실제 멀티모달 환경에서의 지시 대용어 처리

최맹식○ 이세희 김학수
강원대학교 컴퓨터정보통신공학전공
nlpmschoi@kangwon.ac.kr, nlpshlee@kangwon.ac.kr, nlpdrkim@kangwon.ac.kr

Resolution of Deictic Anaphora in Real Multimodal Environments

Maengsik Choi○, Sehee Lee, and Harksoo Kim
Program of Computer and Communications Engineering,
Kangwon National University

요 약

언어기반 대화 시스템에서는 시스템과 사용자의 대화가 발화 자체만으로 이루어지기 때문에 사용자가 사람과 대화하는 것처럼 지시 대용어를 사용할 수 없어서 불편하다. 그리고 사용자의 발화 의미를 시스템이 정확하게 해석하기가 어렵다. 하지만 이런 언어기반 대화 시스템과는 달리 멀티모달 대화 시스템에서는 발화 자체의 정보뿐만이 아닌 제스처와 같은 발화 이외의 행위 정보들이 포함되는데 이 정보를 이용하면 지시 대용어의 처리가 가능해짐으로 시스템과의 대화가 좀 더 자연스러워진다. 본 논문에서는 군집화와 격틀을 이용하여 여러 사물들 중에서 지시 대용어가 될 가능성이 있는 지시 후보 선정을 한다. 그리고 특출성 점수와 엔트로피를 이용하여 후보 사물들 중에서 지시 대용어가 될 수 있는 대상을 선택하는 알고리즘을 제안한다. 시뮬레이션 환경에서의 실험결과 평균 2.8번의 상호작용으로 지시 대용어를 처리할 수 있었다.

1. 서 론

사람의 대화는 발화 자체의 정보뿐만이 아니라 다른 중요한 정보들을 수반하는데 그 중에서 행위 정보는 발화의 내용을 이해하는데 중요하다. 멀티모달 대화 시스템(multimodal dialogue system)은 사용자와 시스템이 발화의 정보만을 이용하여 상호작용하는 것이 아닌 발화, 표정, 제스처(gesture) 등의 다양한 입력 채널을 가지는 대화 시스템이다. 이러한 멀티모달 대화 시스템은 사용자에게 보다 편리한 인터페이스를 제공하지만 사용자의 발화 의도를 정확하게 파악하기 위한 방법이 개선되어야 한다. 본 논문에서는 대용어 중 지시 대용어 처리 중심으로 알아본다. 지시 대용어는 사용자가 '저거'라는 대용어와 함께 사물을 가리키는 지시 행위를 수반하는 대용어로 발화 자체의 의미만으로는 해석이 불가능하다. 이럴 경우 사용자 발화에 사용자의 지시 행위가 함께 수반된다. 본 논문의 멀티모달 시스템의 입력 채널은 크게 사용자의 발화를 받아들이는 청각 채널(auditory channal)과 사용자의 지시 행위를 받아들이는 시각 채널(visual channal)로 구분한다. 그러므로 사용자는 "저거 가져와."라고 말하면서 특정 물건을 가리킬 수 있고, 시스템은 이를 해석할 수 있다.

멀티모달 대화 시스템에서 지시 대용어를 처리하는 기존의 연구는 터치스크린 환경에서 사물과 매핑하는 것이었다[1,2]. 그러나 이러한 방법은 지시 행위가 정확히 사물에 일치할 경우만 가능하므로 실제 지능형 인간 로봇과 같은 시스템에 적용하기에는 부적합한 면이 있다. 또한 지시 대용어의 해석이 틀렸을 경우 사용자와 시스템간의 대화(상호작용)를 많이 일으키게 되므로 좀 더 구체적인 알고리즘이 필요하다. 본 논문에서는 사용자 발화와 지시 행위로부터 최소한의 상호작용을 통한 지시 대용어의 해석을 위해 군집화(clustering), 격틀(case-frame), 특출성 점수(salience score) 및 엔트로피(entropy)를 통한 방법을 제안한다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서 지시 대용어를 효과적으로 처리하기 위한 방법을 제안한다. 3장에서 구현된 시스템을 통하여 지시 대용어 처리 예제를 살펴보고, 4장에서 결론을 맺는다.

