

Database Interaction Using Automatic Speech Recognition

S.Nareshkumar, N.Mariappan, K.Thirumoorthy

Dept of Computer science, Mepco Schlenk Engineering College, Sivakasi, India.

Dept of Computer science, SSN College of Engineering, College, Chennai, India.

Assistant Professor, Dept of Computer science, Mepco Schlenk Engineering College, Sivakasi, India.

Abstract—Interaction with standard databases are possible only if we know about the standard SQL queries. This paper focuses on interacting with the DBMS with speech. Automatic speech recognition is becoming famous now a days and it is widely used in many applications. Here users can interact with the database with their voice for retrieving details from it. Hence it is not necessary that user must have a prior knowledge about the SQL queries, they could retrieve details with their knowledge. The main purpose of this paper is that novice users who have no knowledge about the sql queries can use it for retrieving the details from the database.

I.INTRODUCTION

Speech is one among the way of exchanging information among people. Many interfaces are being developed for human-machine interaction.Voice/speech recognition is one of the widely developing area now.

Automatic speech recognition (ASR) is becoming famous nowadays. Many speech recognition systems can recognize lot of words. ASR, is has a lot of applications in many aspects of our daily life, for example, telephone applications, applications for the physically handicapped and illiterates and many others in the area of computer science.

II.RELATED WORK

Qiang Fu, Yong Zhao [1] proposed a system in which bayes decision theory is considered as the main approach for achieving the performance objective. Errors in automatic speech recognition is difficult to found,hence they have considered non-uniform classification error cost . Minimum classification error method is used for optimization here.They provided various approaches to test the effectiveness of their approach. Non-uniform

error criteria have been proposed by them for automatic speech recognition. They have designed a new method to support a consistent modeling and decoding process. TIDIGITS database was used to conduct experiments and non-uniform error cost matrix was created. The drawback in their approach is that the word error rate and sentence rate was more and they have not specified any criteria how to select a suitable error cost function for specific user requirements.

Q.Fu, D. Mansjur, and B.-H. Juang[2] proposed a system design for pattern recognition. This system involves non-uniform class dependent error cost issue. Inorder to maximize the rate of correct recognition, now adays there are many pattern recognition systems. Non-uniform error cost arises in many pattern recognition applications. Error is not minimized with the conventional techniques. Hence the proposed method has some steps to alleviate this limitation.This paper, mainly focuses on minimum error cost and uses the Bayes decision theory and finds a solution to solve the problem of classifier optimization with non-uniform error cost. Non-uniform error cost issue is raised since the uniform treatment of pattern recognition does not meets the expectation. Expected error cost reduction is a great issue to be considered now a days

Bayes decision theory is considered in this paper for applications with non-uniform error criteria . Inorder to implement the decision policy, they have addressed the two fundamental aspects in Bayes decision theory. Inorder to obtain the recognizer's parameters, they have followed a optimization procedure. Inorder to test how the system work for their designed system, they have tested it in Gaussian mixture classifiers. Bayes decision theory is used in this paper to solve the problems related to non-uniform error cost. They have verified their work using both computer-generated data samples and database . The drawback is that the non-uniform error cost learning criteria can gradually reduce the error cost

even when the model of the classifier is not exactly consistent with the true data distribution.

Song You, Huang Lu[3] proposed a specific program for different requirements, the program is very difficult to understand and difficult to use. Text processing is famous in the field of computer science and its applications. In order to overcome the complicated programs and the hardness in reusing the existing programs, here they have defined some rules to describe how text processing is to be done. Hence using this, they have designed an engine to parse the rules and execute it. So by this they simplified the text processing from programming to writing rules. In order to verify their implementation, they have done an experiment for extracting the web topic text.

Since the programming is strongly connected with logic in which we process, it is difficult to reuse the existing applications of text processing. Hence they have proposed a solution that initially rules are to be defined and then text processing is done by the appropriate design. They used two steps for achieving this, define the rules and then design the engine.

