World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:3, 2009

An Integrated Natural Language Processing Approach for Conversation System

Zhi Teng, Ye Liu, and Fuji Ren

Abstract—The main aim of this research is to investigate a novel technique for implementing a more natural and intelligent conversation system. Conversation systems are designed to converse like a human as much as their intelligent allows. Sometimes, we can think that they are the embodiment of Turing's vision. It usually to return a predetermined answer in a predetermined order, but conversations abound with uncertainties of various kinds. This research will focus on an integrated natural language processing approach. This approach includes an integrated knowledge-base construction module, a conversation understanding and generator module, and a state manager module. We discuss effectiveness of this approach based on an experiment.

Keywords—Conversation System; integrated knowledge-base construction; conversation understanding and generator; state manager.

I. Introduction

A Conversation system is a computer system intended to converse with a human, with a coherent structure. Conversation systems have employed text, image, speech, emotion, gestures and other modes for communication on both the input and output channel.

Man-machine-conversation system has become more important in 21 century. Several speech dialogue systems and web agent have been developed in order to provide a very easeful interaction for dissimilar users. The primary aim is to satisfy the user without or at least with reduced human involvement [1]. A few well known scientists believe that computers will achieve capabilities comparable to human reasoning and understanding of languages by 2020 [2]. We think a more humanoid interactive system will be achieved along with recent advances in Natural Language Processing (NLP) and Artificial Intelligence (AI) field.

II. Related Work

The ability for computers to converse with users using natural language would arguably increase their easiness. Research in practical conversation systems has gained much attention in recent years [3]. Most of the conversation systems today typically focus on helping users to complete a specific task, such as information search, manual, planning, sightseeing or diagnosis [4].

Zhi Teng and Ye Liu are with the Graduate School of Advanced Technology and Science, University of Tokushima 2-1 Minamijosanjima, Tokushima, 770-8506, Japan. e-mail: (teng@is.tokushima-u.ac.jp; liuye@is.tokushima-u.ac.jp)

Fuji Ren is with Institute of Technology and Science, University of Tokushima 2-1 Minamijosanjima, Tokushima, 770-8506, Japan and School of Information Engineering Beijing University of Posts and Telecommunications Beijing, 100876, China. e-mail: (ren@is.tokushima-u.ac.jp)

However original conversation systems were designed like a embodiment of Turing's vision, it usually to return a predetermined answer in a predetermined order and usually used the handcrafted corpus. And the original conversation system which as a system that was designed to converse like a human as much as their intelligent allows did not used the comprehensive information such as semantic information, context information, emotion information, image information or environment information etc.

In recent years, some researchers did lots of works to develop a more natural and intelligent conversation system. Teruhisa Misu built a speech-based interactive information guidance system based on context information and Face Registration function. However, this system did not use semantic or emotion information, and only used a handcrafted corpus [5]. Hidekazu Kubota presented a computational approach to understanding and augmenting the conversational knowledge process that was a collective activity for knowledge creation, management, and application where conversational communications were used as a primary means of interaction among participating agents. This system was designed like a Turing, the conversation manner is one question and one answer, and he did not discuss the semantic or context processing [6]. For a conversation systems that using emotion information, Laurence Devillers worked aimed at relating multi-level dialog annotations with meta-data annotations for a corpus of real human-human dialogs. A corpus of 100 agent-client dialogs had been annotated with three types of annotations and five emotions types were annotated at the utterance level [7]. Aims at conversations system abound with uncertainties of various kinds, Tim Paek propose a task independent, multimodal architecture for supporting robust continuous spoken dialog. He used four interdependent levels of analysis, and described representations, inference procedures, and decision strategies for managing uncertainties within and between the levels [8].

The problem of conversation systems can be approached from different dimensions [9], [10], [11]. Generally, conversation systems can be categorized into two groups based on their approaches. The first approach is based on simple natural language processing and information retrieval. The second one is based on natural language understanding and reasoning [4]. In this paper, we propose a new approach for conversation system except the two groups. This new approach includes an integrated knowledge-base construction module, a conversation understanding and generator module, and a state manager module. The main aim of this research is to investigate a novel technique for implementing a more natural and intelligent conversation system. Table I summarizes the

TABLE I
CHARACTERISTICS OF THE TWO APPROACHES AND PROPOSED IN CONVERSATION SYSTEM.

Dimensions	Simple Natural Language Processing and Information Retrieval	Natural Language Understanding and Reasoning	Our proposed
Technique	Syntax Processing, Named Entity Tagging and Information Retrieval	Semantic Analysis, and Reasoning	Emotion Recognition, Image Processing, and Multimedia
Source	Text documents	Knowledge base	Web and Knowledge base
Domain	Open Domain	Special Domain	Semi-open Domain
Response	Extracted snippets	Synthesized responses	Intelligent responses

characteristics of the two approaches and our proposition with respect to different dimensions.

