On an Understanding System that Supports Human-Computer Dialogue

Samir Rustamov
Cybernetics Institute of ANAS, Baku, Azerbaijan
samir.rustamov@gmail.com

***Abstract*— The problem of understanding of information and its application in human-computer dialogue systems is investigated in the paper. The problem of understanding user intention in dialogue systems for concrete object is investigated. For solution of this problem is applied Hidden Markov Model and received satisfactory results. Mathematical algorithms and software of the system have been developed for automatically defining user intention in human-computer dialogue.**

Keywords*— *dialogue systems; speech understanding; hidden markov model; speech recognition

I. INTRODUCTION

The basic core of the human-computer dialogue systems is the understanding natural language by computer. *Natural-language understanding* the processing of utterances in human language in order to extract meaning and respond appropriately. The human-computer dialogue can be hold written and oral form.

Speech understanding – the processing of speech that involves the mapping of the acoustic signal, usually derived from some form of speech recognition system, to some form of abstract meaning of the speech. The systems have been defined as computer systems with which humans interact on a turn-by-turn basis are called *spoken dialogue systems* (SDS). The main purpose of a spoken dialogue system is to provide an interface between a user and a computer-based application.

The problem of understanding of information and its application in human-computer dialogue systems is investigated in the paper. A system performing the functions of the information center has been taken as a object of research. The basic functions of this center are acceptance of calls, carrying out dialogue with a user, identification of user's intention and connection to appropriate department according to user's intent. The most important component in spoken dialogue systems is the learning user intention for speech understanding. For the solution of this problem applied HMM and received satisfactory results.

II. THE MAIN MODULES OF THE SPOKEN DIALOGUE SYSTEM

The main modules of the spoken dialogue system are followings: speech recognition, speech understanding, dialogue manager and speech generation (fig 1) [1][9].

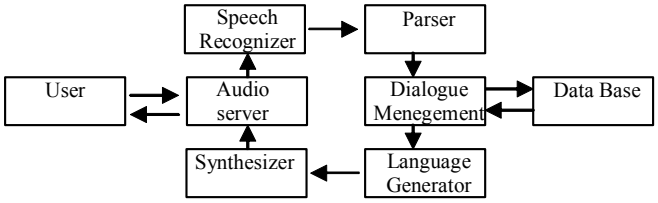


Figure 1. System architecture of the spoken dialogue system.

Speech Recognizer. The main purpose of this module is to convert user's speech corresponding to the sequence of the words in text. Apparently, the working quality of this module influences the correct working of all systems directly.

Speech recognition component can hold recognition by two principles: speech recognition by whole words and recognition by segmentation speech units (speech token). Experiments show that, recognition accuracy for whole words is considerably better than recognition accuracy for any sub-division of words. Research has tended to focus on word sub-divisions for their obvious storage advantages and less restrictive applications. If the application is small enough, however, real time systems can be built to perform useful tasks employing whole word token [2]. Most commercial recognisers fall into this category.

Speech recognition module has complex structure and solves the following problems:

Pre-processing. The pre-emphasing, endpoint detection, framing and windowing processes are carried out in this sub-module.

Speech feature extraction. The efficiency of this stage is one of the significant factors affecting behaviour of the next stages and exactness of speech recognition. Using the time function of the signal as feature is ineffective. The reason for this is that when the same person says the same word, its time function varies significantly.

To extract the impulse response of the vocal tract equivalent filter from speech signal are applied MFCC (Mel Frequency Cepstral Coefficients) and LPC (Linear Predictive Coding) algorithms.

A speech signal may be subjected to some channel noise when recorded, also referred to as the channel effect. A

problem arises if the channel effect when recording training data for a given person is different from the channel effect in later recordings when the person uses the system. The problem is that a false distance between the training data and newly recorded data is introduced due to the different channel effects. The channel effect is eliminated by subtracting the mel-cepstrum coefficients with the mean mel-cepstrum coefficients.

The LPC and MFCC cepstrals combined use in speech recognition system for calculating speech features [3].

Feature training and recognition. Different mathematical models are used for training and recognition of the features. For example: artificial neural networks (ANN), hidden Markov model (HMM), dynamic programming.

