Exploiting knowledge base to generate responses for natural language dialog listening agents

Sangdo Han, Jeesoo Bang, Seonghan Ryu, Gary Geunbae Lee
Pohang University of Science and Technology (POSTECH), South Korea
hansd, jisus19, ryush, gblee@postech.ac.kr

Abstract

We developed a natural language dialog listening agent that uses a knowledge base (KB) to generate rich and relevant responses. Our system extracts an important named entity from a user utterance, then scans the KB to extract contents related to this entity. The system can generate diverse and relevant responses by assembling the related KB contents into appropriate sentences. Fifteen students tested our system; they gave it higher approval scores than they gave other systems. These results demonstrate that our system generated various responses and encouraged users to continue talking.

1 Introduction

Dialog systems can be separated into task-oriented dialog systems and nontask-oriented dialog systems. Task-oriented dialog systems have mainly been intended to communicate with devices like cellphones or televisions. Nontask-oriented dialog systems are intended for use as entertainment, or to provide casual dialog. In this paper, we studied the listening agent, which is one nontask-oriented dialog system.

The main objective of the listening agent is to analyze user's utterances and to generate appropriate response that satisfies user's desire to speak (Meguro et al., 2009). To satisfy this desire, the system should emulate actual 'listening' by responding appropriately to user utterances in ways that make the user feel that the system is responding specifically to the utterances.

Listening agents should generate various responses to encourage the user to continue the dialog. If responses are monotonous, a dialog can be boring, and a user may lose interest in talking to the system. In previous work, listening agents generated system responses to content extracted from user utterances (Weizenbaum, 1966; Han et al., 2013; Han et al., 2015). For example, when a user talk about the footballer Lionel Messi "I like Messi", the system responses are "Why do you like Messi?", or "You like Messi". However, by using only extracted contents from user utterances, system responses are too restricted to encourage the user to engage in conversation. To increase the user's motivation to interact with the system, the diversity and relevance of the external knowledge that it uses must be increased.

Our objective of this study is to increase the variety of system responses. For the previous example, our system could generate responses like: "What is Messi's position?", "Do you like David Beckham, too?", or "You like Messi, a football player". We also expected encouraging dialog by talking about related information, and increasing dialog satisfaction by pin-pointing the content that user want to talk about. The system extracts named entities from a user utterance, recognizes them, and extracts related information from a knowledge base (KB) to guide generation of responses.

2 Related Work

2.1 Listening Agent

Two main types of listening agents have been developed: non-verbal agents and verbal agents. Non-verbal listening agents generate multimodal responses from multimodal user input (Schroder et al., 2012). Verbal listening agents get text input from user and generate a text response (Weizenbaum, 1966; Han et al., 2013; Han et al., 2015). Our study focused on a verbal listening agent.

2.2 ELIZA & Counseling Dialog System

ELIZA (Weizenbaum, 1966) is a natural language conversation program that interacts with

Proceedings of the SIGDIAL 2015 Conference, pages 129–133,
Prague, Czech Republic, 2-4 September 2015.

a speaker as a psychoterapist would. The system models person-centered therapy, a counseling technique based on the reflective listening strategy (Rautalinko and Lisper, 2004), which aims to encourage a user to continue talking. It includes encouragement, recapitulation, questioning, and reflecting emotion. Because the system generates a response by matching keywords and replaces slot with the contents for user utterance, the variety of responses that it can generate is limited.

Han et al. (2015) developed a listening agent that uses a dialog strategy based on microskills (Ivey et al., 2013), which is a basic communication technique that includes attending, paraphrasing, questioning, and reflecting feeling. This is similar to the reflective listening strategy used in ELIZA. Han's system encourages users to continue talking. Because the system also generates a response based only on information extracted from user utterances, the variety of responses that it can generate is also limited.

ELIZA and Han's dialog strategies are both based on effective listening. In this study, we designed our dialog strategy, focusing on knowledge driven response generation while simultaneously communicating using microskills.

