User Information Extraction for Personalized Dialogue Systems

Toru Hirano, Nozomi Kobayashi, Ryuichiro Higashinaka,
Toshiro Makino and Yoshihiro Matsuo
NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation
1-1 Hikarinooka, Yokosuka-Shi, Kanagawa, 239-0847, Japan
{hirano.tohru, kobayashi.nozomi, higashinaka.ryuichiro, makino.toshiro, matsuo.yoshihiro}@lab.ntt.co.jp

Abstract

We propose a method to extract user information in a structured form for personalized dialogue systems. Assuming that user information can be represented as a quadruple ⟨predicate-argument structure, entity, attribute category, topic⟩, we focus on solving problems in extracting predicate argument structures from question-answer pairs in which arguments and predicates are frequently omitted, and in estimating attribute categories related to user behavior that a method using only context words cannot distinguish. Experimental results show that the proposed method significantly outperformed baseline methods and was able to extract user information with 86.4% precision and 57.6% recall.

1 Introduction

Recent research on dialogue agents has focused extensively on casual conversations or chat (Ritter et al., 2011; Wong et al., 2012; Meguro et al., 2014; Higashinaka et al., 2014) because chat-oriented conversational agents are useful for entertainment or counseling purposes. To make users want to talk to such conversational agents more, users and systems need to know each other well since it is important to build relationships of trust between users and systems (Bickmore and Picard, 2005).

In casual conversations between people, people sometimes talk about themselves such as mentioning their hobbies or experiences. Our manual examination of text-based casual conversation between two people indicated that 26% of utterances are self-disclosure utterances that convey information about the speaker. We also observed the same tendency in casual conversations between a person and a system.

On the basis of these findings, we propose a method to extract information about the speaker, that is, user information, from utterances in order to develop personalized dialogue systems using the extracted user information. For instance, for a user who said "I went to London.", "I live in Tokyo.", and "I love One Direction." (a pop band), we want to personalize conversations as follows.

Ex. 1 Telling users that the system remembers conversations in the past.
USER: I would like to go traveling!
SYSTEM: You **went to London** the other day, didn't you?

Ex. 2 Complementing unknown conditions with user information.
USER : What time does Frozen start?
SYSTEM : At 10 AM at **Tokyo Theater**.

Ex. 3 Providing information related to user interests.
USER : It's time to practice karaoke.
SYSTEM : A new song by **One Direction** is coming out soon.

To implement personalized dialogue systems such as this, the extracted user information should satisfy the following requirements.

1. It should have information to reproduce what users said in order to tell users their past utterances.
2. It should have information to complement unknown conditions.
3. It should have information that can be searched to determine which information to provide.
4. It should have information to determine when systems will use which user information.

To satisfy requirement 1, we extract the predicate-argument structure (PAS), which represents "who did what to whom." PAS is useful for representing the basic content of an utterance. Higashinaka et al. (2014) proposed a method to generate system utterances from PAS. For requirements 2 and 3, we extract entities such as person

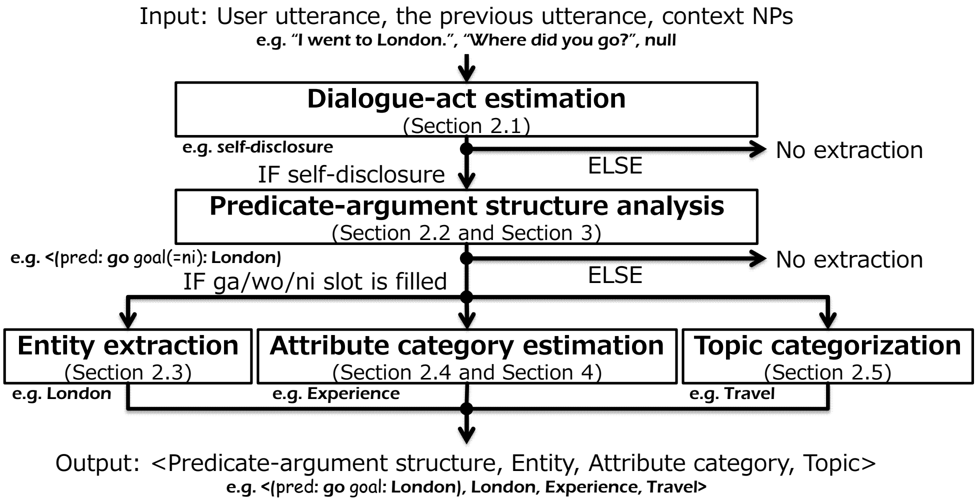

Figure 1: Overview of user information extraction.

and location names, which represent keywords. For requirement 4, we extract attribute categories such as hobbies and experiences, which represent aspects of users, to determine which information will be used, and we extract topics such as music and travel, which represent main subjects, to determine when to use the information.

