Application of Neuro-Fuzzy Model for Text and Speech Understanding Systems

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

***Abstract*— The problem of speech and text understanding and its application to the spoken dialogue systems is investigated in the paper. The Neuro-Fuzzy model has been applied for solution this problem and received satisfactory results. Mathematical model and software developed for building human-computer dialogue system.**

***Keywords*— dialogue systems; speech understanding; neuro-fuzzy model; learning user intention**

I. Introduction

A development of the Artificial Intelligence (AI) has stimulated a creation of the intellectual means for serving people in society. The main feature of the fifth generation computers is their having an artificial intelligence. Currently, serious and outstanding works are performed towards the development of artificial intelligence computer systems and their application to human activities. An information recognition and its understanding by computer stands in the centre of the AI. At first for creation human-computer inteaction must be solved the information recognition and understanding problems. Because information understanding by computer is very difficult problem generally, scientists apply this problem in specific fields. One of these fields is the different purpose human-computer dialogue systems. The first issue in dialogue systems is the understanding user intention. Determination of the user intention from the first sentence of the dialogue has wide application. At present, the most information centers of companies employ interactive voice response systems. User choose one item of menu which systems presented him (her) and push some button of a phone during call time. At offices and organizations carrying out wide-range of activities this menu is greater. The user applying to these companies is forced to listen to such menu every time and eventually it makes troubles and is exhausting for the user. Therefore, there is a necessity for systems making intellectual speech intercourse with a human replacing such systems.

Along with mathematical methods and algorithmic application it is necessary to take into account the peculiarities of natural language in development of these systems. As the modern dialogue systems have been designed and developed for languages such as English, German, Japanese etc. they don't support Azerbaijani.

The issue related to development of such systems based on strong mathematical models especially for Azerbaijani language researched in the article. Since, the initial problem in dialogue systems regarding the identification of which content does a user request corresponds to have been resolved. A specific object has been chosen and Neuro-Fuzzy models have been suggested for understanding process.

Note that different approaches have been suggested for solution of such kind of problems [1-8]. Dr. Chung and his colleagues applied HMM for learning user intention to hold dialogue in transaction systems [2]. They proved that correct identification of user intention seriously improves the dialogue system performance and its decision-making function. As another approach for identification of user intention has been used Hidden Topic Markov Models [7]. It should be noted that this mathematical method is the combination of Hidden Markov and Latent Dirichlet Allocation Models. Note that as another method theory of fuzzy sets is applied for understanding problem [1,8,9].

II. Problem Statement

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 main modules of the spoken dialogue system are followings: speech recognition, speech understanding, dialogue manager and speech generation [10].

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. Neuro-fuzzy models have been suggested for solution of this problem.

III. NEURO-FUZZY APPROACH FOR LEARNING USER INTENTION

Due to a complex structure of natural language, linguists use different type of "soft" technologies for its analysis. We have used fuzzy theory for solving this task, because there are fuzzy relations between words and their meanings [16].

The algorithm that we suggested for understanding of users query in human-computer dialogue by applying fuzzy set theory involves two phases: training and understanding process.

The following operations are carried out *in the training process*[1]:

1. The users queries are collected for training and classified according to their context;
2. $A = ((a_{ij}))$, ($i = 1..P, j = 1..N$) - term-document frequency matrix is constructed, where P is the number of salient terms, N is the number of destinations and an element a_{ij} represents the number of times the term i occured in calls to destination j. This number indicates the degree of association between term i and destination j.
3. In order to balance the contribution of each term, term-document frequency matrix is normalized so that each term vector is of unit lenght. Let B be the result of normalizing the term-document frequency matrix, whose elements are given as follows:
$$b_{ij} = \frac{a_{ij}}{\left(\sum_{e=1}^{N} a_{ie}^2\right)^{1/2}}, \; i = \overline{1,P}, j = \overline{1,N}.$$
4. The vector of inverse-document frequency (IDF) is calculated[17].
$$IDF_i = \log_2 \frac{N}{z_i}, \; i = \overline{1,P}.$$
where i is a term, n is the number of documents in the corpus, and z_i is the number of documents containing the term t. Using this weighting scheme, terms that occur in every document will be eliminated.
5. Membership function values μ_{ij}, for words containing user queries classified according to topics are defined.
$$\mu_{ij} = IDF_i \cdot b_{ij}, \; i = \overline{1,P}, j = \overline{1,N}.$$
6. Sentences selected according to topics are trained by neuro-fuzzy model. The aim function is defined as follows:
$$E(y) = \frac{1}{2}\sum_{i=1}^{P}\left(\frac{\sum_{j=1}^{N}\mu_{ij}y_j}{\sum_{j=1}^{N}\mu_{ij}} - d_i\right)^2 \to \min_{y \in R^N}. \quad (1)$$

The partial derivatives of this function calculate following form:

$$\frac{\partial E(y)}{\partial y_j} = \sum_{i=1}^{P}\frac{\mu_{ij}}{\sum_{j=1}^{N}\mu_{ij}}\left(\frac{\sum_{j=1}^{N}\mu_{ij}y_j}{\sum_{j=1}^{N}\mu_{ij}} - d_i\right). \quad (2)$$

Equation (1) is minimized by conjugate gradient methods and defined optimal values of y^*.