III. OVERVIEW OF PROPOSED APPROACH

An overview can be seen in Fig.1. The approach is broken down into three main modules: integrated knowledge-base construction, conversation understanding and generator, and state manager.

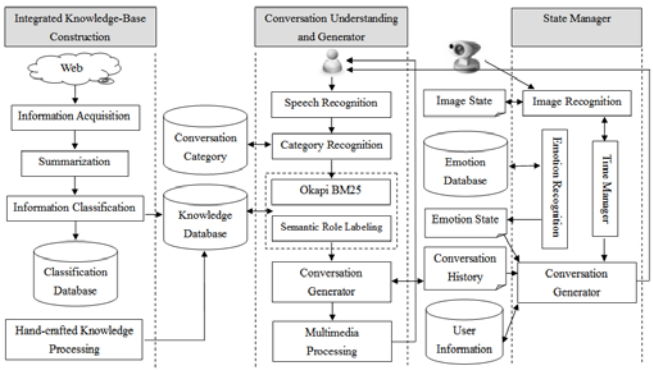

Fig. 1. Overview of proposed approach

Integrated knowledge-base construction is included by a hand-crafted knowledge processing and a special domain knowledge collections module. The special domain knowledge collections module is designed to be self sufficient, meaning it does not require an editor to be involved in organizing the information nor does it require any intervention during its operation. Summarization and information classification are the heart the main focus of this part.

Conversation understanding and generator module performs three-phase processing: First, the module predicts the categories of conversation content through the classification processing. Second, find out the befitting responses from the knowledge database in the same categories through similarity computing. Finally, the system will respond the user by multimedia technique.

State manager module is designed to achieve a natural conversation, at least in particular situations, seems to the human user of a conversation system as though it could come from another human. This module controls the conversation flow by the emotion information, image information and time.

IV. INTEGRATED KNOWLEDGE-BASE CONSTRUCTION

A. Hand-crafted Knowledge Processing

The integrated Knowledge Database construction is designed to build Knowledge Database by hand and web. There are four kinds of knowledge in our knowledge database. Three kinds of knowledge are about the self-introduction, the lab introduction and the phrase; they made by hand and we input them into the knowledge database through the interface. The residuary knowledge is the disaster information; they are automatically made by our system from the internet RSS information everyday. They will be explained in more detail in the section B.

An example of self-introduction: My name is ”UB”.

An example of lab introduction: I am born in A1 laboratory. The A-1 Laboratory’s research deals with Multi-function, Multi-lingual, Multi-media intellectual systems using Machine Translation and Speech Recognition as the core technologies.

An example of phrase: Nice to meet you.

B. Special Domain Knowledge Collections

In recent years, web is increasingly seen as a good source for creating corpora, as seen by the birth of workshops such as ACL’s ”Web as Corpus.” Researchers are also more and more seeing the Web as a way of creating special corpora, such as [12], [13], [14]. The special domain knowledge collections module is made up of three subsystems; Information Acquisition, Summarization and Information Classification.

1) Information Acquisition: RSS is a lightweight XML format designed for sharing headlines and other Web content. RSS solves myriad problems webmasters commonly face, such as increasing traffic, and gathering and distributing news. RSS can also be the basis for additional content distribution services [1].

RSS is a dialect of XML. Each RSS text file contains both static information about your site, plus dynamic information about your new stories, all surrounded by matching start and end tags. A channel may contain any number of $\langle item \rangle$s. An item may represent a ”story”: much like a story in a newspaper or magazine. Each story is defined by an $\langle item \rangle$ tag, which contains the required channel elements about a headline Title, URL, and Description. Here’s an example:

...
$\langle item \rangle$
$\langle title \rangle$ Venice Film Festival Tries to Quit Sinking$\langle /title \rangle$
$\langle link \rangle$ http://nytimes.com/2004/12/07FEST.
html$\langle /link \rangle$
$\langle description \rangle$ Some of the most heated chatter at the Venice Film Festival this week was about the way that the arrival of the stars at the Palazzo del Cinema was being staged. $\langle /description \rangle$
$\langle /item \rangle$

[1] http://www.webreference.com/authoring/languages/xml/rss/intro/

...

In our system, we do parsing and retrieving web corpus containing news articles across all categories from XIN HUA online, BBC Chinese.com, Baidu RSS News and SOHU RSS. We only do parsing, retrieving and download title, date and URL from the item. Then we use the URL to download the HTML code. The web page is downloaded and read using the encoding specified in the template. Next, the article is extracted using regular expressions and the extracted articles and information are saved in MySQL Database format.

2) Summarization: After the article acquisition, the automatic information classification measure will be used to determine which of the articles are in the disaster domain and which are not. We did an experiment using the general classification measure by news's title, but we can not get good results. We think there are some shortcoming to class only use the title. It is generally believed that the news's title is short and laconic, but we think we can not comprehend completely essential information from news through the title only. The summarization can describe the important and complete information of the news. We think we can use the summarization to improve the classification and get a better result.