Therefore, we extract a quadruple ⟨PAS, entity, attribute category, topic⟩ as user information from a user utterance. From the above examples, we extract the following information.

- "I went to London." → ⟨(pred: go goal: London), London, Experiences, Travel⟩
- "I live in Tokyo." → ⟨(pred: live locative: Tokyo), Tokyo, Place of residence, House⟩
- "I love One Direction." → ⟨(pred: love accusative: One Direction), One Direction, Hobbies/Preferences, Music⟩

In this paper, the work is done in Japanese although we want to apply our method to other languages in the future. For languages other than Japanese, instead of PASs, semantic role labeling (SRL) can be used (Palmer et al., 2010).

2 User Information Extraction

An overview of the method we propose to extract user information, ⟨PAS, entity, attribute category, topic⟩, from user utterances is shown in Figure 1. The method has five parts:

- Dialogue-act estimation
- Predicate-argument structure analysis
- Entity extraction
- Attribute category estimation
- Topic categorization

We focus on solving problems in analyzing predicate argument structures of question-answer pairs in which arguments and predicates are frequently omitted, and in estimating attribute categories related to user behavior that a method using only context words cannot distinguish. In this section, we outline the overall functionality of a user information extraction system; further methods to solve the problems are described in sections 3 and 4.

2.1 Dialogue-act estimation

We identify the dialogue-act of utterances to determine whether the input user utterance contains information about the user him/herself. We use the dialogue-act tag set consisting of 33dialogue-acts listed in Table 1, proposed by Meguro et al. (2014). Their tag set is designed for annotating listening-oriented dialogue, but because speakers in listening-oriented dialogue are allowed to speak freely, the tag set can cover diverse utterances, making it suitable for casual conversation.

We evaluate whether the input user utterance contains information about the user in two cases, as follows.

1. the dialogue-act of the user utterance is one of the self-disclosure tags: No. 3–11.
2. the dialogue-act of the user utterance is one of the sympathy/agreement tags: No. 22–23, and the dialogue-act of the previous utterance is one of the question tags: No. 14–19.

No.	Dialogue-acts
1	greeting
2	information
3	self-disclosure.fact
4	self-disclosure.experience
5	self-disclosure.habit
6	self-disclosure.preference.positive
7	self-disclosure.preference.negative
8	self-disclosure.preference.neutral
9	self-disclosure.desire
10	self-disclosure.plan
11	self-disclosure.other
12	acknowledgment
13	question.information
14	question.fact
15	question.experience
16	question.habit
17	question.preference
18	question.desire
19	question.plan
20	question.self-questioning
21	question.other
22	sympathy/agreement
23	non-sympathy/non-agreement
24	confirmation
25	proposal
26	repeat
27	paraphrase
28	approval
29	thanks
30	apology
31	filler
32	admiration
33	other

Table 1: Dialogue-act tag set.

We use a method proposed by Higashinaka et al. (2014) to estimate a dialogue-act. They trained a classifier using a support vector machine (SVM). The features used are word N-grams, semantic categories obtained from the Japanese thesaurus Goi-Taikei (Ikehara et al., 1999), and character N-grams.

2.2 Predicate-argument structure analysis

Predicate-argument structure (PAS) analysis involves detecting predicates and their arguments. A predicate can be a verb, adjective, or copular verb, and the arguments are noun phrases (NPs) associated with cases in case grammar. As cases, we use ga (nominative), wo (accusative), ni (dative), de (locative/instrumental), to (with), kara (source), and made (goal).