2. 지시 대용어 처리

지시 대용어 포함 발화에서 사용자가 "저거 가져와."라고 하면서 '저거'에 해당하는 사물을 김학수-2000[1]에서처럼 정확히 가리키면 좋겠지만, 사용자가 해당 사물을 정확히 가리킬 수도 있고 또는 그 근처를 가리킬 수도 있다. 그리고 사용자가 정확히 가리켰다고 해도 실제 시스템의 이미지가 3차원 공간상이라면 정확한 좌표를 알아내기가 힘들다. 이럴 때 사람의 경우에는 적당한 추측을 통해서 '저거'의 의미를 파악하지만 시스템의 경우에는 사람과 같은 추측을 할 수가 없기 때문에 사람끼리의 대화에서처럼 지시 대용어의 사용이 어렵다.

지시 대용어의 처리는 다음과 같은 순서로 진행된다. 우선 사용자의 발화가 태깅[3]되어서 시스템에 문자열로 입력되고 지시 행위 또한 좌표로 시스템에 입력된다는 가정 하에 사용자의 발화 내에 지시 대용어가 있으면 사물들을 군집화하고 격틀을 이용하여 지시 후보를 선정한다. 그리고 사용자의 행위 정보와 사물들의 정보를 이용하여 특출성 점수를 계산하여 지시 대용어의 대상이 될 사물을 선택한다. 만약 선택 된 사물이 틀리다면 엔트로피 계산을 이용하여 지시 대용어를 처리한다. 사용자 발화 태깅 방법은 그림 1과 같다.

그림 1에서 MA(main action)는 발화 내의 동사 분류이고, NE(named entity)는 발화 내의 목적어에 해당하는 명사를 나타낸다.

```
<frame>
  <utterance>창문 좀 닫아.</utterance>
  <slot type='MA'>close</slot>
  <slot type='NE'>창문</slot>
</frame>
<frame>
  <utterance>저거 좀 가져와.</utterance>
  <slot type='MA'>get</slot>
  <slot type='ANAPHORA'>저거</slot>
</frame>
```

그림 1 사용자 발화 태깅 결과

그림 2에서 애매성 존재의 판단은 지시 대용어가 될 수 있는 후보 사물의 수가 1개면 애매성이 없는 것이고, 2개 이상이면 애매성이 있는 것이다.

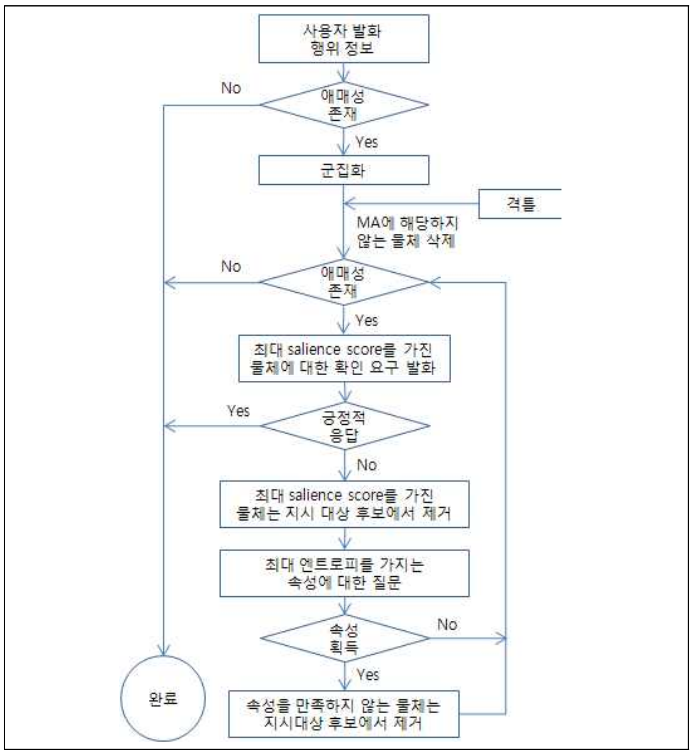

그림 2 시스템 흐름도

2.1 지시 후보 선정

지시 대용어 처리에서 가장 먼저 할 일은 지시 대용어가 될 수 없는 사물들을 제거 하고 지시 대용어가 될 수 있는 사물들만을 선정하는 것이다. 지시 후보 선정은 군집화와 격틀을 사용한다.