Using ANN's and HMM together for speech recognition can sometimes give better results than either of the techniques alone. In hybrid HMM/ANN models the neural networks are used to estimate posterior probabilities of classes given the input data[4].

We use Multilayer Artificial Neural Network for training and recognition processes. The neural networks of developed system were trained by conjugate gradient method[5].

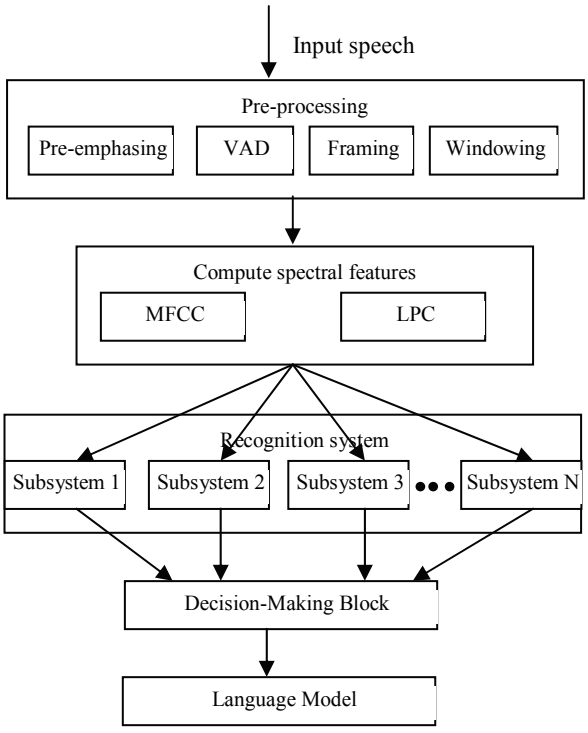


Figure 2. Structure scheme of the Recognizer Module.

One of main requirements in speech recognition system is reliability of recognition. To improve reliability of the system is offered combine using different structured or different features systems. This recognition systems can be work independently in one system and we called them conditionally subsystems of the main system (fig. 2).

Speech signal is trained by different mathematical models in each subsystem separately. The recognition results of the subsystems passed to decision making block in during recognition process.

The speech recognition system depending on the aim of a user presents him a recognition system of different quality [5].

Language model. Depending on grammar of language recognized words are corrected.

Parser. According to their meaning, the recognized words are parsed for speech understanding. There are applied standard CFG parsing algorithm and semantic HMM model for parsing operation.

Dialogue management. This module is considered as "thinking brain" of the dialogue systems. The decisions are made according to the features of the investigating object. So, before building spoken dialogue system about the flight, the calls to the company are investigated. User's standard questions are classified and the information about all possible cases is written to the knowledge base. The system sends information corresponding to user's messages during automatic dialogue and generates answers [6].

In order to complete required information system takes some strategies [7]:

Mixed initiative strategy. The system usually guides the user to give the required information for a specific intension.

Confirmation strategy. To make sure the information that system gets is corrects, the user must confirm the information he/she provided. Therefore, at the end of the dialogue, the system will make the final check of the information in the semantic slots.

Repair strategy. The user can correct the information he/she provided at any time.

Recovery strategy. If the user changes the content of the semantic slot, the system will update and response properly according to the new content in the semantic slot.

Language Generation. The parsed words are grammatically converted to sentence for delivering the information, determined by dialogue management module to the user.

Synthesizer. In this module text generalized by system is nearly synthesed to natural speech. Here the naturality of the synthesed speech is main factor So, for user correct understanding the system and avoiding irrelevant questions during dialogue, the methods serving for providing special intonation are applied[8].

III. The Application of HMM to Define User Intention in SDS

As an approach discrete HMM has been applied for defining user intention in a spoken dialogue system. The calls incoming to companies selected as a specific research object have been applied in the article. The main duty of the set system is the identification and automatically routing of calls incoming to the company on the basis of several sentences thereof throughout appropriate departments. Here the initial phase is the automatic speech to text converting process carried

out by a speech recognizer. One or few sentences from output of speech recognizer are taken as the user query.

The parameters of the HMM applied in speech understanding system introduced in the work are as follows [9,10].