Regular expression is used to identify strings which we needed from the given text. Apache Lucene is a text search engine library which helps to solve the problems in specified areas of text processing and the drawback is that it is not applicable for the universal ones. These drawbacks are overcome in this paper by defining rules and designing engines. This paper compares the various text editors with their proposed methodology. The principle of rule definition and an engine is implemented and the various rules are tested in the engine. The methodology proposed in the system is highly reliable, reusable and also easy to implement for processing the text but it is not a reasonable solution for a problem.

Biing-Hwang Juang, Wu Chou, and Chin-Hui Lee[4] proposed a new Minimum Cost Error (MCE) approach based on learning for discrimination. The MCE approach aims at minimizing the error rate. They have elaborated the issues and solutions associated with the new MCE approach in this paper in the context of HMM-based recognizer designs.

In recent years of speech recognition, hidden Markov models are more prevalent in various applications. This paper addresses the problems in traditional speech recognition techniques in the context of revisiting the recognition problems. Statistical formulation of any data has the base in the Bayes decision theory. Lack of complete knowledge in training the data and the way they get distributed is considered as an issue here. The Hidden Markov Model plays a key role in characterizing the speech signals. Generalized Probabilistic Descent algorithm based on Hidden Markov Model[7] is discussed in detail in this paper. The proposed system provides 30-50% reduction error rate compared to the previous recognizer designs. The main difficulty in this new approach is with means to formulate the error rate estimate as a smooth loss function.

III. OVERVIEW OF PROPOSED SYSTEM

We propose an interactive database system approach in which speech is recognized using speech recognizer and the recognized voice is converted into text. Now the text is converted into a standard SQL query. For the constructed SQL query necessary details are retrieved from the database. This process is repeated until the user is satisfied with the retrieval results.

LIST OF MODULES:

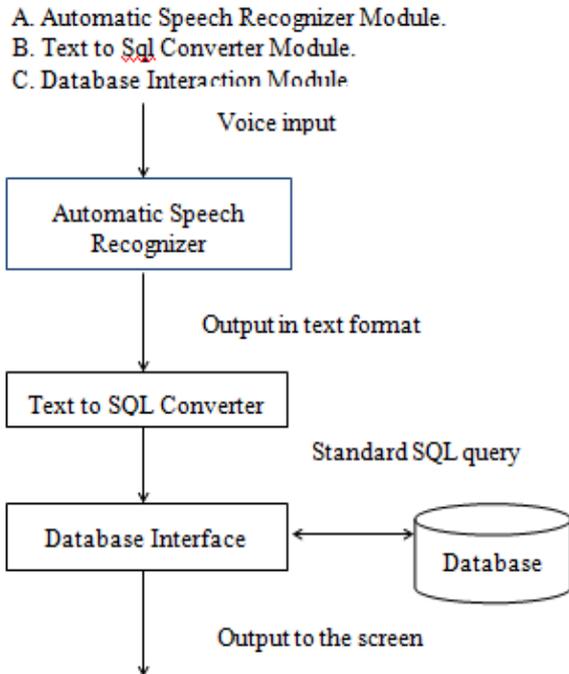


Fig. 1. Block diagram of the proposed system

A. Automatic Speech Recognizer Module:

The automatic speech recognizer takes user voice as the input and produces the corresponding text[3] as the output. It is a speaker independent continuous speech recognition system. It accepts voice input from the user as SQL or NON-SQL commands and converts into text. Sphinx4[5] is the tool which is used to implement this module. In order to implement this module, the voice of the user is taken as the input. This input is recorded using the microphone.

The voice of the users are stored as wav file. This wave file is taken as the input for processing. Here we need to consider the configuration file for processing the user voice input. The configuration file contains the required details for processing a wav file, the sampling rate[8] for processing the speech signal is stored in that file. Sphinx-4 is an HMM-based speech recognizer[7]. HMM stands for Hidden Markov Models, which is a type of statistical model.

1) Role of Sphinx4

Sphinx4[5] helps to recognize the speech of the user. It is used to recognize what speaker speaks and if there is any noise in their speech, it is discarded by this system. It gives the required string for further processing. In order to accomplish this task, we record the voice of user using microphone. The role of Sphinx4 is vital here. Sphinx4 has an in-built recognizer, decoder, feature extractor in it. It checks with the linguist for any rule violations[6].