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

1. The user's query is divided into words and membership degrees defined in the training process of words are recalled from database. If the word is not found in the database then its membership degree is taken as zero.
2. These membership degrees are input elements of the neuro-fuzzy model. In next layer maximum values of membership degrees of each word according to content in the user query and the number of that content are defined:
$$\overline{\mu}_k(y_k) = \max_i \mu_k(y_i),$$
where $\mu_k(y_i)$ is membership degree of k-th word to i-th topic, $i = 1,...,N$, $k = 1,...,M$, M – is the number of words contained in the query.
3. The minimum of maximum values obtained for the same department according to each content is defined:
$$\overline{\overline{\mu}}_i = \min(\overline{\mu}_{k_1}, \overline{\mu}_{k_2}, ..., \overline{\mu}_{k_l}),$$
where $\overline{\mu}_{k_i}, i = 1,2,...,l$ is the maximum values for i-th topics.
4. The next block of neuro-fuzzy network consists of type of IF-THEN rules. The membership of user's query context given by means of Center of Gravity Defuzzification Method is defined in the final layer[16]:
$$\overline{\overline{y}} = \frac{\sum_{k=1}^{M}\overline{\overline{\mu}}_k \overline{y}_k}{\sum_{k=1}^{M}\overline{\overline{\mu}}_k} \quad (3)$$

$\overline{\overline{y}}$ quantity obtained in the result shows the context addressed by the user query (fig 1).

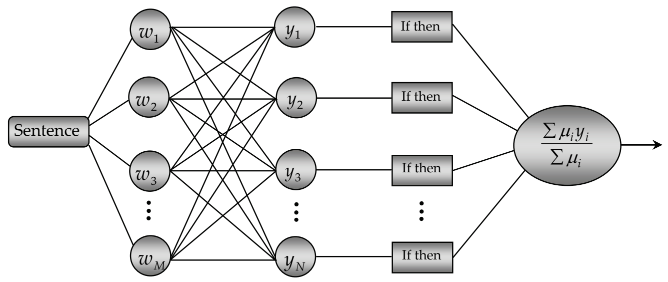

Figure 1: Learning user intention using neuro-fazzy model

IV. CONCLUSION

The understanding of a query from the initial question of calls incoming to information center 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 center 2) accounting department 3) test exams center 4) service departments.

180 queries have been taken for training process. During the application of neuro-fuzzy model the membership functions of queries have been calculated by dividing into words in a parallel way according to departments. Due to the restrictions put on the volume of the article we don't show the database of queries and the values of membership degrees for each query to fuzzy sets

After training of neuro-fuzzy sets we have the following values for $\bar{y}_i$, $i = 1,...,M$: $\bar{y}_1 \approx 1.028575$; $\bar{y}_2 \approx 1.772186$; $\bar{y}_3 \approx 3.000383$; $\bar{y}_4 \approx 4.000012$.

Incoming user's request is calculated by expression (3) in understanding process. If value of $\bar{\bar{y}}$ belong to the interval $(N-\Delta, N+\Delta)$, then system connects user to N-th department. Here $\Delta \in [0;0.5)$ is the main quantity, which influences to the reliability of the system. The value of Δ round the context addressed by the query increases meanwhile both understanding and misunderstanding rate enhances.

There is used kind of IF-THEN rules and according middle centre defuzzyfication method to determine the user intention for given appeal.

Let's apply neuro-fuzzi model to one user's request. 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"). Note that, the word "bazar" ("Sunday") doesn't belong to database.