3 System Architecture

Our system (Figure 1) includes five modules: emotion detection, natural language understanding, related information extraction, dialog management, and natural language generation module. The natural language understanding module includes user intention detection, triple extraction, and named entity recognition module.

3.1 Emotion Detection

Our emotion detection module uses a keyword-based method (Guinn and Hubal, 2013). We assembled an emotional keyword lexicon, which includes 170 keywords with 7 basic emotions: sadness, anger, happiness, fear, disgust, contempt, and surprise. Emotional keywords were collected from Ivey's list of 'feeling words' (Ivey et al., 2013). We detect these basic emotion when a user utterance includes one or more of these keywords.

3.2 Natural Language Understanding

3.2.1 User Intention Detection

We detected user intention in collected listening agent training data. We collected dialogues with

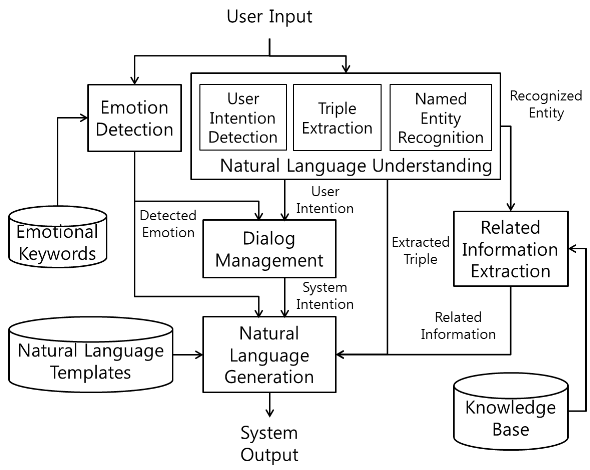

Figure 1: System Architecture. Components and processes are described in the text.

15 students who generated a total of 77 dialogues in English. Students worked in pairs to generate dialogues; one student had the role of speaker or the other had the role of listener. Listeners responded based on listening techniques of microskills. They communicated by text through the internet. The dialog topic was chosen freely by the speaker. Each conversation was restricted to 10 min. This corpus collection process was inspired by Meguro et al. (2009).

We defined five user intentions: 'greeting' (say 'hello' to user), 'self-disclosure' (express users preference and experience), 'informating' (providing information for the dialog), 'questioning' (asking questions to the listener), and 'else' (other utterances). Our definition of user intention also referenced Meguro et al. (2009). In total, 1281 utterances were collected from the speakers; 51.2% were self-disclosure, 32.7% were information, 7.6% were else, 5.7% were greetings, and 2.7% were questions.

We used the maximum entropy classifier (Ratnaparkhi, 1998) with word-n grams (uni-gram, bi-gram, and tri-gram) features to detect user intention.

3.2.2 Triple Extraction

We extracted arguments and their relation (triple) from user utterances. For example, a triple [I, like, Messi] is extracted from "I like Messi". These words are the subject, verb, and object of the sentence. We used ClausIE (Del Corro and Gemulla, 2013) to extract triples, then sent them to the natural language generation module.

3.2.3 Entity Recognition

To extract related information from the KB, the named entities in the user utterances were detected and recognized. Each entity was recognized by matching to an entity name in DBpedia, which is a structured database that contains data from Wikipedia. For example, when "I like Messi" is the input, the module detects "Messi" and matches it with "Lionel Messi", an entity of DBpedia (Auer et al., 2007). We used DBpedia Spotlight (Mendes et al., 2011) for entity detection and recognition. Recognized entities are sent to the related information extraction module.

3.3 Related Information Extraction

The related information extraction module takes a recognized entity as input, then extracts related information from the KB. We used Freebase (Bollacker et al., 2008) as our KB. Freebase is a database system which stores a public repository of the world's knowledge. Because Freebase includes DBpedia, we easily converted DBpedia entities to Freebase entities.

We should choose appropriate related information from Freebase. For example, when a user utterance includes the name of a football player, the topics of the system responses should also be about football players, or the player's position.

