

Data Mining Techniques and Applications in Speech Processing - A Review

Authors

¹ K.A.Senthildevi, ² Dr.E.Chandra

Address for Correspondence:

¹ Gobi Arts and Science College, Gobichettipalayam, Tamil Nadu, India.

² Dr.SNS.Rajalakshmi College of Arts and Science, Coimbatore, Tamil Nadu, India.



Abstract: - Data mining can be defined as an activity that automatically extracts hidden knowledge from large set of data. Data mining research has emerged in the areas of speech, audio processing and spoken language dialog. The research has gained interest due to audio data that are available in plenty. In this paper, we have discussed about different kinds of mining with speech, voice and audio processing and tasks of data mining in speech processing. Also a brief summary about existing data mining applications with speech and audio processing is given.

Key Words: Data Mining, Speech Processing, Automatic Speech Recognition, Speech Data Mining, Pattern Discovery.

I. INTRODUCTION

Data mining techniques are concerned with discovering patterns and extracting useful information automatically from data. Data mining is involved in the fields of statistics, artificial intelligence, and machine learning [1]. As a result of the recent advances in data mining algorithms, data mining research has started to penetrate new grounds in areas of speech and audio processing as well as spoken language dialog [2].

Fundamental research in areas of prediction, classification, summarization, knowledge retrieval, learning and language understanding of speech and audio data are become important in

business processes [1], [3]. Automatic learning techniques will become an essential component of these new services. The research can be fueled by audio data that are becoming more widely available from a variety of multimedia sources including Web casts, conversations, music, meeting, voice messages, lectures, television and radio [1], [2]. Algorithmic advances in automatic speech recognition have also been a major, enabling technology behind the growth in data mining.

Effective techniques for mining speech, audio and voice data can impact numerous business and government applications. The technologies for discovering patterns from conversational audio, capturing useful trends, generating alarms are essential for law enforcement organizations and call center operation. Data mining is also an essential tool for searching through large volumes of audio warehouses to find information, document, and news [3].

II. SPEECH PROCESSING

Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop techniques and systems for speech input to machine [4]. Automatic speech recognition systems today find widespread application in tasks that require human machine interface.

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 [5]. A speech recognizer converts the observed acoustic signal into the corresponding [written] representation of the spoken [words].i.e., it converts spoken into text. Speech recognition is the ability of a machine or program to recognize and carry out voice commands or take dictation.

Voice or Speaker recognition is the process of recognizing the speaker that can simplify the task of translating speech [6]. Voice recognition is used to refer the recognition systems that must be trained to a particular speaker. The recognition process relies on features influenced by both the physical of an individual's vocal tract and the behavioral characteristics of the individual.

Speech synthesis is defined as a part of speech or voice recognition which is used to read aloud the contents of a display screen automatically to a blind user [5]. Speech synthesizers are text-to-speech systems used with computers. The synthesizer can be a card that is inserted into the computer, a box attached to the computer by a cable, or software.

III. DATA MINING METHODS OF SPEECH

Speech recognition is the system for conversion and recognition of spoken words to text without being

targeted at single speaker that can recognize arbitrary voices. Audio files contain speech and music contents. Data mining methods of speech can be classified into the following categories:

A. *Speech data mining*

Large amount of multimedia data, including speech are being recorded and stored on digital media. The existence of such large amounts of data has created a need for efficient and accurate data mining tools for extracting useful information content from the data. The typical need for mining is to search or browse through the data, scanning for specified topics, words, phrase, or speaker. Mining of multi speaker data collected from broadcasts, recorded meetings and telephone conversations are essential areas in research [7]. Mining heterogeneous spoken dialog data for the purpose of improving system operation and extracting business intelligence is a fertile growth area for new research initiatives at many research and industrial labs around the world.