1. N – is the number of states. As the states in understanding problem we have taken Word Semantic Sets (WSS). For building these sets dialogue examples are taken for learning problem in advance and the words taken part in these dialogues are included to the same WSS in terms of certain meaning. Note that the same polysemantic word can be included to different corresponding sets.
2. M – is the number of different words of dialogues taking part in the training process for the given problem.
3. V – is the all possible observations set, $V=\{v_1,...,v_M\}$. The elements of these sets in understanding problem are different words in the dialogues taking part in the training process.
4. $\pi=\{\pi_i\}_{i=1}^{N}$ – initial state distribution: $\pi_i=P(q_1=i)$.
5. $A=\lfloor a_{i,j}\rfloor$ – is the state transition probability distributions, $a_{ij}=P(q_{t+1}=j\,|\,q_t=i)$, $1\le i,j\le N$. A matrix in understanding process is the transition probabilities to corresponding WSS that words sequence are related to transition from one corresponding state to another.
6. $B=\{b_j(o_t)\}_{j=1}^{N}$ – is the probability functions of observation elements in states. Here for every j state, $b_j(o_t)=P(o_t\,|\,q_t=j)$ is the probability distribution of words taking part in WSS.
7. $O^{(r)}=\left[o_1^{(r)},o_2^{(r)},...,o_{T_r}^{(r)}\right]$ – observation sequences, $R-$ is the number of observed sequences, T_r – is the length of r-th observed sequence, $T_r\le T$, T– is the given quantity, $r=1,2,...,R$.

Note that HMM is briefly represented as $\lambda=(A,B,\pi)$.

The algorithm that we proposed for understanding of users queries in human-computer dialogue system with the application of HMM involves two phases: training and understanding process.

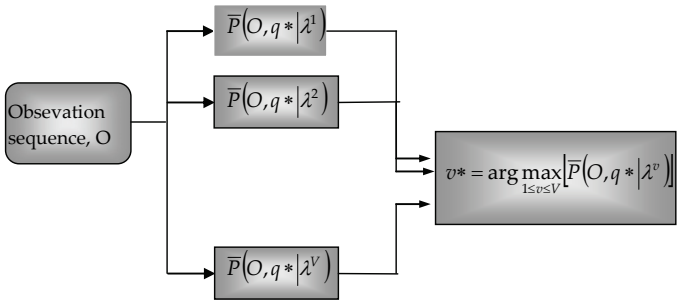
The following operations are carried out in the *training process*.

Collection of Word Semantic Sets. Words in the dialogues selected for understanding process are included to the same set from the point of view of having certain meaning and these sets are called Word Semantic Sets. WSS are taken as states in HMM. Note that, a similar polysemantic word can belong to different appropriate sets.

Evaluation of the Transition Matrix A. As the values of transition matrix A transition probabilities from one state to another corresponding to WSS belonging to the sequence of words in the sentence are calculated in the given problem.

Evaluation of probability distribution B. Distribution of user's query words to corresponding WSS have been taken as the probability distribution of observation elements in states.

Note that parameters of the HMM are estimated according to each corresponding department of the selected company. Probabilities found on the basis of parameters of HMM of all departments corresponding to each query are calculated by Viterbi algorithm[12]. The calculated probabilities are passed to a decision-making block. The probabilities of HMM are compared according to the departments in decision-making block (fig. 3). If the calculated maximum probability is less than a limit value found for appropriate problem as a result of experiments, computer rejects to receive calls or connects with human operator.



Learning user intention using HMM

IV. CONCLUSION

The understanding of a query from the initial question of calls incoming to information centre of an educational company and routing according to its intention was taken as the test problem to be solved.

Calls must be routed to one of the 4 departments of the company or connected to an operator or be rejected. These departments are: 1) information centre 2) accounting department 3) test exams centre 4) service departments.

180 queries have been taken for training process. Words contained in the user query in human-computer dialogue are taken as observation sequence in HMM. HMM have been built for every department of the company. The user queries are divided into the words and the parameters of HMM are estimated according to departments. The probabilities found on the basis of HMMs of all departments for each query are calculated, compared and result forwarded to a decision-making block.