We use the PAS analyzer described by Imamura et al. (2014) to analyze PASs for general utterances. The analyzer works statistically by ranking NPs in the context using supervised learning with an obligatory case information dictionary and a large-scale word dependency language model. On the other hand, the analyzer cannot extract PASs correctly in order to analyze them for question-answer pairs in which arguments and predicates are frequently omitted. For example, predicate ellipsis is not targeted by the analyzer. Therefore, we use a method described in section 3 to analyze the PAS for question-answer pairs.

No.	Attributes	No.	Attributes
1	relationship:family	18	occupation
2	relationship:partner	19	place of business
3	relationship:lover	20	position in company
4	relationship:other	21	journey to work
5	name	22	biography
6	gender	23	earnings
7	age	24	expenditure
8	blood type	25	possessions
9	birthday	26	knowledge
10	constellation	27	hobbies/preferences
11	Chinese zodiac	28	habits
12	characters	29	experiences
13	physical description	30	strong points
14	home town	31	abilities
15	place of residence	32	opinions/feelings
16	house mate	33	desires
17	house type	34	other

Table 2: Attribute category tag set.

To extract user information, we need to select a PAS from the ones in the input utterances. We select the last PAS in the utterance on the basis of the observation that important information comes last in many Japanese utterances. Additionally, we should not output insufficient PASs in which argument slots are not filled at all because the extracted PASs would be used to generate system utterances. Therefore, we output PASs only when at least one of the argument slots (ga, wo, or ni) of the predicate is filled.

2.3 Entity extraction

We define an entity as a noun phrase (NP) that denotes the center word of a conversation. To extract the entity from the input user utterance, we use the center word extraction method proposed by Higashinaka et al. (2014). They extracted an NP from an utterance and trained a conditional random field (Lafferty et al., 2001); NPs are extracted directly from a sequence of words without creating a parse tree. The feature template uses words, part-of-speech (POS) tags, and semantic categories of current and neighboring words.

When no NP is extracted from the input user utterance, we try extracting an NP from previous utterances.

2.4 Attribute category estimation

We identify an attribute category, which represents aspects of users, for self-disclosure utterances of the user, e.g. "I went to London." → experi-

No.	Topics	No.	Topics
1	travel	23	disaster prevention
2	events	24	volunteering
3	movies	25	health
4	music	26	post-retirement
5	TV	27	beauty
6	entertainment	28	fashion
7	talent	29	shopping
8	computers	30	gourmet dining
9	games	31	anime
10	telephone	32	occult
11	business	33	gardening
12	study	34	sports
13	school	35	art
14	money	36	books
15	animals	37	cars/bikes
16	home	38	history
17	housekeeping	39	fishing
18	appliances	40	fortune-telling
19	family	41	religion
20	friends	42	general
21	love	43	other
22	politics		

Table 3: Topic category tag set.

ences. As an attribute category set, we define 34 categories of attributes in Table 2 on the basis of a questionnaire conducted in a market research study and on the analysis of personal questions (Sugiyama et al., 2014). The inter-annotator agreement with 200 self-disclosure utterances was 90.5% (Cohen's κ = 0.885). Because κ is more than 0.8, we can say the agreement is high.

We used a logistic-regression-based classifier to estimate attribute categories. We describe in section 4 the features used to estimate attribute categories related to user behavior that a method using only context words cannot distinguish.

2.5 Topic categorization

We identify a topic category, which represents the main subject, of the input user utterances, e.g. "I went to London." $\rightarrow$ travel. As a topic category tag set, we use 43 categories listed in Table 3 based on categories used on a Japanese question and answer communication site[1]. The inter-annotator agreement with 200 utterances was 93.0% (Cohen's κ = 0.925). Because κ is more than 0.8, we can say the agreement is high.

To categorize topics, we trained a classifier in the same way as done with attribute category estimation.

3 Analyzing Predicate-argument Structure of Question-answer Pairs

As mentioned in section 2.2, analyzing PASs for question-answer pairs in which predicates and arguments are frequently omitted is problematic. Although many prior studies have been done on PAS analysis (Taira et al., 2010; Hayashibe et al., 2011; Yoshikawa et al., 2011; Imamura et al., 2014), the methods they use could not be applied to analyze PASs of question-answer pairs with ease. For example, they could not extract the following PAS because of predicate ellipsis. The PAS (pred: read accusative: Fashion magazines) should be extracted from the example.