In here, we used an approach for automatic summaries based on sub-topics identification and word frequency methodology. Documents can be clustered into local topics after sentences similarity calculating, which can be sorted by the scoring. Then sentences from all local topics are selected by computing the word frequency.

(1). Sub-topics Identification

One document can be composed of some subdocument. The described contents in each of the documents laid special emphasis on different aspects although these documents were all surround the same topic. It is obvious that document is composed of different side information and the different side information is the sub-topics. The sub-topics structural document is a more logical and convenient structure format. This approach can deal with the redundancy in the large documents too.

After the similarity of sentences is measured by the VSM method, the sub-topics are found by sentence clustering.Here we defined a threshold value named η. After the similarity measurement, we incorporate the sentences in the same local topics if the similarity value of sentence is larger than η. The optimal value of η, 0.6, was determined using the similarity values of 2000 pairs of sentences.

(2). Sub-topics Score

Here we have to compute and order each sub-topic, circular pick the best sentence from the best sub-topics if the desired summary length has not been reached.

SumFocus: In the multi-document summarization system that made by Lucy Vanderwende [15], a new approach they named SumFocus, captures the information conveyed by the topic description by computing the word probabilities of the topic description. The weight for each word is computed as a linear combination of the unigram probabilities derived from the topic description, with back off smoothing to assign words not appearing in the topic a very small probability, and the unigram probabilities from the document, in the following manner:

$$WordWeight = (1-\lambda) \times DocWeight + \lambda \times TopicWeight \quad (1)$$

The optimal value of λ, 0.9, was empirically determined using the DUC2005 corpus, manually optimizing on ROUGE-2 scores.

Sentence Location Feature: Sentence Location Feature is also important except words occurring frequently. According to the statistics, in Chinese news, the probability that the first sentence can be picked and make summarization is about 85%. The probability that the last sentence can be picked and make summarization is about 7% [16]. Thus we have to adjust the sentence weight that in the especial location. We use the following algorithm:

$$Weight(L_j) = (1 + \frac{(P - 0.75m)^2}{m^2}) \times ST_j \quad (2)$$

For each sentence Lj in the local topics, P is the serial number of sentence Lj in the document. m is the number of sentence in the document. STj is the similarity value between sentence Lj and title of the document.

An Improved SumBasic: The improved SumBasic approach is described in the following manner:

Step1. Compute the probability distribution over the words W_i appearing in the document used formula (1).

Step 2. For each sentence S_j in the local topics, compute the Weight(S_j) and Weight(L_j), calculate the scoring of each sentence use

$$Score(LocalTopic) = \sum(\alpha Weight(S_j) + \beta Weight(L_j)) \quad (3)$$

and the scoring of each local topic. (We find out that we can get the best result when $\alpha = 0.9$ and $\beta = 0.1$ through some experimentations by used different α and β.)

Step 3. Pick the best scoring sentence that contains the highest probability word from the best scoring local topic.

Step 4. Delete this local topic and all sentences in this local topic.

Step 5. For each word wi in the sentence chosen at step 3, update their probability:

$$P_{new}(w_i) = P_{old}(w_i) \times P_{old}(w_i) \quad (4)$$

Step 6. If the desired summary length has not been reached, go back to Step 2.

Step 7. Use the sentences that picked up by Step 3 to generate summarization.

3) Information Classification: Category classification is a type of text classification problem in which news articles are assigned one or more predefined news categories. Typical information categories only include disaster and non-disaster, total two categories news information. For classification, we prepared 600 example sentences that the category of sentence has been rightly classed by us. The category classification algorithm used Support Vector Machines (SVMs).

In recent years, SVMs have been successfully applied to a lot of applications such as particle identification, face identification and text categorization to engine knock detection, bioinformatics and database marketing. We used libsvm [17],

[18], a library for SVMs, because of its support for multi-class classification. It is a based classification method that only requires positive and unpositive examples for training data. It has the added ability that it is capable of updating the training information as well as easily adding new categories without a complete retraining. This means that as the needs of the users grows or changes, categories can be easily added or deleted from the system.

V. Conversation Understanding and Generator

A. Speech Recognition and Synthesis

When we converse with a computer, we hope for a natural conversation, we mean that at least we can not just satisfy with use the text. Speech recognition is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program. The first part of conversational processing is speech recognition. We used Microsoft Speech SDK to make a user's speech signal into text. Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech (TTS) system converts normal language text into speech.

B. Category Recognition

There are four kinds of knowledge in our knowledge database, the self-introduction, the lab introduction, the phrase, and the disaster information. For category recognition, we used the SVMs theory that has been explained in the section IV. For training data, we prepared 500 example sentences that the category of sentence has been rightly classed by us for the four categories. These example sentences have been processed, the keywords have been transformed into the specified format of Libsvm and the "Model" file would be generated. Then, category recognition module predicts the conversation category used the "Model" file by Libsvm.