For the scenarios above, we extracted type, instances of the type, and properties of the type. For example, when the user talked about a football player, 'Lionel Messi', the system extracts type 'football player', instances of type 'David Beckham', 'Pélé', and other players, and properties such as 'position', 'matches played'.

We used 'notable type' of Freebase. Because an entity can have many types, we used a type that could be the best disambiguator. For example, 'Barack Obama' has multiple types: 'US President', 'Person', 'Politician', 'Author', 'Award Winner'. The 'notable type' that is the best disambiguator is 'US President'.

To generate a system response, we chose one instance and one property. The instance was chosen randomly from top-10 popular instances to find an instance that the user will find relevant interesting. We also chose one property randomly from properties whose object instance is in the top-10 popular instances. We used Freebase popularity score to get top-10 popular instances. Extracted information is sent to the language generation module.

3.4 Dialog Management

The dialog management module returns system intention based on interpretation of emotion and user intention. We generated a rule-based management strategy based on microskills (Algorithm 1) (Evans et al., 2010). Each system intention is given below:

Greeting: Say hello to user.

Attending: Encourage users to continue talking. For example, when a user enters "I watched Avatar", the system responses "Tell me more", "I see", or "Tell me about Avatar".

Paraphrasing: Reflect contents of user utterance. For example, "You watched Avatar", or "You watched Avatar, a movie".

Questioning: Ask questions from user utterance. For example, "Did you watch Titanic, too?", or "Do you know the genre of Avatar?" for same above input.

Reflect Feeling: Acknowledges the user's feeling. When user enters "I was so angry", or "I was annoyed", then the system response could "You were so mad".

Questioning Response: This is our additional strategy to deal with user questions. It changes topic or asks users to talk about themselves. For example, when user asks "Who is your mother?" to system, then questioning response could be "Let's just keep talking about Avatar", or "Why don't we talk about you?".

Algorithm 1 Dialog Management Strategy

```
if User Intention = Greeting then
  return Greeting
else if User Intention = Questioning then
  return Questioning Response
else if User Intention = Information then
  return Paraphrasing or Questioning
else if User Intention = Self Disclosure then
  if Emotion Detected then
    return Reflect Feeling
  else
    return Paraphrasing or Questioning
  end if
else
  return Attending
end if
```

3.5 Natural Language Generation

Our system response is generated using templates. A template is chosen by system intention, entity

System Intention	Entity Location	Related Information	Emotional Keyword	Response Template
Greeting	*	*	*	Nice to meet you.
Attending	*	type	*	Please tell me about the [type], [entity].
Paraphrase	*	*	*	[S] [V] [O].
Paraphrase	S	property	*	[entity], whose [property] is [object], [V] [O]
Questioning	O	instance, type	*	[S] [V] another [type], [instance], too?
Reflect Feeling	*	*	Sad	Oh, cheer up.
Questioning Response	*	*	*	Why don't we talk about you?

Table 1: System Utterance Template Examples. Slots: [S] subject of triple; [V] verb of triple; [O] object of triple; [object] object instance of Freebase triple ; * for don't care condition

Speaker	Utterance
User	Hello.
System	Hi. What do you want to talk about?
User	I want to talk about Messi.
System	Lionel Messi, a football player?
User	Yes. Messi is my favorite.
System	Is David Beckham your favorite, too?
User	No, I just like Messi.
System	What is Lionel Messi's position?
User	It's forward.

Table 2: Dialog Example

location, related information, and detected emotion (Table1). Entity location is the triple slot that includes entity. For example, when a user says, "I like Messi", then the entity location is the object; if the user says, "Messi is my favorite player", then the entity location is the subject. The system response is generated with filling slots with the extracted entity, an extracted triple, and related information. We changed words "I, my, myself, you" to "You, your, yourself, me" by applying simple rules. For example, when a user enters "I like Messi" and a system template "[S] [V] another [type], [instance], too?" is chosen, then system response generated is "You like another football player, David Beckham, too?".