사물을 가리키는 행위를 수반한 발화에서 대상 물건이 무엇인지 확인할 때 사람은 여러 가지를 고려하여 판단한다. 그 중에서 사물들의 배치도 포함되는데 여러 사물이 모여 있는 것보다는 따로 떨어져있는 한 두 개의 사물이 대상 사물이 될 가능성이 높다. 그래서 지시대용어를 처리하기 위해서 사물들의 군집화가 필요한데 군집화에 필요한 유사도는 각 사물들의 거리지표를 사용한다. 그림 3은 본 논문에서 사용하는 군집화 알고리즘이다.

```
1. 지시 행위로부터 가장 가까운 사물 선택
2. 새로운 군집의 생성
3. 선택된 사물을 군집에 추가
4. 군집내의 모든 사물들의 평균 좌표를 구함
5. 군집에 포함되지 않은 사물 중에서 현재
     군집과 가장 가까운 사물 선택
6. 두 좌표 사이의 거리가 군집화 영역 안이면
     3번으로 돌아감.
   두 좌표 사이의 거리가 군집화 영역 밖이면
     2번으로 돌아감.
7. 모든 사물이 군집화 되었으면 종료.
```

그림 3 군집화 알고리즘

군집화 알고리즘에서 모든 사물이 하나의 군집에 포함되는 것을 막기 위해 고정된 임의의 군집화 영역을 설정한다.

군집화가 끝나면 그림 4와 같은 격틀을 이용하여 지시 후보를 선정한다.

```
<DomainList>
  <Domain specification="errand">
    <MA="get">
      <NE>리모컨</NE>
      <NE>쿠션</NE>
      <NE>꽃병</NE>
      <NE>초</NE>
   </MA>
  </Domain>
  <Domain specification = "action">
    <MA="start">
      <NE>TV</NE>
      <NE>형광등</NE>
      <NE>초</NE>
    </MA>
    <MA="end">
      <NE>TV</NE>
      <NE>형광등</NE>
      <NE>초</NE>
    </MA>
    <MA="open">
      <NE>창문</NE>
      <NE>서랍</NE>
    </MA>
    <MA="close">
      <NE>창문</NE>
      <NE>서랍</NE>
    </MA>
  </Domain>
</DomainList>
```

그림 4 격틀

그림 4에서 NE는 지시 대용어가 될 수 있는 후보 사물을 의미한다. 지시 후보를 선정하기 위한 방법은 사용자 발화의 의미태깅 결과로부터 MA를 추출하고 격틀의

MA와 비교하여 추출된 MA와 같은 격틀의 MA에 있는 NE리스트에 해당하지 않는 사물들을 군집화 결과에서 제외시킨다. 그러면 NE리스트에 속하는 사물들만이 군집화 결과로 남게 된다.

2.2 지시 대상 결정

지시 대용어가 될 수 있는 지시 후보 선정이 되었으면 이 지시 후보 중에서 어떤 것을 지시 대용어의 실제 대상인지를 결정해야 한다. 지시 대용어 사물 선택에는 특출성 점수와 엔트로피를 이용한다.

특출성 점수는 사용자의 지시대용어 포함 발화와 함께 나타나는 행위 정보(사물을 가리키는 포인팅 행위)를 이용하여 지시대용어의 대상이 될 수 있는 점수를 나타낸다.