C. Similarity Computing

After category recognition processing, we have to find out the befitting responses from the knowledge database in the same categories. Our method is based on the Semantic Role Labeling (SRL) and the Okapi BM25 measure slightly modified. The similarity score measure is as follows:

$$S(c,d) = \alpha \sum SR + BM25(c,d) \quad (5)$$

In equations (5) $S(c,d)$ meaning the score of conversation c and document d; SR is the same semantic role in the conversation c and document d. $BM25(c,d)$ is the similarity score by used Okapi BM25 measure; α are set to 0.5.

After the similarity processing, system will find out the befitting responses from the knowledge database in the same categories. Finally, system will respond the user by multimedia processing.

1) Okapi BM25: The Okapi BM25 measure slightly modified measure is as follows:

$$bm25(q,d) = \sum_{t \in q} \log\left(\frac{N - f_t + 0.5}{f_t + 0.5}\right) \times \frac{(k_1+1) f_{d,t}}{K + f_{d,t}} \quad (6)$$

In equations (6) terms t appears in query q; the collection contains N documents d; f_t documents contain a particular term and a particular document contains a particular term $f_{d,t}$ times; K is $k_1((1-b) + b \times Ld/AL)$; constants k_1 and b respectively are set to 1.2 and 0.75; and Ld and AL are measurements in a suitable unit for the document length and average document length respectively [19].

2) Semantic Role Labeling: SRL problem has been the topic of the both the CoNLL-2004 and the CoNLL-2005 Shared Tasks. General, the labels used in the Propbank annotation. The Propbank use predicate-specific labels ARG0, ARG1, ... ARGn for arguments and ARGM combined with a secondary tag to mark adjunct-like elements. For Chinese SRL we used the Chinese PropBank (CPB) [20]. The labeling theory based on the SVMs.

We created our training corpora by the Chinese Propbank 2.0; in total we selected 13761 sentences for training sets. Semantic roles in the Chinese PropBank are grouped into two major types [21]:

(1) Arguments: which represent central participants in an event. A verb may require one, two or more arguments and they are represented with a contiguous sequence of numbers prefixed by arg, as arg0 (Subject), arg1 (Object).

(2) Adjuncts: which are optional for an event but supply more information about an event, such as time, location, reason, condition, etc. An adjunct role is represented with argM plus a tag. For example, argM-TMP (Temporal) stands for temporal, argMLOC (Locative) for location.

The assignment of numbered argument labels is illustrated in Fig.2., where the predicate is the verb "investigate".

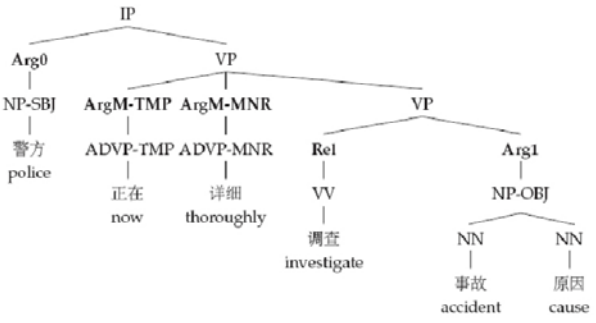

Fig. 2. An instance of the verb "investigate"

In some preliminary work on Chinese SRL, some researchers have successfully adapted a number of features to Chinese. The features that we have used are listed below:

Position: The position is defined in relation to the predicate verb and the values are before and after.

Path: The path between the constituent in focus and the predicate.

Head word and its POS: The head word and its part-of-speech

is often a good indicator of the semantic role label of a constituent.
Predicate and its POS: The verb itself and its part-of-speech.

D. Multimedia Processing

The information feedback mode is very important to conversation system too. A conversation system will not only use the text to talk with the user but also can use the media data such as speech, picture, video, and Web page etc. In our system we used four kinds of media data: text, speech, video and Web page. If conversation content included a Web link or a video, a prompt can be showed on the interface and click the prompt the user can see the relevant WEB page or the video.

The speech synthesis is an indispensable part for multimedia conversation system. In this system a speech interface for the conversation system was implemented along with the text interface, using an acoustic model HMM (Hidden Markov Model), a pronunciation lexicon and a language model FSN (Finite State Network) on the basis of the feature of Chinese sentence patterns [22].

VI. State Manager

A. Image Recognition

The image recognition is to judge the status of the user through the motion range of the learner's face in an image. The system performs a three-phase processing. First, time manager takes 7 seconds of continuous facial images of the user by a web camera and records the images to memory. Secondly, the image date is transmitted to the Data Analyzer system to be analyzed. The analyzer checks the user's facial movement to see if it falls within an acceptable range. If the user's facial movement falls within an acceptable range, we think the user is here yet. If the user's facial movement dose not fall in the acceptable range then it is assumed that the user have left.