For example: User calling to the Educational Center asks a question in this way: "Bazar günü keçirilmiş sınaq imtahanının cavabını öyrənmək istəyirəm" ("I want to know results of exam held on Sunday"). We have not provided the initial information and intermediate results in the article as they occupy more space in the calculation of HMM probabilities for this query according to departments. The calculated probabilities of the

HMMs matching the departments for the recorded query are as follows:

1. Information center: $P = 4 \cdot 10^{-7}$.
2. Accounting department: $P = 0$.
3. Test exams department: $P = 1{,}9 \cdot 10^{-4}$.
4. Service department: $P = 0$.

Apparently, maximum probability department is the "test exams department" the computer connects this query with the relevant department as this value is more than the experimentally defined limit value.

REFERENCES

[1] Aida-Zade K.R. Rustamov S.S., Mustafayev E.E. Principles of Construction of Speech Recognition System by the Example of Azerbaijan Language. International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Tutkey. 2009, pp. 378-382.

[2] G. Holmes. Natural Language Processing in Speech Understanding Systems, www.cs.waikato.ac.nz/pubs/wp/1992/uow-cs-wp-1992-06.pdf

[3] Aida-Zade. K.R., Ardil C., Rustamov S.S. Investigation of combined use of MFCC and LPC Fetures in Speech Recognition Systems. IJSP "International Journal of Signal Processing" Open Access Refereed Research Journal Volume 3:2006, ISSN 1304-4478.

[4] Veera Ala-Keturi. Speech Recognition Based on Artificial Neural Networks. Helsinki University of Technology. www.cis.hut.fi/Opinnot/T-61.6040/pellom-2004/project-reports/project_07.pdf

[5] S.S.Rustamov. On using an ambiguity of training neural networks in systems of speech recognition. (Azerbaijani). Transactions of Azerbaijan National Academy of sciences. "Informatics and control problems". Volume XXVI, №2. Baku, 2006, pp.256-260.

[6] Rustamov S.S., Mustafayev E.E. Construction Principles of the Speech Understanding Computer System. 24-th Mini EURO Conference "On Continuous Optimization and Information-Based Technologies in the Financial Sector" MECEurOPT 2010, Izmir, Turkey. 2010, pp. 294-299.

[7] Chung-Hsien Wu, Gwo-Lang Yan, and Chien –Liang Lin. Spoken Dialogue System Using Corpus-Based Hidden Markov Model. www.shlrc.mq.edu.au/proceedings/icslp98/PDF/AUTHOR/SL980219.PDF

[8] Rustamov S.S., Saadova A.V.On an Approach to Computer Synthesis of Azerbaijan speech. The second internation conferance "Problems of Cybernetics and İnformatics" dedicated to the 50th Anniversary of the ICT in Azerbaijan. Volume I. Baku, Azerbaijan. 2008, pp. 267-270.

[9] Aida-zade.K.R, Rustamov S.S., Baxishov U.Ch. The Application of Hidden Markov Model in Human-Computer Dialogue Understanding System. Transactions of ANAS. Series of physical-mathematical and technical sciences. Vol XXXII, No 3, pp. 37-46, Baku, 2012. (in Azerbaijani).

[10] Ch. Wu, G. Yan, and Ch. Lin. Spoken Dialogue System Using Corpus-Based Hidden Markov Model. 1998. The 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney Convention Centre, Sydney, Australia, ISCA. Volume 4, pp. 1239-1243.

[11] Jurafsky, Daniel, and James H. Martin. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 2nd edition. Prentice-Hall.

[12] M. Nilsson, First Order Hidden Markov Model - Theory and Implementation Issues, Tech. Rep., Department of Signal Processing, February 2005, ISSN: 1103-1581.

[13] Pieraccini, R.,Levin, E., and Lee, C.-H.(1991). Stochastic representation of conceptual structure in the ATIS task. In Proceeding DARPA Speech and Natural Language Workshop, Pacific Grove, CA, pp. 121-124.

[14] J.Haas, J. Hornegger, R. Huber, H. Niemann. Probabilistic Semantic Analysis of Speech. www5.informatik.uni-erlangen.de/Forschung/Publikationen/1997/Haas97-PSA.pdf

[15] H. Cuayáhuitl, S. Renals, O. Lemon, H. Shimodaira. Human-computer dialogue simulation using hidden markov models. In Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2005. pp. 290-295.

[16] B. H. Juang; L. R. Rabiner. Hidden Markov Models for Speech Recognition. Technometrics, Vol. 33, No. 3. (1991), pp. 251-272.