특출성 점수 =

$$\frac{1}{1 + kc \times \log Ci + kr \times \log Ri + kl \times \log Li + ko \times \log Oi}$$

Ci: 군집 내의 사물의 개수
군집 내의 사물의 개수가 많아질수록
점수가 작아진다.
Ri: 입력좌표와 사물의 거리 순위
입력좌표와 가까운 사물일수록 점수가 커진다.
Li: 한계영역 내에 같은 색의 사물의 개수
한계영역 내에 같은 색의 사물들은
점수가 작아진다.
Oi: 한계영역 내에 같은 종류의 사물의 개수
한계영역 내에 같은 종류의 사물들은
점수가 작아진다.
kc, *kr*, *kl*, *ko*: 각 요소의 중요도를 나타내며,
값이 클수록 중요도가 크다.
실험적으로
kc=0.2, kr=0.4, kl=0.15, ko=0.15로 설정

그림 5 특출성 점수 계산식

지시 행위의 위치에서 한계영역 밖의 사물들은 특출성 점수 계산에서 제외되는데 이는 근처에 사물이 없을 경우, 관련 없는 사물이 지시 대상으로 결정되는 것을 방지하기 위해서이다.

지시 후보 중에서 특출성 점수가 가장 높게 나타난 후보가 지시 대상인지를 확인하여 맞으면 지시 대용어가 제대로 처리 된 것이지만, 틀리다면 지시 후보 중에서 현재 선택된 지시 후보를 제거하고 새로운 지시 대상을 결정하여야 한다. 이때 엔트로피를 사용하는데 엔트로피는 무질서함, 복잡함 등으로 생각할 수 있다. 본 논문에서는 사물들의 정보에 대해서 엔트로피를 계산하고 엔트로피가 높은 정보의 속성을 획득하여 사용자와 시스템과의 상호작용을 최대한 줄이도록 한다.

$$H(X) = -\sum_{x=1}^{n} p(x)\log_2 p(x)$$

$H(X)$는 사물 정보에 대한 엔트로피
(예: 색의 엔트로피, 종류의 엔트로피)
x는 정보에 대한 속성
(예: 색 정보에 대한 속성 : 빨간색, 노란색 등)

그림 6 엔트로피 계산식

그림 6으로 지시 후보들에 대해 엔트로피를 계산하여 가장 높은 엔트로피를 가지는 정보에 대해 시스템은 사용자에게 질문을 하고 사용자는 속성을 알려준다. 그러면 시스템은 현재 지시 후보들 중에서 획득한 속성을 갖지 않는 지시 후보들을 제거한다. 그리고 속성을 가지는 지시 후보들만 새로운 지시 후보로 재분류하고 다시 특출성 점수를 계산하여 지시대용어를 찾게 된다.

3. 구현

3.1 구현 환경

본 논문의 최종 목적은 3차원 환경의 지능형 인간 로봇 시스템을 대용어가 포함된 발화로 제어하는데 있다. 그러나 본 논문에서는 실제 로봇 환경을 구축할 수 없어서 그림 7과 같은 터치스크린 기반의 시뮬레이션 환경을 구축하였다.