> "Do you read books?" - "Fashion magazines." $\rightarrow$ (pred: ϕ)

Types	Rate
completed	47.1% (115/244)
argument ellipsis	28.3% (69/244)
predicate ellipsis	7.8% (19/244)
yes-no	16.8% (41/244)

Table 4: Types of question-answer pairs.

To solve the problem, we break question-answer pairs down into the following four types, and we propose a method to analyze PAS for each of them except for the completed type.

Completed: Both the predicate and its arguments are included. (e.g. "What is your hobby?" - "My hobby is playing tennis.")

Argument ellipsis: A predicate is included, but the argument is omitted. (e.g. "Did you go to London last year?" - "I went with friends.")

Predicate ellipsis: A predicate is omitted, but the argument is included. (e.g. "Do you read books?" - "Fashion magazines.")

Yes-no: An answer is either "yes" or "no". (e.g. "Do you like to read books?" - "Yes.")

Table 4 lists the percentage of these four types among 244 question-answer pairs and indicates that predicates or arguments are omitted in 52.9% ($=$ 100% $-$ 47.1%) of the question-answer pairs.

The question-answer pairs had some typical forms such as "What do you like?" - "(I like) x.". We can accurately extract PASs from these typical cases using predefined extraction patterns. On the basis of these extractions, we propose a four-step method to analyze the PASs of question-answer pairs.

- pattern-based extraction: all types
- argument complement: argument ellipsis
- complete sentence generation: predicate ellipsis
- question PAS copying: yes-no

[1] http://oshiete.goo.ne.jp/

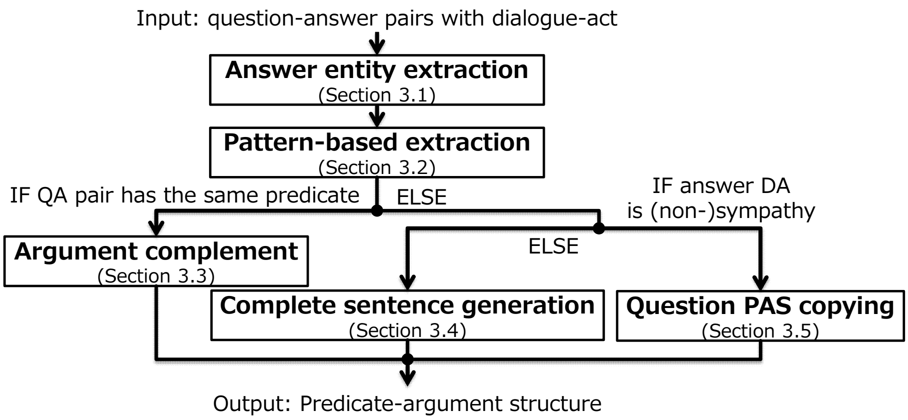

Figure 2: Process of PAS analysis.

Note that the method can be used to analyze the completed type as well as other types, because it cannot determine if it is the completed type before analyzing the PAS. Figure 2 outlines the PAS analysis process.

3.1 Pre-process: answer entity extraction

As a pre-process, we extract an answer entity from an answer utterance using named entity recognition since the answer is likely to be regarded as a named entity. We use the named entity recognition method proposed by Sadamitsu et al. (2013), which is based on Sekine's Extended Named Entity Hierarchy[2].

3.2 Pattern-based extraction method

In the pattern-based extraction step, an attempt is made to extract predicate-argument structures using pre-defined extraction patterns. If a pattern can extract a PAS, the extracted PAS is output as an answer.

We collected frequently appearing patterns in the Person-Database (Sugiyama et al., 2014) using the frequent-pattern mining method (Pei et al., 2001) and assembled 20 regular expression patterns by checking the collected frequent patterns. The Person-Database consists of a number of question-answer pairs created by 42 questioners and includes 26,595 question-answer pairs, which cover most of the questions related to the information about users.

The following is an example of a regular expression pattern.

> "What .* do you like?" - "$answer\ entity$"
> $\rightarrow$ (pred: like accusative: $answer\ entity$)

Here, $answer\ entity$ denotes an answer entity detected in the answer entity extraction.

We show the example, "What kind of food do you like?" - "Sushi." "Sushi" in the answer utterance is extracted as an answer entity. Therefore, (pred: like accusative: Sushi) is extracted as a PAS from this example.