B. *Voice data mining*

Voice data mining (VDM) is a multi-lingual voice processing system that is proficient in mining specific keywords from a large audio repository. It deals with the need to organize, search and retrieve collection of spoken documents such as recorded telephony conversations, TV or radio archives in an effective, efficient manner [8]. The technology can be very useful when handling with highly secured audio information and documents in major enterprises. By recognizing critical speech portions, it assists human operators in real time voice to improve operational efficiency by reducing the time, cost and effort. Generally, it automatically extracts portions of a speech or conversation that are of interest or importance.

C. *Audio mining*

Audio of common birds and pet animals have been recorded casually. Audio mining is a technique that is used to search audio files for occurrences of spoken words or phrases. Speech technology is used to recognize the words or phonemes that are spoken in an audio or video file and audio mining searches can then be carried out to locate specific words and phrases within the audio. There are two main approaches to audio mining[9].

1. Text-based indexing

Text-based indexing, also known as large-vocabulary continuous speech recognition (LVCSR), converts speech to text and then identifies words in a dictionary that can contain up to several hundred thousand entries. If a word or name is not in the dictionary, the LVCSR system will choose the most similar word it can find.

2. Phoneme-based indexing.

Phoneme-based indexing doesn't convert speech to text but instead works only with sounds. The system first analyzes and identifies sounds in a piece of audio content to create a phonetic-based index. It then uses a dictionary of several dozen phonemes to convert a user's search term to the correct phoneme string.

D. Video Mining

Video mining can be defined as the unsupervised discovery of patterns in audio-visual content. The motivation for such discovery comes from the success of data mining techniques in discovering non-obvious patterns. Furthermore, surveillance video often consists of events that are not known beforehand, and is hence an obvious target for unsupervised discovery of patterns, which in this case are events [10]. With video mining we would hope to discover the interesting events in the video without a priori knowledge of what those events are.

E. Conversation Mining

The conversations in a call center are analyzed to understand various issues being discussed, the intent of the caller, issue resolution and such things. To address the problems, speech recognition, Natural Language Understanding (NLU), machine learning, topic modeling and other techniques are used.

IV. DATA MINING TASKS IN SPEECH RECOGNITION

The tasks of data mining can be classified into some broad groups. The task groups are as follows: (i) prediction, (ii) classification, (iii) clustering, (iv) search & retrieval and (v) pattern discovery [11]. Speech recognition research is concerned with temporal data mining (sequential data) where audio records are ordered or indexed by time. The ordering is important and is central to the speech description or modeling. In this section, we provide a brief overview of data mining techniques which relevant to classification, clustering and search & retrieval of speech applications.

A. Classification.

Classification classifies data to pre-defined classes. The classification application builds a model from the trained classes and uses that model to classify new objects into one of the predefined classes automatically. It is one of the most widely used data mining methods. The main focus of data mining techniques is on classification performance with large amounts of data.

Speech recognition is one of the sequence classification applications. Time series matching and classification have received much attention in speech recognition research activity. The task of a speech recognition system is to transcribe speech signals into their corresponding textual representations. The classification task is preceded by feature extraction step. In speech recognition, the standard analysis

method is to divide the speech pattern into frames and apply a feature extraction method on each frame.

The feature extraction step in sequence recognition applications typically generates, for each pattern such as a speech utterance, a sequence of feature vectors that must then be subjected to a classification step. Sequence classification applications use pattern based as well as model-based methods [11]. Popular methods for sequence recognition are Hidden Markov Models (HMMs) and Dynamic Time Warping (DTW). Many other model-based methods have been explored for sequence classification.

B. Clustering.

Clustering provides an attractive mechanism to automatically find some structure in large data sets that would be otherwise difficult to summarize. Clustering of sequences or time series is concerned with grouping a collection of sequences based on their similarity. There are a variety of methods for clustering sequences. In case of sequence clustering, ARMA models or Hidden Markov Models are in popular use. The other broad class in sequence clustering uses pattern alignment-based scoring or similarity measure to compare sequences. In speech applications, a data mining system with clustering algorithm is used to find useful patterns from speech database.