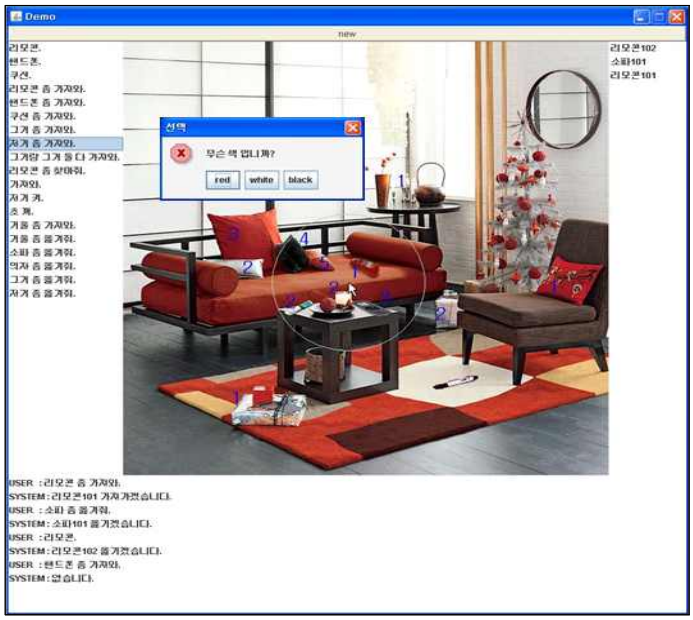

그림 7 시뮬레이션 환경

그림 7 실험 이미지는 일반적인 거실 환경에 사용자가 지시할 수 있는 지시 대상 사물들은 랜덤하게 배치되도록 하였다. 각각의 사물 중에서 이름이 같은 사물들을 구분 짓기 위해서 사물 옆에 번호를 부여하였다. 번호가 없는 사물은 그림에서 하나만 존재하는 것들이다. 사용자 발화는 그림 7의 왼쪽 부분과 같이 미리 태깅되어 있는 예제 발화들 중에서 하나를 선택하도록 하였고, 지시

행위는 터치스크린 모니터를 이용하여 위치를 선택하도록 하였다. 위치 선택 시에 사물을 정확히 선택하는 것이 아니라 실제 환경과 유사하게 하기 위하여 사물의 근처를 선택하는데 사물이 한계영역 안에는 항상 위치하도록 하였다.

3.2 구현 예제

그림 8은 기본시스템에서의 지시 대용어 처리의 예를 보여준다. 기본시스템은 본 논문의 제안 시스템에서 사용한 특출성 점수와 엔트로피는 사용하지 않았다. 지시 대상의 결정은 지시 후보에서 선택된 위치와의 거리가 가까운 것을 선택하였다.

사용자: (초2와 리모컨1 사이의 영역을 선택하며) 저거 가져와.
(‘쿠션4, 쿠션5, 리모컨1, 리모컨2, 리모컨3, 초2, 소파’ 사물이 한계영역 안에 있음, 소파는 격틀의 ‘가져와’에 해당하지 않으므로 제거)
시스템: 초2 맞습니까?
사용자: 아니요. (초2 삭제)
시스템: 리모컨1 맞습니까?
사용자: 아니요. (리모컨1 삭제)
시스템: 리모컨3 맞습니까?
사용자: 예.
시스템: 리모컨3 가져가겠습니다.

그림 8 대화예제 - 기본 시스템

그림 9는 본 논문에서 제한한 방법으로 인해 사용자와 시스템간의 상호작용이 얼마나 줄었는지를 보여준다. 그림 9에서 보는 것과 같이 제안 시스템은 그림 8과 동일한 환경에서 3번의 상호작용만으로 지시 대상을 찾아준다. 실제로 모든 경우를 다 따져보아도 지시 후보가 6개 존재하지만 최소 1번에서 최대 3번의 상호작용으로 지시 대용어를 처리할 수 있다.

사용자: (초2와 리모컨1 사이의 영역을 선택하며) 저거 가져와.
(‘쿠션4, 쿠션5, 리모컨1, 리모컨2, 리모컨3, 초2, 소파’ 사물이 한계영역 안에 있음, 소파는 격틀 ‘가져와’에 해당하지 않으므로 제거)
(특출성 점수 계산)
시스템: 초2(최대 특출성 점수 사물) 맞습니까?
사용자: 아니요. (초2 삭제)
(사물색의 엔트로피가 큼)
시스템: 흰색, 검은색, 빨간색 중에 무슨 색 입니까?
사용자: 검은색.(속성 획득-검은색 이외의 사물 제거)
(특출성 점수 재계산)
시스템: 리모컨3(최대 특출성 점수 사물) 맞습니까?
사용자: 예.
시스템: 리모컨3 가져가겠습니다.

그림 9 대화예제 - 제안 시스템

제안 시스템의 성능을 평가하기 위해 표 1과 같은 다양한 상황을 가정하여 상호작용 수를 계산하였다.