When the pattern-based method extracts a predicate-argument structure, the steps described in the following subsections would be skipped.

3.3 Argument complement method

If an answer utterance has the same predicate that appeared in the question utterance, the argument complement step is executed.

This step compares the question PAS and the answer PAS that were analyzed using an existing predicate-argument structure analysis method, and complements the arguments that only appear in the question PAS. For example, when the question PAS is (pred: go goal: London) and the answer PAS is (pred: go with: friends), (pred: go goal: London with: friends) is generated by copying "goal: London" from the question PAS.

3.4 Complete sentence generation method

If an answer utterance does not have the same predicate that appeared in the question utterance and the dialogue act of the answer utterance is not "(non-)sympathy/agreement", the complete sentence generation step is executed.

When there is a predicate-ellipsis example, we generate a complete sentence by replacing a question expression with an answer entity. A ques-

[2] https://sites.google.com/site/extendednamedentityhierarchy/

tion expression consists of a question word (such as "what" or "how") and suffixes or nouns (such as "food" or "meter"). For example, given the question-answer pair "What kind of food do you like?" - "Sushi", "(I) like Sushi." is generated as a complete sentence by replacing the question expression "What kind of food" with the answer entity "Sushi" and then converting a question sentence into an affirmative sentence.

The question expression is extracted with a predefined question word list and extraction rules. The rules extract suffixes or nouns attached to a question expression as a question expression. We can obtain a PAS applying existing predicate-argument structure analysis methods to the generated utterance. From the above example, we can obtain the PAS (pred: like accusative: Sushi).

3.5 Question PAS copying method

If an answer utterance does not have the same predicate that appeared in the question utterance and the dialogue act of the answer utterance is "(non-)sympathy/agreement", the question PAS copying step is executed.

A yes-no type answer PAS is empty because the answer utterance is expressed by an interjection such as "yes" or "no". This case is regarded as a case in which a predicate and its arguments are both omitted, so the question PAS is output as the answer PAS. For the example, "Did you go to London?"- "Yes.", the question PAS (pred: go goal: London) is extracted as the answer PAS.

4 Estimating Attribute Categories Related to User Behavior

Attribute category estimation is used to identify an attribute category for self-disclosure utterances of the user. For example, the utterance "I went to London." should be categorized with an experiences tag. A simple approach to estimate the attribute category for self-disclosure utterances of the user is a logistic-regression-based classifier with word N-gram features and semantic category features obtained from the Japanese thesaurus Goi-Taikei (Ikehara et al., 1999), which are used for topic categorization. These context features are important clues for identifying 26 categories, No. 1–26 in Table 2, but they are not important clues for identifying the other categories, No. 27–34, which are related to user behavior.

For instance, the baseline method incorrectly classifies the utterance "I played tennis a little while ago.", which should be classified with an experience tag, and the utterance "I always play tennis.", which should be classified with a habit tag, because the context words in both utterances are the same, "play" and "tennis".

To solve this problem, we need to use features representing whether the user behavior has ended, is continuing, or was repeated. Therefore, we propose using semantic information of functional words and adverbs as features to classify attribute categories related to user behavior.

4.1 Semantic information of functional words

We use semantic information of functional words in our proposed method. In Japanese, "-ta" is a past tense expression that means the action was completed, and "-teiru" is a present tense expression that means the action is continuing, so semantic information of functional words would be important clues to classify attribute categories related to user behavior. In this paper, we use semantic labels of function words by analyzing function words using the method proposed by Imamura et al. (2011) as features. We assume that semantic labels of functional words would be important clues to classify attribute category, especially the semantic labels "completion" for the experiences category, "continuance" for the habits category, "supposition" and "admiration" for the opinions/feelings category, and "request" and "desire" for the desires category.

4.2 Semantic information of adverbs

We use semantic information of adverbs such as "a little while ago" or "always" in our proposed method. For instance, in our method, "always" expresses that the action is done on a daily basis, and "a little while ago" expresses time information about when the action was done. In attribute category estimation, we expect that adverbs expressing that the action is done on a daily basis would be important clues for the habits category, and adverbs expressing the time in which the action was done would be important clues for the experiences category.