C. Search and retrieval

Searching for sequences in large databases is an important task in data mining. Sequence search and retrieval techniques play an important role in interactive explorations. In content-based retrieval, the task is to search a large database of sequential data and retrieve from it sequences or subsequences similar to the given query sequence. In speech or audio applications, the individual elements of the sequences may be feature vectors of real numbers. When the sequential elements are feature vectors,

Euclidean distance may be used for measuring similarity between two elements.

In speech or audio signals, similar sounding patterns may give feature vectors that have large Euclidean distances and vice versa. An elaborate treatment of distortion measure for Speech and Audio signals can be found. In speech applications, Dynamic Time Warping (DTW) is a systematic and efficient method that identifies which correspondence among feature vectors of two sequences is best when scoring the similarity between them.

D. Periodicity detection

Periodicity detection has been a much researched problem in signal processing for many years. For example, there are many applications that require the detection and tracking of the principal harmonic in speech and other audio signals. Standard Fourier and autocorrelation analysis-based methods form the basis of most periodicity detection techniques that are currently in use in signal processing [11].

E. Pattern discovery

Pattern Discovery is a pioneer in data mining and predictive analytics. Unlike in search and retrieval applications, in pattern discovery there is no specific query in hand with which to search the database. However, one concept that is found very useful in data mining is that of frequent patterns. Data mining is concerned with formulating useful pattern structures and developing efficient algorithms for discovering all patterns which occur frequently in the data. In some applications, acoustic pattern discovery methods are used for automatically discovering words from speech using a combination of graph clustering, and base form searching.

V. EXISTING APPLICATIONS OF DATA MINING IN SPEECH PROCESSING.

Data mining researches in areas of prediction, classification, summarization, knowledge retrieval, learning and language understanding of speech and audio data are become important in revolutionizing business processes [1],[3]. Research of Data mining with Speech has involved in the following areas:

Essential Research Areas of Speech with Data mining

- Call center management
- Error detection in Dictation Speech recognition
- Speech Quality Measurement.
- Pattern discovery from speech database.
- Mining knowledge from Noisy data.
- Personalizing Therapy of Speech Disorders.
- Integration of Asynchronous knowledge sources.
- Voice document retrieval and indexing
- Multimedia and telephony applications
- Telephony conversions in English automatically
- Searching large audio/media archives
- Automate quality control aspects of the business.
- Captioning of TV and other video/media content.

Automatic speech recognition systems today find widespread application in tasks that require human machine interface, such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc. Speech recognition technology is increasingly used within telephone networks to automate as well as to enhance the operator services.

A. Information retrieval in noisy audio messages

Information can be extracted from large databases of audio messages by using speech data mining strategy. High Word-Error-Rate transcripts are often obtained on speech documents containing bad audio conditions. Call centers recordings and telephone survey corpora contain a large variety of speakers with bad audio quality due to cell phones and surrounding noises, unconstrained speech, variable utterance length and numerous disfluencies like hesitations, repetitions and corrections[12]. Extraction of business intelligence from call center recording or extraction of opinion from telephone surveys are very difficult tasks on such corpora. The potential applications of speech mining in this context is important.

The system can work on very noisy automatic transcriptions of spoken messages. There is a need to quantify the extent of similarity between any two (sub)sequences. An application was developed to extract the distribution of user's opinions from telephone surveys. The system works on very noisy automatic transcriptions of spoken messages [13]. Another one application was developed to the removal of continuous noise from old music records. The data mining process allows one to discern between signal and noise portions of the audio material, The masking threshold level can be determined for each data frame allowing one to make the noise inaudible. Data Mining for Detecting Errors in Dictation Speech Recognition.