Whenever an acoustic model is being created, the speech signals are initially transformed to sequence of vectors to represent the signal characteristics. Front end is used to generate these vectors which is present inside the recognizer. The features extracted from the speech signals are present in the acoustic model of Sphinx4. It has a dictionary[9] to store the words used in feature extraction. In order to represent the grammar used in speech recognition, we use language model. Microphone has the sampling rate of 16kHz and 16bit mono track. In this tool they use the feature extractors[6] such as windower and MFCC – Mel-frequency cepstral coefficients. The MFCCs are the amplitudes of the resulting spectrum.

B. Text to Sql Module:

Input recognized in the module1 is converted into a standard text format in this module. The administrator have to write an xml configuration file for the created database. Once the configuration file is written, it is taken as an input for processing module2. Here the xml configuration file is converted into a adjacency list like structure. Each line of the configuration file is processed and the required details are taken to produce the output for text to SQL converter.

Initially parsing[3] is done to separate each words in the sentences and now we get different set of tokens as the output. Now in this tokens we have to remove the stop words[3] such as a, the ,for, is. Once the stop words have been removed, now we have to categorize the tokens in to different labels.

- SELECT (select, display, view, show, get, ...)
- UPDATE(change, update, increase, decrease)
- DELETE(delete, remove, drop)
- WHERE(who, whose, where, has, having, have)
- MATH(average, minimum, maximum, sum, count, total, least,.....)
- RELATION(is, equal, greater than, less than,.....)
- TABLE(table names from database matrix)
- ATTRIBUTE(attribute names from database matrix)

TABLE 2: Relations used in the adjacency list

Relation	Representation
Attribute to Table	Tbl
Table to Attribute	Attr
Attribute to Attribute	Join

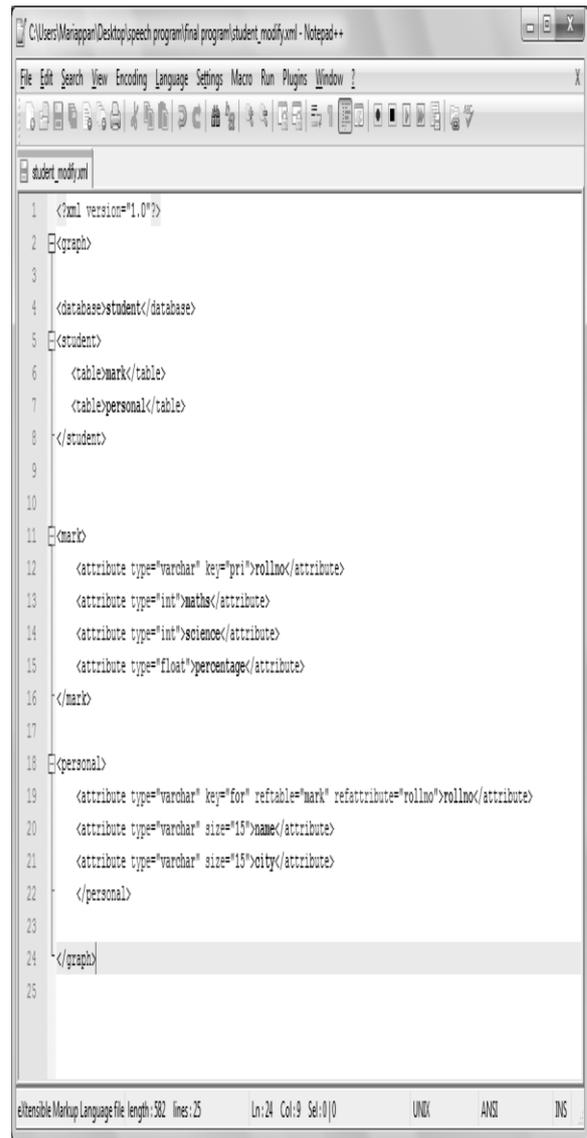


Figure 2: XML configuration file

i) XML configuration file:

The XML configuration file is a xml file. In this file, we store all the details of about our database. Name of the database is stored initially, followed by the tables created in the database. Then the attributes of the table created are stored. If there exists multiple tables, then we need to specify each table attributes separately. When the attributes are specified in the xml file, we need to mention their datatype also. Here we use the keyword “type” to denote the attribute’s datatype. The attribute which acts as the primary key must also be mentioned using the keyword “key”. If key=”pri” then it denotes that the attribute is the primary key of the table . If key =”for” then it denotes that the attribute is the foreign key of the table and it refers to some attribute in another table. Hence we need to mention the reference table name and the reference attribute in the xml file

The above table lists the various relations to be used in the construction of adjacency list structure If there is a any relation between two attributes ie. an attribute may act as primary key in one table and foreign key in another

table, hence we need to specify it by using the representation ‘join’ as listed in the above table.

Example:

Mark(rollno, maths, science, percentage)

Personal(rollno, name, city)

Let mark and personal be two tables and rollno acts as a key here. We can construct the adjacency list structure for the above relation and we obtain results such as

1)If the table name is given, attributes can be obtained.

2)If the attribute is given, corresponding table name can be obtained.

3)If two tables are given, join relation between them can be obtained.

The adjacency list constructed for the above example is given below. The table names are listed initially and then the attributes of the table are mentioned. Index starts from 0 here. Attributes index starts from 2 in our example since two tables occupy index 0 and 1 respectively. Relation specifies the type of relation i.e. the attributes of the table ‘mark’ are rollno, maths, science, percentage, we denote this by “attr” in the relation entry for mark table in the list. In the same way, the attributes of the mark table have their relation entry as “tbl” and the index value of the attribute can be obtained from ‘mark’ table. In our example, rollno is the primary key, hence it occurs in the both table. So we denote this by using the keyword ‘join’ in the list structure created.

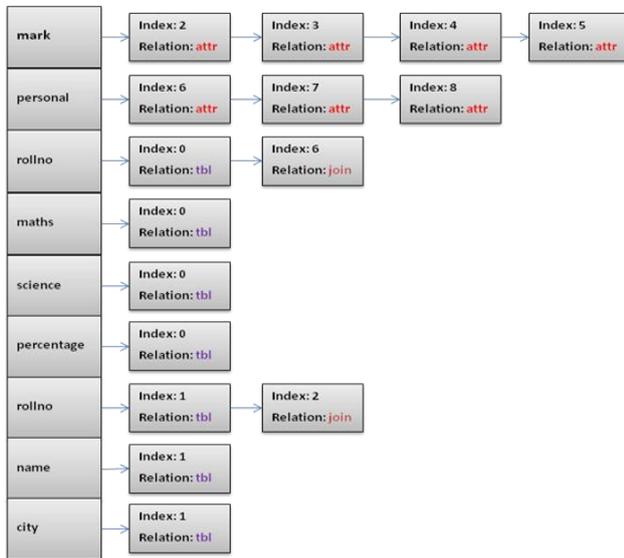


Fig. 3: Adjacency List like structure

ii) Sample patterns and conversion methods:

TABLE 3: Sample patterns and conversions methods .

Pattern	Conversion method
<select> <table>	select * from <tablename>
<select> <attribute>	select <attribute> from <tablename>
<select><table><where> <attribute><relation> <value>	select * from <tablename> where <attribute><relation><value>
<select><math> <attribute>	select <math>(<attribute>) from <tablename>
<select><tablelist>	<select> * from <tablelist> where <join conditions>

Same pattern for the queries are listed above in the table. From the sequence pattern of the tokens, the required query is generated.

Example:

1) select table (ex: employee)

Query is: select * from employee;

2)select attribute(ex:name) where(whose) attribute(id) relation(is) value(2001)

Query is : select name from employee where id=2001

C. Database Interaction Module

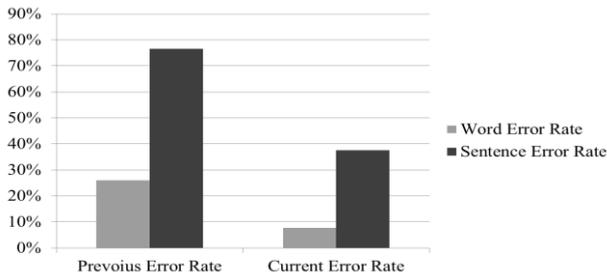
In this module, we have to consider the sql query obtained in the previous module as the input and it is used to retrieve the necessary details from the database. It is similar to normal database interaction, but here we are not using any keyboard to give the query, rather we just use the queries said by the user through their voice. Finally, the result is displayed into the screen.