We prepare in advance two lists of adverbs that are used in order to extract semantic information of adverbs: (A) a list of adverbs expressing that the action is done on a daily basis, e.g. "always" and "every day", and (B) a list of adverbs expressing the time the action was done, e.g. "a little while ago" and "before". Such lists represent the semantic information of adverbs, so we use the lists of extracted adverbs as features.

Types	Baseline Precision	Recall	F	Proposed Precision	Recall	F
completed	**89.2**% (149/167)	**87.6**% (149/170)	**0.884**	85.3% (139/163)	81.8% (139/170)	0.835
argument ellipsis	43.1% (66/153)	40.7% (66/162)	0.419	**63.1**% (94/149)	**58.0**% (94/162)	**0.605**
predicate ellipsis	26.3% (5/19)	19.2% (5/26)	0.222	**41.7**% (10/24)	**38.5**% (10/26)	**0.400**
yes-no	40.0% (30/75)	28.0% (30/107)	0.330	**67.0**% (61/91)	**57.0**% (61/107)	**0.616**
total	60.4% (250/414)	53.8% (250/465)	0.569	**71.2**% (304/427)	**65.4**% (304/465)	**0.682**

Table 5: Comparison of PAS analysis of question-answer pairs for baseline and proposed methods.

5 Experiments

5.1 Predicate-argument structure analysis of question-answer pairs

We investigated how effective the proposed method described in section 3 was in analyzing PAS of question-answer pairs by comparing it with a baseline method. The baseline method used was that of Imamura et al. (2014), which is described in section 2.2. This method analyzes the PASs of question-answer pairs as well as the other utterances.

We used 478 question-answer pairs in 480 casual dialogues between a person and a system (Higashinaka et al., 2014) to evaluate whether the system could extract PASs correctly.

Table 5 lists the performance results of both methods for various types of question-answer pairs. Precision is defined as the percentage of correct PASs out of the extracted ones. Recall is the percentage of correct PASs from among the manually extracted ones. The F measure is the harmonic mean of precision and recall.

A comparison between the baseline and proposed methods indicates that the F measure of the proposed method improved by 0.113 points. The use of a statistical test (McNemar Test) demonstrably showed the proposed method's effectiveness. Specifically, the proposed method increased the F measure by 0.186 points in the argument ellipsis type and by 0.286 points in the yes-no type.

Our error analysis indicated that 40% of errors consisted of a failure to include complementing arguments of predicates. For examples, from the pair "Did you have dinner tonight?" - "I ate a little while ago." the system would not extract the correct PAS because the predicate is not the same in the question (have) and the answer (eat). To solve this problem, we plan to evaluate whether two predicate-argument structures have the same meaning by applying paraphrase detection methods such as using recursive autoencoders (Socher et al., 2011). In addition, we plan to improve the handling of ellipsis and anaphora by incorporating methods that utilize syntactic structures (Dalrymple et al., 1991; Iida et al., 2007).

Method	Accuracy
baseline	76.0% (14,120/18,579)
proposed	**88.9**% (16,523/18,579)
upper bound (ref.)	90.5% (181/200)

Table 6: Accuracy of attribute category estimation.

5.2 Attribute category estimation

We also investigated the effectiveness of the proposed method described in section 4 when using semantic information of functional words and adverbs as features by comparing its results with those of the baseline method. To train a logistic-regression-based classifier, we used LIBLINEAR[3] with both methods.

We used 18,579 self-disclosure utterances as well as previous utterances from 4,160 casual dialogues: 3,680 dialogues between two people and 480 dialogues between a person and a system, annotated with 34 categories listed in Table 2. The number of utterances annotated for each category in decreasing order was: opinions/feelings 5,580 (30%); experiences 4,758 (25%); habits 2,414 (12%); and hobbies/preferences 2,234 (12%). We used the above self-disclosure utterances for training and testing by ten-fold cross validation.

Table 6 gives the accuracy of the baseline and proposed methods in estimating the attribute category, and the inter-annotator agreement as a referential upper bound. A comparison between the baseline and proposed methods indicates that the proposed method using semantic information of functional words and adverbs improved the accuracy by 12.9 points. The use of a statistical test (McNemar Test) demonstrably showed the proposed method's effectiveness. With the proposed method, the accuracy was greatly improved to 86.9% from 66.4% in the habits category and to 89.0% from 68.9% in the experiences category.