B. Emotion Recognition

In our system, we used four broad emotion categories, impassibility, happiness, anger and sadness. Happiness class would have "blessedness", "happiness", "pleasure" and other positive emotions. The theory of emotion recognition is emotion classification with the Rough Set Theory and SVMs theory[25].

For emotion recognition, we prepared 800 emotional sentences for trainings. There were 200 sentences of impassibility, 200 sentences of happiness, 200 sentences of anger and 200 sentences of sadness. The 800 emotional sentences have been rightly classed by us.

C. Responses Generator

The state manager module will be activated by two kinds of situations: a sudden change in the emotion of user and unanswered from user. In the first case, the conversation generation module will automatically generate a dialogue by an emotional responses rule. The emotional responses rule is a easy rule for example: if the emotion state of user is the anger, the system will respond by "why are you angry?", the rest may be deduced by analogy.

TABLE II
The responses rule

Responses content	Index
"What is your name?"	1
"How old are you? "	2
"What is your hobby? "	3
"Do you know my name? "	4
"Do you know my lab? "	5
Generate by a similarity computing	6
Generate by the big news today	7

In the other case, the time manager will be activated if the user did not carry out any operation in 8 seconds. If the user did not carry out any operation, the system will actively generate a dialogue used a responses rule by index if the image recognition module judged user be, in an opposite direction system will not do any responses if the image recognition module judged user have left. The responses rule included two kinds of situations: if it is the first time user use system, the conversation generator module will actively generate some dialogue which can start understanding each other by index. For example: "Do you know my name?" or "What is your hobby?" etc. If the user and system have known each other, the conversation generator module will actively generate a dialogue by a similarity computing that find the most similar topic which was automatically made by system today with the conversation history. The approach of similarity computing is same as the approach that we have described in the section V. If system can not find out the right topic, the system will use the big news today to generate a response. Table II describes the responses rule. We think that the system can make a more natural and intelligent conversation with user by our state manager.

VII. Experimentation

The experiment object is that we want to test the feasibility and validity of the proposed approach.

A. Corpus

For summarization experimentation, a corpus of 341 disaster texts in Chinese was selected from a natural disaster database. The natural disaster database contains approximately 366 texts and all of them were downloaded from news website. Of the 366 natural disaster texts, 341 particular texts were deliberately selected as training data. For non-disaster training data, we searched and extracted 355 general news such as business, securities, sports, health, travel, science, etc., total 696 classification training data. About 150 number of web information were extracted as disaster data from about 2000 news by our system everyday. A 5 days web news data was used to evaluate the experimental results (Table III).

For integration performance experimentation of conversation system, we prepared 680 number of data as knowledge data. The Table IV shows the number of integrated knowledge-base content of each category.

TABLE III
THE 5 DAYS INFORMATION DATA.

$1^{st}Day$	$2^{nd}Day$	$3^{rd}Day$	$4^{th}Day$	$5^{th}Day$	$Total$
158	146	144	144	163	755

TABLE IV
THE NUMBER OF KNOWLEDGE DATA OF EACH CATEGORY.

Self-introduction	Lab introduction	Phrase	Disaster	Total
150	150	80	300	680

B. Evaluation

The summarization results have an important influence on the quality of the classification. In order to evaluate our proposed approach, we used two evaluation criterion, $Vpercent$ and $Hpercent$ [23]. $Vpercent$ is the coverage ratio of valid word (exclude stopwords). $Hpercent$ is the coverage ratio of high frequency word (the valid word that appears more than twice). The measure is as follows:

$$Vpercent = \frac{SVWords}{TVWord} \times 100\%$$

$$Hpercent = \frac{SHWords}{THWord} \times 100\%$$

$TVWord$: the number of valid words in the original texts
$THWord$: the number of high frequent words in the original texts
$SVWords$: the number of valid words in the summarization
$SHWords$: the number of high frequent words in the summarization

For integrated knowledge-base construction experimentation we get a result report on accuracy, precision, recall and F1. Accuracy-This measures the portion of all decisions that were correct decisions. Precision-This measures the portion of the assigned categories that were correct. Recall-This measures the portion of the correct categories that were assigned. F1-This measures an even combination of precision and recall. It is defined as $\frac{2\times p\times r}{p+r}$.

There are not have a standard evaluation approach in the conversation system field now. In this paper, we present a solution to evaluate the conversation system quality: a approach through observation, and classification with a scoring mechanism. This black-box approach is based on the work of Diekema [24]. We used four kinds of data to evaluate the system; there are system performance, responses, knowledge-base content, and expectations. System performance is the category that deals with system speed and system availability. For The users not only wanted responses to be accurate, they also wanted them to be complete. Users also shared thoughts about the knowledge-base content that are used for responses. They find it important that these contents are reputable. They also shared concerns about the size of the knowledge-base, fearing that a limit in size would restrict the number of responses. Another interesting aspect of user criteria is expectations, user hope use the system and get some information by some instrumentality. For example, speech recognition and synthesis, the video, the Web page etc. Expectations can be captured by survey so that it can be established whether these expectations are reasonable and whether they can be met.