표 1 예제 상황

상황1		상황2		상황3	
리모컨	빨간색	리모컨1	빨간색	리모컨	흰색
주전자	노란색	리모컨2	노란색	주전자	흰색
핸드폰	파란색	리모컨3	파란색	핸드폰	흰색
연필	검은색	리모컨4	검은색	연필	흰색
지우개	흰색	리모컨5	흰색	지우개	흰색
쿠션	주황색	리모컨6	주황색	쿠션	흰색
상황4		상황5		상황6	
리모컨1	검은색	리모컨1	빨간색	리모컨1	빨간색
리모컨2	검은색	리모컨2	검은색	쿠션1	흰색
리모컨3	검은색	쿠션1	빨간색	쿠션2	검은색
리모컨4	검은색	쿠션2	흰색	핸드폰1	검은색
리모컨5	검은색	쿠션3	노란색	연필1	흰색
리모컨6	검은색	핸드폰1	검은색	연필2	빨간색

표 2는 표 1의 상황에 대해 제안시스템의 상호작용 수를 보여준다. 기본 시스템에서 후보 사물이 6개이지만 최대 5번의 상호작용만이 필요한 이유는 마지막 남은 후보사물의 경우에는 사용자와의 상호작용 없이 그냥 후보사물을 지시 대용어로 처리하기 때문이다.

표 2 예제 상황 결과

	기본 시스템		제안 시스템	
	최소	최대	최소	최대
상황1	1	5	1	2
상황2	1	5	1	2
상황3	1	5	1	2
상황4	1	5	1	5
상황5	1	5	1	3
상황6	1	5	1	3

표 2에서 보는 것과 같이 제안 시스템(평균 2.8번)은 기본 시스템(평균 5번)에 비하여 2.2번 적은 수의 상호작용만으로도 지시 대용어를 처리할 수 있음을 알 수 있었다.

4. 결론 및 향후 과제

본 논문에서는 멀티모달 대화 시스템에서의 지시 대용어 처리를 위해서 군집화와 격틀을 사용하여 1차적으로 지시 후보를 선정하고 특출성 점수를 이용하여 지시 대용어를 처리하였다. 여기에서 지시 대용어로 선택 된 사물이 틀렸을 경우 엔트로피를 이용하여 사용자와 시스템 간의 상호작용을 최소로 하고 지시 대용어를 처리 할 수 있는 방법을 제시하였다.

향후 연구과제는 다음과 같다. 먼저 제안 시스템은 군집화 할 때 사물의 중심좌표만을 사용하기 때문에 사물

의 크기는 반영이 되지 않는다. 이는 사물이 특정 군집의 영역에 포함이 될 수도 있지만 중심좌표가 포함이 되지 않는다면 다른 군집으로 될 가능성이 있으므로 사물의 크기를 반영하는 방법을 생각해야 한다. 또한 격틀의 경우, 사용자와 시스템 간의 상호작용을 통해서 수정이 가능한데 사용자마다 발화 내 동사에 따른 지시대용어 대상 사물의 종류가 달라질 수 있으므로 이를 통해 각각의 사용자에게 특화된 격틀을 구성한다면 좀 더 효과적일 것이다.

감사의 글

김학수의 이 연구(논문)는 지식경제부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

참고문헌

[1] 김학수, 서정연, *다중모드 대화 시스템에서 이중 캐시 모델의 센터링 알고리즘을 이용한 명사 대용어구 처리*, 정보과학논문지: 소프트웨어 및 응용 제27권, 제11호, pp.1133-1140, 2000

[2] 이세희, 김학수, “S-list를 이용한 멀티모달 참조대용어 처리”, *한국정보과학회 강원지부 제1회 학술대회 논문집*, pp.149-152, 2007

[3] 정민우, 이근배, *음성 언어 이해를 위한 기계 학습*, 정보과학회지 논문지 제27권, 제3호, pp.21-27, 2007

[4] 조은경, 서정연, “대화 시스템에서의 조응어 해석”, *제16회 한글 언어 인지 학술대회*, pp.283-289, 2004

[5] 김명자, 채숙희, 조은영, 이정민, “지시사 대용적 용법의 대조연구”, *제15회 한글 및 한국어 정보처리 학술대회*, pp.127-133, 2003