1. $\bar{\mu}_1 = \max(\mu_1(y_1),...,\mu_1(y_4)) = \max(0;0;0;0) = 0;$
2. $\bar{\mu}_2 = \max(\mu_2(y_1),...,\mu_2(y_4)) = \max(0{,}97;\ 0;\ 0{,}24;\ 0) = \mu_2(y_1) = 0{,}97;$
3. $\bar{\mu}_3 = \max(\mu_3(y_1),...,\mu_3(y_4)) = \max(0{,}4;\ 0{,}07;\ 0{,}07;\ 0) = \mu_3(y_1) = 0{,}4;$
4. $\bar{\mu}_4 = \max(\mu_4(y_1),...,\mu_4(y_4)) = \max(0;\ 0;\ 2;\ 0) = \mu_4(y_3) = 2;$
5. $\bar{\mu}_5 = \max(\mu_5(y_1),...,\mu_5(y_4)) = \max(0;\ 0;\ 0;\ 0) = 0;$
6. $\bar{\mu}_6 = \max(\mu_6(y_1),...,\mu_6(y_4)) = \max(0;\ 0;\ 2;\ 0) = \mu_6(y_3) = 2;$
7. $\bar{\mu}_7 = \max(\mu_7(y_1),...,\mu_7(y_4)) = \max(0{,}95;\ 0;\ 0{,}32;\ 0) = \mu_7(y_1) = 0{,}95;$
8. $\bar{\mu}_8 = \max(\mu_8(y_1),...,\mu_8(y_4)) = \max(0;\ 0;\ 0;\ 0) = 0;$

$\bar{\bar{\mu}}_1 = 0$, $\bar{\bar{\mu}}_2 = 0$, $\bar{\bar{\mu}}_3 = \min(\bar{\mu}_4, \bar{\mu}_6) = 2$; $\bar{\bar{\mu}}_4 = 0$.

$$\bar{\bar{y}} = \sum_{i=1}^{4} \bar{\bar{\mu}}_i \bar{y}_i \Big/ \sum_{i=1}^{4} \bar{\bar{\mu}}_i \approx 3 .$$

It implies that computer system connects the user with the 3-rd department.

Table 1 shows that understanding indicators of the system which is given in percentage terms. It appears that as the value of Δ round the context addressed by the query increases meanwhile both understanding and misunderstanding rate enhances. Depending on destination of dialogues different strategies is chosen by the companies. The query address is defined by asking new questions by the system for the rejected queries. The results given in the Table 1, corresponds to understanding of intention from the first user question. Software has been developed for the intellectual computer system.

TABLE I. RESULT OF UNDERSTANDING OF USER INTENTION FROM FIRST SENTENCE OF USER'S REQUEST

Δ	Correct understanding	Misunderstanding	Rejection
0,2	66,3%	1,66%	32,04%
0,3	87,29%	8,84%	3,87%
0,4	88,95%	9,94%	1,56%

REFERENCES

[1] Aida-zade.K.R, Rustamov S.S., Ismayilov E.A., Aliyeva N.T. Using Fuzzy Set Theory for Understanding User's Intention in Human-Computer Dialogue Systems. Transactions of ANAS. Series of physical-mathematical and technical sciences. Vol XXXI, No 6, pp. 80-90, Baku, 2011. (in Azerbaijani).

[2] 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, p. 1239-1243.

[3] 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.

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

[5] W. Eckert, H. Niemann. Semantic Analysis in a Robust Spoken Dialog System. www5.informatik.uni-erlangen.de/Forschung/Publikationen/1994/Eckert94-SAI.pdf

[6] 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).

[7] Hamid R. Chinaei, Brahim Chaib-draa. Learning user intentions in spoken dialogue systems. ICAART 2009 - Proceedings of the International Conference on Agents and Artificial Intelligence, Porto, Portugal, January 19 - 21, 2009. INSTICC Press 2009, ISBN 978-989-8111-66-1. p. 107-114.

[8] Subasic P., Huettner A. Affect analysis of text using fuzzy semantic typing. Fuzzy Systems, Volume 9, Issue 4, pp. 483-496.

[9] V. Salvador, M. Andrade, A. Kawamoto. Fuzzy theory applied on the user modeling in speech interface. IADIS International Conference Interfaces and Human Computer Interaction. 2007. ISBN: 978-972-8924-39-3. www.iadis.net/dl/final_uploads/200710C029.pdf.

[10] Figure 2. SVM-Bc (Support Vector Machine algorithm is improved using cubic Bezier curves function).

[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] Aida-Zade. K.R., Ardil C., Rustəmov 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.

[13] 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.p. 378-382.

[14] R. Lopes-Cozar, A.J. Rubio, J.E. Diaz Verdejo, A.De la Torre. Evaluation of a Dialogue System Based on a Generic Model that Combines Robust Speech Understanding and Mixed-Initiative Control. www.lrec-conf.org/proceedings/lrec2000/pdf/101.pdf

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

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

[17] Kaufmann A., Gupta M. M. Introduction to Fuzzy Arithmetic Theory and Applications. N: Van Nostrand Reinhold, 1991. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=940962

[18] J. Chu-Carroll and B. Carpenter, Vector-based natural language call routing. Comput. Linguist., vol. 25, no. 3, pp. 361–388, 1999.