C. Results

In order to compute the $Vpercent$ and $Hpercent$, we made a program to automatically evaluate the result. To compare the result of our approach with existed methods, the Top-N method was used to build a reference summarization used the same disaster texts. The method of Top-N is that the first sentence in each paragraph in document is taken in turn until the number of sentences is satisfied. The order of paragraphs is random. If the number of sentences is not satisfied, repeat the process on the second sentence of each paragraph. The Table V shows the coverage ratio of valid word and high frequency word based on Top-N and our approach.

TABLE V
THE VPERCENT AND HPERCENT RESULTS BASED ON TOP-N AND OUR SYSTEM.

	Vpercent TOP-N (%)	Proposed (%)	Hpercent TOP-N (%)	Proposed
Data 1	31.9	39.5	19.2	29.8
Data 2	27.5	38.8	15.6	28.7
Data 3	28.7	38.1	15.9	27.9
Data 4	29.4	39.2	16.5	28.5
Data 5	30.6	38.9	18.7	29.1
Average	29.6	38.9	17.2	28.8

For comparison purposes, a TFIDF based collection method and general classification method were used. To use the TFIDF based algorithm, a 3,000 document corpus containing news articles across disaster and non-disaster categories and topics was created. The general classification method used the same training data presented in Corpus (section VII.A). The averaged precision, recall and F1 of non-disaster and disaster for TFIDF, general classification and proposed approach were presented in Table VI.

TABLE VI
THE AVERAGED PRECISION, RECALL AND F1 OF TFIDF, GENERAL CLASSIFICATION AND PROPOSED.

			TFIDF			
	Data 1	Data 2	Data 3	Data 4	Data 5	Average
Precision (%)	63.5	62.2	59.2	60.5	62.2	61.5
Recall (%)	67.3	65.8	60.8	61.3	64.7	64.0
F1 (%)	65.35	63.9	60.02	60.9	63.4	62.7
			Classification			
	Data 1	Data 2	Data 3	Data 4	Data 5	Average
Precision (%)	74.1	72.6	63.8	71.5	74.2	71.2
Recall (%)	81.2	75.7	67.2	74.6	79.1	75.6
F1 (%)	77.5	74.1	65.5	73.0	76.6	73.3
			Proposed			
	Data 1	Data 2	Data 3	Data 4	Data 5	Average
Precision (%)	85.0	82.6	80.8	81.5	83.5	82.7
Recall (%)	94.5	89.8	92.8	87.0	90.1	90.8
F1 (%)	89.5	86.0	86.4	84.2	86.7	86.6

To evaluate the conversation system quality, we propose an evaluation approach based on four kinds of data. This makes the scheme generally applicable to evaluation of mostly conversation systems with different approaches by general users. Under this scheme, we define three general categories BQ, CQ and LQ, which represent the best, commonly and the lowest quality response for each data respectively. And we invited 20 students who will give the evaluation by use this system. Before the students converse with this system, they

TABLE VII
THE RESULT OF SYSTEM QUALITY

System Perform	BQ	CQ	LQ
Speed	18	2	0
Availability / reliability	14	5	1
Responses	BQ	CQ	LQ
Completeness	13	4	3
Accuracy	12	4	3
Knowledge-Base Content	BQ	CQ	LQ
Provenance / Source quality	19	1	0
Updatedness	18	2	0
Expectations	BQ	CQ	LQ
Natural and Intelligent	19	1	0
Multimedia	20	0	0
Percent	83.1%	11.9%	5%

didn't know the theory of it. The conversation system quality was presented in Table VII and the number of BR, CR and LR is the number of students who judged the quality of system in each data.

D. Discussion

1) Integrated Knowledge-Base Construction: In the integrated knowledge-base construction part we create a collection in disaster domain.We think everyone can create different domain collections using our approach. Therefore, we have left both the variable value and threshold can be tunable. In the local topics identification (Section IV.B) we defined a threshold value named η. To determine η we used 2000 pair similarity values of disaster sentences. And in the an improved SumBasic section, we used a combination value of α and β. Fig.3. shows results for different η. Fig.4. shows $Vpercent$ and $Hpercent$ results for different α and β combination value.

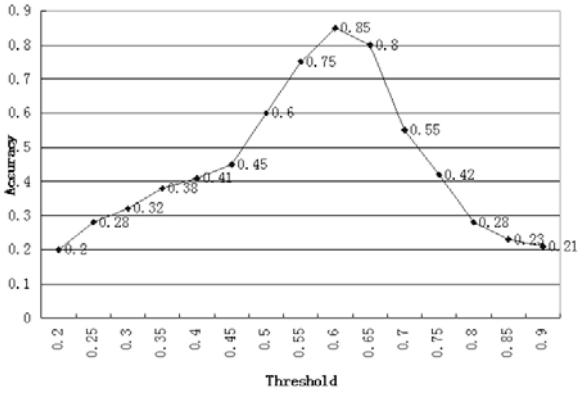

Fig. 3. The local topics identification results for different η.