A comparison between the referential upper bound and the proposed method indicates that the

[3]http://www.csie.ntu.edu.tw/ cjlin/liblinear/

	Precision	Recall	F
baseline	57.4% (171/298)	34.2% (171/500)	0.429
proposed	**86.4**% (288/333)	**57.6**% (288/500)	**0.691**

Table 7: Performance of user information extraction.

proposed method is very close to the upper bound accuracy.

5.3 Overall performance of user information extraction system

To evaluate the overall functionality of a method implemented with the user information extraction system described in section 2, we used 500 user utterances randomly selected from 3,680 casual dialogues (Higashinaka et al., 2014) between two people, and annotated with PAS, entity, attribute categories, and topics.

Table 7 lists the performance results of the baseline and proposed methods in extracting user information from user utterances. A comparison between the two methods indicates that the proposed method improved the F measure by 0.262 points. The use of a statistical test (McNemar Test) demonstrably showed the proposed method's effectiveness.

This result demonstrates that the proposed method was able to extract user information with high precision, 86.4%, and moderate recall, 57.6%. User information extracted with such high precision would be useful for personalized dialogue systems, because when the extracted information is wrong, the system personalizes it in the wrong way.

The proposed method could not extract user information from 167 ($= 500 - 333$) utterances because of incorrect dialogue-act estimation (18 utterances) and PAS analysis (149 utterances). We need to solve these problems, especially in the PAS analysis, to extract more user information.

6 Related Work

Several studies have been done on extracting user information from user utterances (Weizenbaum, 1966; Wallace, 2004; Kim et al., 2014; Corbin et al., 2015). Chat bot systems such as ELIZA (Weizenbaum, 1966) and ALICE (Wallace, 2004) extract the user information, name, and hobby of the user by using predefined pattern rules to personalize casual dialogues. In these systems, since the extracted information is limited to match predefined pattern rules, new rules need to be added in order to extract new information.

In a dialogue system used to find out information on the colleagues of the user (Corbin et al., 2015), the system extracts where the user sits in an office and uses the extracted information to search a database for a personalized information service. In this study, the same problem exists in that the extracted user information is limited to where the user sits.

Kim et al. (2014) used open information extraction (OpenIE) techniques (Banko and Etzioni, 2008) to solve this problem. OpenIE extracts triples ⟨NP, relation, NP⟩ that include relation expressions between NPs, without using predefined pattern rules. Using this framework, the system was able to extract ⟨I, like, apples⟩ in a structured form from the utterance "I like apples." in order to generate system utterances. In this study, because their purpose is only to generate utterances directly from extracted user information, they do not extract attribute categories and topics. Thus, it can be said that our work expands the types of personalized conversation by extracting quadruples ⟨PAS, entity, attribute category, topic⟩.

Much research has been done on information search (Shen et al., 2005; Qiu and Cho, 2006) and recommendation (Ardissono et al., 2004; Jiang et al., 2011) in the research area of personalization. These studies represent user interests with word vectors by comparing a vector of user interests and document vectors and selecting a document that has a similar vector to a user interest vector. These methods can roughly capture user interests, but they cannot precisely capture user information.

7 Conclusion

This paper proposed a method to extract user information in a structured form, ⟨predicate-argument structure, entity, attribute category, topic⟩, for personalized dialogue systems. We focused in particular on the tasks of extracting predicate argument structures from question-answer pairs and estimating attribute categories from self-disclosure utterances of the user. The experiments demonstrated that the proposed method outperformed a baseline method in both tasks and that the method was able to extract user information from human-human dialogue with 86.4% precision and 57.6% recall.

In future, we plan to implement a personalized dialogue system using extracted user information. We also want to solve the problems in PAS analysis to extract more user information and to apply our method to other languages.

References

Liliana Ardissono, Cristina Gena, Pietro Torasso, Fabio Bellifemine, Angelo Difino, and Barbara Negro. 2004. User modeling and recommendation techniques for personalized electronic program guides. In *Personalized Digital Television - Targeting Programs to Individual Viewers*, volume 6 of *Human - Computer Interaction Series*, pages 3–26.

Michele Banko and Oren Etzioni. 2008. The tradeoffs between open and traditional relation extraction. In *Proceedings of the 46th Annual Meeting on Association for Computational Linguistics: Human Language Technologies*, pages 28–36.

Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327.

Carina Corbin, Fabrizio Morbini, and David Traum. 2015. Creating a virtual neighbor. In *Proceedings of the 2015 International Workshop Series on Spoken Dialogue Systems Technology*.

Mary Dalrymple, Stuart M. Shieber, and Fernando C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452.

Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2011. Japanese predicate argument structure analysis exploiting argument position and type. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 201–209.

Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an open domain conversational system fully based on natural language processing. In *Proceedings of the 25th International Conference on Computational Linguistics*, pages 928–939.

Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2007. Zero-anaphora resolution by learning rich syntactic pattern features. *ACM Transactions on Asian Language Information Processing*, 6(4):1–22.

Satoru Ikehara, Masahiro Miyazaki, Satoru Shirai, Akio Yoko, Hiromi Nakaiwa, Kentaro Ogura, Masafumi Oyama, and Yoshihiko Hayashi. 1999. *Nihongo Goi Taikei (in Japanese)*. Iwanami Shoten.

Kenji Imamura, Tomoko Izumi, Genichiro Kikui, and Satoshi Sato. 2011. Semantic label tagging to functional expressions in predicate phrases. In *Proceedings of the 17th Annual Meeting of Association for Natural Language Processing (in Japanese)*, pages 518–521.

Kenji Imamura, Ryuichiro Higashinaka, and Tomoko Izumi. 2014. Predicate-argument structure analysis with zero-anaphora resolution for dialogue systems. In *Proceedings of the 25th International Conference on Computational Linguistics*, pages 806–815.

Yechun Jiang, Jianxun Liu, Mingdong Tang, and Xiaoqing Liu. 2011. An effective web service recommendation method based on personalized collaborative filtering. In *Proceedings of the 2011 IEEE International Conference on Web Services*, pages 211–218.

Yonghee Kim, Jeesoo Bang, Junhwi Choi, Seonghan Ryu, Sangjun Koo, and Gary Geunbae Lee. 2014. Acquisition and use of long-term memory for personalized dialog systems. In *Proceedings of the 2014 Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*.

John Lafferty, Andrew McCallum, and Fernando C.N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289.

Toyomi Meguro, Yasuhiro Minami, Ryuichiro Higashinaka, and Kohji Dohsaka. 2014. Learning to control listening-oriented dialogue using partially observable markov decision processes. *ACM Transactions on Speech and Language Processing*, 10(4):15:1–15:20.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103.

Jian Pei, Jiawei Han, Behzad Mortazavi-asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei chun Hsu. 2001. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *Proceedings of the 17th International Conference on Data Engineering*, pages 215–224.

Feng Qiu and Junghoo Cho. 2006. Automatic identification of user interest for personalized search. In *Proceedings of the 15th International Conference on World Wide Web*, pages 727–736.

Alan Ritter, Colin Cherry, and William B. Dolan. 2011. Data-driven response generation in social media. In *Proceedings of the Conference on Empirical Methods in Natural Language*, pages 583–593.

Kugatsu Sadamitsu, Ryuichiro Higashinaka, Toru Hirano, and Tomoko Izumi. 2013. Knowledge extraction from text for intelligent responses. *NTT Technical Review*, 11(7):1–5.

Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 824–831.

Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Proceedings of the Advances in Neural Information Processing Systems 24*, pages 801–809.

Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Higashinaka, and Yasuhiro Minami. 2014. Large-scale collection and analysis of personal question-answer pairs for conversational agents. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents*, pages 420–433.

Hirotoshi Taira, Sanae Fujita, and Masaaki Nagata. 2010. Predicate argument structure analysis using transformation-based learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 162–167.

Richard S. Wallace. 2004. *The Anatomy of A.L.I.C.E. A.L.I.C.E.* Artificial Intelligence Foundation, Inc.

Joseph Weizenbaum. 1966. Eliza — a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, 9:36–45.

Wilson Wong, Lawrence Cavedon, John Thangarajah, and Lin Padgham. 2012. Strategies for mixed-initiative conversation management using question-answer pairs. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 2821–2834.

Katsumasa Yoshikawa, Masayuki Asahara, and Yuji Matsumoto. 2011. Jointly extracting japanese predicate-argument relation with markov logic. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 1125–1133.