은 더 이상 파싱을 진행하지 않는다. 따라서 첫 후보가 파싱은 되나 의미가 틀린 문장과 첫 후보가 아닌 다른 후보에서 잘못된 의미를 선택한 문장을 합쳐 11%는 잘못된 의미의 후보가 먼저 나와서 그 의미로 결정해 버린 에이전트의 에러율을 나타낸다. 잘못된 음성 인식결과이지만, 의미·구문적으로는 옳은 문장으로 파싱에 성공하여 결과적으로 에이전트의 오동작을 유발하는 유형의 예를 다음에 보였다.

1. 엉뚱한 명령을 수행하는 경우
사용자 명령 문장에서 명령부분이 잘못 인식되어서 다른 행동을 취하는 경우로 전체 예에서 다음 2가지만이 여기에 해당되었다.
- "어제 받은 전자 우편 찍어줘"에서 "찍어줘"를 "지워줘"로 잘못 인식
- "팩스 온 것 뽑아"를 "팩스 온것 보관"으로 "뽑아(print)"를 "보관(save)"로 인식

2. 특정 키워드의 오인식으로 인한 명령의 다상/범위의 오류
- "목요일"을 "토요일", "세시"를 "네시" 등으로, 요일, 시간 따위를 잘못 인식
- "철수 담임 선생님한테 전화를 걸어"를 "철수 선생님한테 전화걸어"로 잘못 인식

3. 1,2 유형의 복합
- "필동 영어선생님께 음성 남겨줘"를 "필동 영어선생님께 횟수를 알려줘"와 같이 대상 및 명령 모두 틀리는 경우

결과적으로 음성 인식기가 완전히 옳은 결과를 첫 후보로 제시한 66.3%에서 자연언어 처리기를 통하여 88.6%까지 성능을 향상시킬 수 있었다. 이와 같이 좋은 결과를 얻을 수 있었던 것은 애매성 해소 규칙이 잘 적용되었다는 점 이외에도 도메인이 상당히 제한적이므로 규칙의 예외사례가 적었고, 튜닝이 잘 이루어 질 수 있었던 점, 컴퓨터에 대한 명령으로 한 문장의 평균 길이가10 단어 정도의 비교적 짧은 문장이라는 점과 문장 set를 검토한 후 귀납적으로 규칙을 만들었다는 점도 크게 작용했으리라 생각된다.

3. 결 론

본 연구의 목적은 음성 인식과 자연언어 처리 기술을 기반으로 사용자의 음성 명령을 인식하고 그 의미를 분석할 수 있는 자연언어 인터페이스 에이전트를 개발하는 것이었다. 본 논문에서는 자연언어 인터페이스 에이전트를 구성하는 음성 인식기와 의미 해석기의 결합시 발생되는 문제점을 확인하고 이를 효과적으로 처리할 수 있도록 의미 해석기를 설계함으로써, 의미 해석기가 본래의 자연언어의 애매성 뿐만아니라 음성 인식기로부터 부가된 애매성 및 오류를 제거하여 사용자 명령의 의미를 올바로 파악할 수 있었다.

통신 환경에서 자연 언어 인터페이스 에이전트의 분석률이 90%에 근접하는 것은, 현재의 음성 인식과 자연언어 처리 기술의 수준을 고려할 때, 그 도메인의 크기가 적절하였다고 볼 수 있다. 한편, 전화, 팩스,전자우편, 호출 등의 일상의 통신 수단이 다양해 졌고, 그 기능 또한 복잡하므로 일반 사용자가 이용하는데에 어려움이 많다. 따라서 본 시스템의 실용성이 아주 크다고 할 수있다.

〔참 고 문 헌〕

〔1〕 고영근,남기심, 표준 국어 문법론.탑 출판사, 1993.
〔2〕 이승배,이종석, N-best 문장 탐색기법을 적용한 연속 음성 인식 시스템.제13회 음성 통신 및 신호처리 워크샵 논문집, pp151-154, 1996.
〔3〕 장석진, 정보기반 한국어 문법, 언어와 정보,1993.
〔4〕 Etzioni,O. and Weld,D., A softbot-based interface to the internet, Communications of the ACM 76, 1994.
〔5〕 Kimball,J., Seven principles of surface structural parsing in natural language, Cognition2, "15-47, 1973.
〔6〕 Lin,D., On the structural complexity of natural language sentence, COLING-96, pp729-733, 1996.
〔7〕 Maes,P., Agents that reduce work and information overload, Communications of ACM, 40, 1994.
〔8〕 Pollard,C. and Sag,I.A., Head Driven Phrase Structure Grammar, CSLI, The University of Chicago Press, 1994.
〔9〕 Schwartz,R. and Chow,Y.L., The N-best algorithm : An efficient and extract procedure for finding the N most likely sentence hypothesis, Proc. OFICASSP, pp. 81-84, 1991.
〔10〕 Wilks,Y. and Huang,X.D., Syntax, preference and right attachment, IJCAI, pp. 779-784, 1985.