The $Vpercent$ and $Hpercent$ results of text summarization experiment show that the information coverage ratio of our system is 38.9%, higher than compression ratios 20%, and 9.3% higher than Top-N method. The coverage ratio of high frequency word is 28.8%, higher than compression ratios 20%, and 12.7% higher than Top-N method. The result showed that our summarization approach can get better result than general approach. Therefore, we think the summarization approach can improve the classification result in the next step.

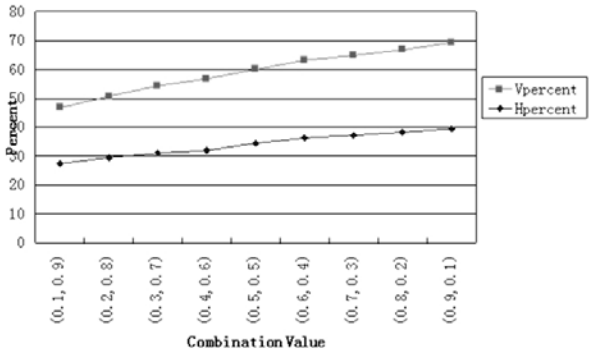

Fig. 4. The Vpercent and Hpercent results for different α and β.

For Article classification, the training set and testing set were constructed to verify the effect of this System. A 5 days news data was used to evaluate the experimental results. It can avoid using the occasional experiment data. As can be seen from the Table III, we can gladly see the highest F1-measures of our system is about 89.5% can be achieved, the average F1-measures is about 86.6%, 23.9% higher than TFIDF method, and 12.2% higher than general classification method. The proposed algorithm greatly outperformed the TFIDF algorithms. The result showed that this special domain news collections system could achieve better results in practice.

2) Similarity Computing: In similarity computing part, we proposed a approach based on the SRL and the Okapi BM25 measure. The Okapi BM25 measure is an effectual similarity computing measure based on word feature, but in this system we want to improve the conventional measure to the semantics level, therefore I used the SRL. To evaluate the effect of SRL, we did a compare experimentation based on Okapi BM25 measure and SRL & Okapi BM25 measure. We prepared 50 test data for each category and total 200 test data. The right numbers of result through the similarity computing were presented in Table VIII.

TABLE VIII
THE RESULTS OF SIMILARITY COMPUTING BASED ON OKAPI BM25 MEASURE AND SRL & OKAPI BM25 MEASURE

	Self introduction	Lab introduction	Phrase	Disaster
Okapi BM25	45	41	46	36
SRL & Okapi BM25	45	43	46	43

From table 8 we can find that the result of Self-introduction and Phrase based on Okapi BM25 is same as the results based on SRL & Okapi BM25, we think the reason is that the Self-introduction and Phrase data are too short and simple, they did not included some information about the semantics. The result of Disaster based on SRL & Okapi BM25 is 86%, 14% higher than Okapi BM25 measure, and we think the reason is that the Disaster data is complex and included some semantic information.

3) System Quality: The table 7 shows that the acceptable quality percent of our conversation system is 95.0% (BQ+CQ) and we think the quality experimentation had a satisfying

result. For speed quality, the system can achieve a response in 2 second, though we used some kinds of technique and lots of multimedia data. We got a very high appraisal value for natural and intelligent, and multimedia of expectations, we think the reason is that we used lots of novel technique and the user approved of these technique too. We get a very well result for knowledge-base content also because we used a special domain knowledge collections module.

VIII. Conclusions and Future Work

This paper outlines an algorithm about an integrated natural language processing approach for a more natural and intelligent conversation system. The three primary contributions of this work are an integrated framework for knowledge-base construction, a conversation understanding and generator approach, and a state manager method. We give the results and discuss the effectiveness of each part based on some experiment. We hope our proposed approach can make a more natural and intelligent conversation system, and the results show that this method could achieve better results in practice. We think the work has potential in this field and can help design and develop a real conversation System in the future.

In recently years, the face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. In the future we hope that use the face recognition technique to drive the development more effective and intelligent conversation system.

Acknowledgments

This research has been partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B), 19300029.