B. Error detection in Dictation Speech recognition

Error detection is the precursor of error correction. Developing effective techniques for error detection can thus lead to improved error correction. The research on error detection is focused mainly on transcription and domain-specific speech. Data mining models for detecting errors in dictation speech recognition (DSR) was modeled. In this system, instead of relying on internal parameters from DSR systems, a loosely coupled approach to error detection based on features extracted from the DSR

output was proposed. Three data mining techniques, including Naïve Bayes, neural networks, and Support Vector Machines (SVMs), were evaluated on DSR corpora [13].

C. Data mining approach to Speech Quality Measurement

Mesuring speech quality by machines overcomes two major drawbacks of subjective listening tests, their low speed and high cost. Rear- time, accurate , and economical objective measurement of speech quality is used in a wide range of applications. In this paper, statistical data mining techniques to improve the accuracy of auditory – model based quality measurement is designed[14]. A large set of speech distortion features is created. Then multivariate adaptive regression splines(MARS) is used to find a set of features to provide best estimate of speech quality.

D. Model of a Data Mining System for Personalized Therapy of Speech Disorders

The children with speech disorder have more and more become object of specialist's attention and investment in speech disorder therapy are increasing. The development and use of information technology in order to assist and follow speech disorder therapy allowed researchers to collect a considerable volume of data. The application presents a data mining system designed to be associated with TERAPERS system in order to provide information based on which one could improve the process of personalized therapy of speech disorders [15].

E. Speech Emotion Recognition based on data mining technology

The speech emotion recognition has very important realistic values in enhancing the intelligence and humanity of computer, developing new human-machine environment and improving speech recognition results. The speech emotion recognition

includes speech signal preprocess, speech feature extraction and speech emotion recognition[16]. The goal of the system is to search the most useful features with analyzing the features related emotions and to find a recognition model to make use of these features.

F. Unsupervised Pattern Discovery in Speech

An approach to speech processing based on the principles of pattern discovery is designed. The system is to classify speech into categories defined by a prespecified inventory of lexical units. It is shown how pattern discovery can be used to automatically acquire lexical entities directly from an untranscribed audio stream[17]. By aggregating information about these matching patterns across audio stream, it is described how to group similar acoustic sequences together to form clusters corresponding to lexical entities.

G. Speech data mining for call center management

A speech data mining system is used in generating a rich transcription having utility in call center management. One of the systems is speech differentiation module which differentiating between speech of interacting speakers. Another is a speech recognition module improving automatic recognition of speech of one speaker based on interaction with another speaker employed as a reference speaker[19]. A transcript generation module generates a rich transcript based on recognized speech of the speakers. Mined speech data includes number of interaction turns, customer frustration phrases, operator politeness, interruptions, and contexts extracted from speech recognition results, such as topics, complaints, solutions, and resolutions. Mined speech data is useful in call center and product or service quality management.

H. Voice of Customer Analytics

Voice of Customer Analytics (VoCA) is developed to provide service in the area of CRM analytics. It provides unique capability to discover actionable business insights across various data sources ranging from calls to unstructured data to structured data. It encompassed advanced data mining algorithms such as data. linking, text clustering, text annotation, sentiment mining and predictive modeling, that allows analysts to come up with actionable insights regarding customer churn, first call resolution, and key customer satisfaction or dissatisfaction drivers. It is used to analyze a variety of heterogeneous data sources such as, agent logs, call records, CSAT survey verbatim, other enterprise logs, CRM records and so on.

I. Semantic Data Mining of Short Utterances

In this application, methodology for speech data mining along with the tools that the methodology requires is introduced. The tools increase the productivity of the analyst who seeks relationships among the contents of multiple utterances. It is shown how data mining techniques that are typically applied to text should be modified to enable an analyst to do effective semantic data mining on a large collection of short speech utterances [18].

J. Delay-coordinates embeddings as a data mining tool for denoising speech signals.

Delay-coordinates embeddings of sets of coefficients of the measured signal is used as a data mining tool to separate structures that are likely to be generated by signals belonging to some predetermined data set [20]. The embedding estimator in a windowed Fourier frame is designed and applied it to speech signals heavily corrupted by white noise. These estimators have performed well for a variety of white noise processes and noise intensity levels.