3.6 Experiment and Results

We recruited another 15 students to evaluate our system, who did not join the dialogue generation task in Section 3.2.1. They chatted with three systems (ELIZA (Weizenbaum, 1966), Counseling Dialog System (Han et al., 2015), and our system) for 10 min, they rated each of them on three questions (Likert scale of 1 [low] to 10 [high]). The first question measured the variety of responses, the second question asked whether the system encouraged the user to continue talking, and the last question asked about overall satisfaction with the dialog. Our system got highest score for all questions (Figure 2).

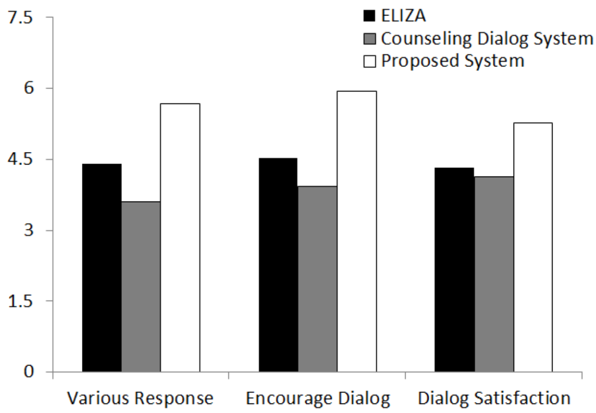

Figure 2: Averaged user experiment score.

3.7 Conclusion

We designed a natural language dialog listening agent that exploits the important and relevant information to the utterance from the KB. Results of our experiment indicated that our usage of a KB generated various responses and encouraged users to continue talking. Related information diversified the contents of system responses, and made users talk with the related information. Dialog satisfaction was increased by pin-pointing the content that user want to talk about.

Acknowledgments

This work was supported by the ICT R&D program of MSIP/IITP. [R0126-15-1117, Core technology development of the spontaneous speech dialogue processing for the language learning] and was partly supported by the ICT R&D progra of MSIP/IITP [14-824-09-014, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)]

References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. *Dbpedia: A nucleus for a web of open data*. Springer.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web*, pages 355–366. International World Wide Web Conferences Steering Committee.

David Evans, Margaret Hearn, Max Uhlemann, and Allen Ivey. 2010. *Essential interviewing: A programmed approach to effective communication*. Cengage Learning.

Curry Guinn and Rob Hubal. 2013. Extracting emotional information from the text of spoken dialog. In *Proceedings of the 9th international conference on user modeling*. Citeseer.

Sangdo Han, Kyusong Lee, Donghyeon Lee, and Gary Geunbae Lee. 2013. Counseling dialog system with 5w1h extraction. In *Proceedings of the SIGDIAL2013 Conference*, pages 349–353.

Sangdo Han, Yonghee Kim, and Gary Geunbae Lee. 2015. Micro-counseling dialog system based on semantic content. In *Proceedings of the IWSDS2015 Conference*.

Allen Ivey, Mary Ivey, and Carlos Zalaquett. 2013. *Intentional interviewing and counseling: Facilitating client development in a multicultural society*. Cengage Learning.

Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. 2009. Analysis of listening-oriented dialogue for building listening agents. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 124–127. Association for Computational Linguistics.

Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. Dbpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*, pages 1–8. ACM.

Adwait Ratnaparkhi. 1998. *Maximum entropy models for natural language ambiguity resolution*. Ph.D. thesis, University of Pennsylvania.

Erik Rautalinko and Hans-Olof Lisper. 2004. Effects of training reflective listening in a corporate setting. *Journal of Business and Psychology*, 18(3):281–299.

Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark Ter Maat, Gary McKeown, Sathish Pammi, and Maja Pantic. 2012. Building autonomous sensitive artificial listeners. *Affective Computing, IEEE Transactions on*, 3(2):165–183.

Joseph Weizenbaum. 1966. Eliza: computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.