References

[1] Will, Thomas, Creating a Dynamic Speech Dialogue. Vdm Verlag Dr. Mller, 2007.
[2] R. J. Lempert, S. W. Popper, and S. C. Bankes, Shaping the next one hundred years: new methods for quantitative, long-term policy analysis. Santa Monica, CA.: RAND, 2003.
[3] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, "Towards conversational human-computer interaction," AI Magazine, vol. 22, 2001.
[4] Ong Sing Goh, Arnold Depickere, Chun Che Fung and Kok Wai Wong, "A Multilevel Natural Language Query Approach for Conversational Agent Systems", IAENG International Journal of Computer Science, vol.33-1, 2007.
[5] Teruhisa Misu and Tatsuya Kawahara, Speech-based Interactive Information Guidance System using Question-Answering and Information Recommendation, the 2007 International Conference on Acoustics, Speech and Signal Processing,vol.10.
[6] Hidekazu Kubota, Ken Saitoh, Ken Kumagai, Yohei Kawaguchi, Satoshi Nomura , Yasuyuki Sumi and Toyoaki Nishida, Conversation quantisation for conversational knowledge process, Inderscience Publishers, Volume 3 Issue 2, pp.134-144, 2007.
[7] L. Devillers et al., Annotations for Dynamic Diagnosis of the Dialog State, LREC'02.
[8] T. Paek & E. Horvitz. Conversation as action under uncertainty. Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI),pp.455-464, 2000.
[9] F. Benamara and P. Saint-Dizier, "Advanced Relaxation for Cooperative Question Answering," in.New Directions in Question Answering: MIT Press, 2004.
[10] H. Chung, K. Han, H. Rim, S. Kim, J. Lee, Y. Song, and D.Yoon, "A Practical QA System in Restricted Domains," presented at the ACL Workshop on Question Answering in Restricted Domains, 2004.
[11] F. Benamara, "Cooperative Question Answering in Restricted Domains: the WEBCOOP Experiment," presented at the ACL Workshop on Question Answering in Restricted Domains, 2004.
[12] William H. Fletcher, "Facilitating the compilation and dissemination of ad-hoc web corpora", in Papers from the Fifth International Conference on Teaching and Language Corpora, 2004.
[13] M. Baroni and S. Bernardini, "Bootcat: Bootstrapping corpora and terms from the web", in Proceedings of LREC 2004.
[14] David B. Bracewell, Fuji Ren and Shingo Kuroiwa, Mining News Sites to Create Special Domain News Collections, INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE VOLUME 4, NUMBER 1, pp.56-63, 2007.
[15] Lucy Vanderwendeet et al. Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. Information Processing and Management 43, pp.1606-1618, 2007.
[16] Harman, D. K. Overview of the fourth text retrieval conference (TREC-4). In D. K. Harman (Ed.), Proceedings of the fourth text retrieval conference. NIST Special Publication 500-236, pp. 1-24.
[17] Zhi Teng, Ye Liu and Fuji Ren, A Multimedia Conversation System with Application in Supervised Learning Methods and Ranking Function, International Journal of Innovative Computing, Information and Control, Volume 4, Number 6, pp.107-119, 2008.
[18] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001.
[19] B. Billerbeck and J. Zobel, Techniques for Efficient Query Expansion.
[20] Nianwen Xue and Martha Palmer. Annotating the Propositions in the Penn Chinese Treebank. In The Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing, pp. 47-54,Japan, 2003.
[21] Nianwen Xue and Martha Palmer. Calibrating features for semantic role labeling. In Pro ceedings of 2004 Conference on Empirical Methods in Natural Language Processing, pp.88-94,Spain, 2004.
[22] H. Zhang, H. Yu, D. Xiong and Q. Liu, HHMM-based chinese lexical analyzer ICTCLAS, Proc. Of the 2nd SigHan Workshop, pp.184-187, 2003.
[23] Qin Bing, Liu Ting Chen Shang-Lin and Li Sheng. "Sentences Optimum Selection for Multi-document Summarization", Journal of Computer Research and Development, Vol.43, No.6 2006-06-01, pp.1129-1134, 2006.
[24] A. Diekema, O. Yilmazel, and E. Liddy., "Evaluation of Restricted Domain Question-Answering Systems," presented at the ACL Workshop on Question Answering in Restricted Domains, 2004.
[25] Zhi Teng, Fuji Ren and Shingo Kuroiwa, "Emotion Recognition from Text based on the Rough Set Theory and the Support Vector Machines", 2007 IEEE International Conference on NLP-KE, ISBN: 978-1-4244-1610-3 pp.36-41, Beijing china, 2007.

Zhi Teng received the M.S. degrees in Graduate School of Advanced Technology and Science, The University of Tokushima in 2007. He is currently a doctoral student in Tokushima University of Japan. His research interests include Conversation system, Information Retrieval and Affective Engineering.

Ye Liu received the B.S. degree in Dalian University of Technology, China, in 2003. Currently she is a master student in Tokushima University of Japan. Her research interests include Question Answering System, Nature Language Processing and Information Retrieval.

Fuji Ren received the Ph.D. degree in 1991 from Faculty of Engineering, Hokkaido University, Japan. He worked at CSK, Japan, where he was a chief researcher of NLP. From 1994 to 2000, he was an associate professor in the Faculty of Information Sciences. From 2001 he joined the faculty of engineering, the University of Tokushima as a professor. His research interests include Natural Language Processing, Machine Translation, Artificial Intelligence, Language Understanding and Communication.