Data mining techniques in speech recognition helps in the areas of prediction, search, explanation, learning, and language understanding. A new class of learning systems can be created that can infer knowledge automatically from data. Effective techniques for mining speech, audio, and dialog data can impact numerous business and government applications. These techniques are also very essentials for searching through large volumes of audio warehouses to find information, documents, and news. Thus data mining technology with speech is an advanced and essential research field.

VII. REFERENCES

- [1] Giampiero Salvi , "Mining speech sound", Machine Learning Methods for Automatic Speech Recognition and Analysis. Phd thesis.
- [2] "Speech Data Mining & Document Retrieval", publication of the IEEE signal processing. www.asru2007.org/submission/EDICS.txt.
- [3] Surendra Shetty, K.K Achar, "Audio Data Mining Using Multi-perceptron ANN" International Journal of Computer Science and Network Security, Vol. 8, No.10 October 2008.
- [4] M.A. Anusuya, S.K. Katti, "Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security, Vol. 6, No.3, 2009.
- [5] Santosh K. Gaikwad, Dr. Babasaheb Ambedkar Marathwada, Bharti W.Gawali, "A Review on Speech Recognition Technique", International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.
- [6] National Science and Technology Council(NSTC), "Speaker Recognition" last updated Aug 2006.

VI. CONCLUSION

- [7] Dai Sheng-Hui, Lin Gang-Yong, Zhou Hua-Qing ,” A Novel Method for Speech Data Mining”, Journal of Software, Vol. 6, No. 1, January 2011.
- [8] A star exploit technologies,” Voice Data Mining – Index & Retrieve Spoken Documents”.
- [9] Neal Leavitt,” Let’s Hear It for Audio Mining”, <http://www.leavcom.com/pdf/Audio.pdf>
- [10] Ajay Divakaran, Koji Miyahara, Kadir A. Peker, Regunathan Radhakrishnan,” Video Mining using Combinations of Unsupervised and Supervised Learning Techniques”, Merl – A Mitsubishi Electric Research Laboratory.
- [11] Srivatsn Laxman and P S Sastry ,” A survey of temporal Data Mining”, Sadhana Vol 2, April 2006.
- [12] Andrzej CZYZEWSKI ,” Mining Knowledge in Noisy Audio Data”, Proceeding in KDD-96.
- [13] Nathalie Camelin, Frederic Bechet, Geraldine Damnati, “ Speech Mining in Noisy Audit Message Corpus” Antwerp Belgium, August 27-31
- [14] Wei Zha , Wai – Yip Chan , “ A Data Mining Approach to Objective Speech Quality measurement”, ICASSP 2004.
- [15] Mirela Danubuanu, Stefan Gheorghe Pentiu ,”Model of a Data mining System for Personalized Therapy of Speech Disorders”.
- [16] Ying Shi; Weihua Song, ”Speech emotion recognition based on data mining technology”, Sixth International Conference on Natural Computation- Aug 2010.
- [17] Park, A.S. Glass, J.R.,” Unsupervised Pattern Discovery in Speech”, IEEE transaction on Audio, Speech, and Language processing- Jan. 2008
- [18] Lee Begeja, Harris Drucker, David Gibbon1, Patrick Haffner, Zhu Liu, Bernard Renger, Behzad Shahraray, “Semantic Data Mining of Short Utterances”, IEEE Trans. On Speech & Audio Proc.: Special Issue on Data Mining of Speech, Audio and Dialog July 1, 2005.
- [19] Marcin Paprzycki, Ajith Abraham, Ruiyuan Guo and Srinivas Mukkamala,” Data Mining Approach for Analyzing Call Center Performance”, srinivas@cs.nmt.edu
- [20] Napoletani D.Struppa DC,Sauer T, Berenstein CA, Walnut D, “ Delay-coordinates embeddings as a data mining tool for denoising speech signals”.