IV EXPERIMENTAL RESULTS

A small database with 335 words and 1212 sentences finally. Trained the database with Word Error Rate=7.8% and Sentence Error Rate = 37.5%. The previous word and sentence error rate were 26.9% and 76% respectively [1]. Word Error Rate[4] is a common metric of the performance of a speech recognition. Word Error Rate is the sum of number of substitutions, number of deletions and the number of insertions divided by the number of words in the reference. Sentence Error Rate[10] is the sum of number of substitutions, number of deletions and the number of insertions divided by the number of sentences in the reference.

TABLE 4: Processing a developed system.

No. of speech sources	56 persons
Total no. of samples taken	1335
Total Samples Accepted	1212
Total no. of words used	335



V. CONCLUSION

Existing speech recognition systems are mostly command based recognition systems[2]. The proposed system gives a common framework for applications which involves database operation. It increases the interaction between user and computer. Users can be relieved from the complexities of the query. Sphinx4[5] toolkit have been studied for speech recognition. Converting the textual representation of queries into standard SQL format is done for some standard SQL queries.. The proposed system is being implemented to verify the given concept.

ACKNOWLEDGMENT

I wish to thank Dr.S. Arivazhagan M.E., Ph.D., Principal and the Management of Mepco Schlenk Engineering College, Sivakasi for extending generous help.

I am very much indebted to Dr. K. Muneeswaran M.E., Ph.D., HOD, CSE for his valuable guidance and helpful suggestions. I like to convey my sincere thanks to my guide Mr.K.Thirumoorthy, M.E.,Assistant Professor, CSE Department for his fruitful suggestions and complete support in my project work .

REFERENCES

- [1] Qiang Fu, Yong Zhao, Member, IEEE, and Biing-Hwang Juang, Fellow, IEEE "Automatic speech recognition based on non-uniform error criteria," *IEEE Trans. On audio, speech, and language processing*, vol. 20, no. 3, march 2012
- [2]Q. Fu, D. Mansjur, and B.-H. Juang, "Empirical system learning for statistical pattern recognition with non-uniform error criteria," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4621–4633, Sep. 2010.
- [3]Song you,Huang lu1,2,liang shixing "Research on rule definition and engine for general text processing"proceedings of 2009 4th international conference on computer science & education)
- [4] Biing-Hwang Juang, *Fellow, IEEE*, Wu Chou, *Member, IEEE*, and Chin-Hui Lee, *Fellow*" Minimum classification error rate methods for speech recognition,"*IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, may 1997.
- [5] Bayoumi, M.A., Lafayette, LA, USA ,Ramakrishna, N.A., Madraswala, T.,Sahu," SPHINX: a high level synthesis system for DSP design" *IEEE International Symposium on Circuits and Systems*, (Volume:1),1992.
- [6] Vojtko, J. Korosi, J., Rozinaj, G , "Comparison of automatic speech recognizer sphinx 3.6 and sphinx 4.0 for creating systems in slovak language systems" 15th international conference on signals and image processing, 2008.
- [7]Abushariah,A.A.M,Gunawan,T.S,Khalifa,O..O, Abushariah, M.A.M. "English digits speech recognition system based on hidden markov models" International conference on computer and communication engineering , 2010.

[8]Gada,Jigar,Rao,Preeti,Samudravijaya,"Confidence measures for detecting speech recognition errors" National Conference on Communications (NCC), 2013.

[9] Slobada, T. , Waibel, Alex "Dictionary learning for spontaneous speech recognition" Fourth International Conference on (Volume:4) Spoken Language, 1996. ICSLP 96. Proceedings.,

[10] Yousafzai, J. , Cvetkovic, Z. , Ager, M , Sollich, P "Redundancy in speech signals and robustness of automatic speech recognition" XIII International Symposium on Problems of Redundancy in Information and Control Systems (RED), 2012.