Speech Therapy Models For Disabled Children Using Data Mining Techniques- A Review

Usha. M
Assistant Professor & Head,
Department of Computer Applications,
K G College of Arts & Science,

Dr. L. Sankari

Associate Professor (SG),
Department of Computer Science,
Sri Ramakrishna College for Women,
Coimbatore-641035
TamilNadu

ABSTRACT

One of the key challenges of the society development is correlated to public health and one of its specific targets includes improved treatments of diseases. It is true that there are affections which by their natural history do not endanger the life of a person, but they may have negative implications during his/her lifetime. Various language or speech disorders are element of this category.

Discovered and treated in time, they can be corrected, most often in childhood. This paper deals with the comparison of various data mining techniques which gives better suggestion in improving the speech with disordered children. Here various techniques like CRISP-DM, LOGO-DM and ASUM DM methods of speech therapy are compared.

*Keywords:-*Speech Therapy, Data mining, disordered children, CRISP-DM, LOGO-DM,ASUM-DM Models

I. INTRODUCTION

Language and speech impairments are affections which have negative collision on individuals' life standard even they do not endanger their life. These problems arise commonly to children under school age and, if they are revealed and proper treated they can be often corrected. During last years, the technological growth has a major influence on speech therapy. First, the communication among persons with various disorders was facilitated by computer or other communication tools. Second, the learning process was enhanced with help of various computer based speech therapy systems which increase the patients' curiosity and their attention in therapy by stimulating strong feelings and affective states [1].

Data mining is defined as the process of extracting appealing and previously unknown information from data, and it is broadly accepted to be a single phase in a complex process known as Knowledge Discovery in Databases (KDD) [2]. KDD consists of following phases:

- business understanding.
- data understanding
- data preparation
- modeling
- evaluation
- deployment

Speech therapy can be viewed as a business in logopaedic area that aims to suggest services for correcting language. A proper treatment of speech impairments ensures improved efficiency of therapy, so, in order to do that, a therapist must frequently learn how to adjust its therapy methods to patient's characteristics`

A first step is a composite examination, which starts with a personal and family anamnesis and continues with cognitive and personality examination. After that, a exhaustive speech investigation is made.

This can reveals the existence of general speech problems concerning hearing, voice or vocabulary, articulation problems as poor coordinating or pronunciation problems.

As an end, complex examination provides data about social, cognitive and affective parameters and point out potential development problems allowing a diagnosis for the patient.
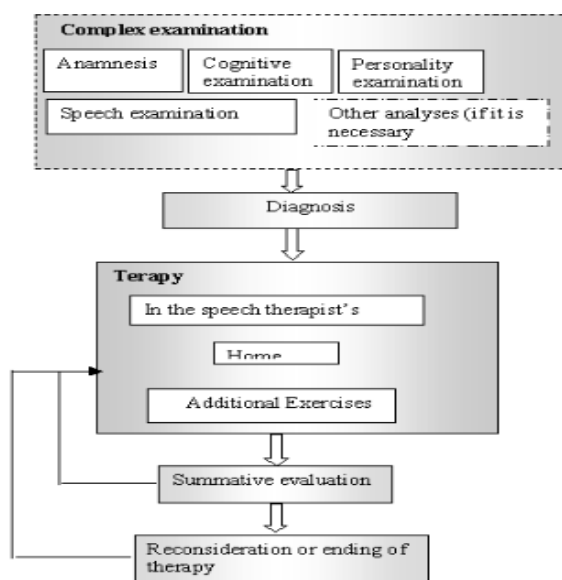
Figure 1 Key issues in speech therapy

II. DATA MINING TECHNIQUES IN SPEECH THERAPY

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.

Speech recognition research is concerned with temporal data mining (sequential data) where audio records are ordered or indexed by time. The ordering is vital and is central to the speech description or modeling

2.1 Data mining – educational tool for speech therapists

Data mining aims to find new patterns in large amount of data. Used mainly in business, data mining has some applications in education, medicine and healthcare. The process of tracking and mining patient's data in order to improve therapists' knowledge for enhanced personalized speech therapy is a new idea.

Knowledge extraction techniques applied in speech therapy area can be considered as formative evaluation techniques which aim to evaluate therapeutically programs while they are still in development with the purpose of continually improving these programs.

Data mining can discover useful information that can be used in formative evaluation to assist therapists to establish a proper basis for decisions when they design or adapt a therapeutically process.

The application of data mining in speech disorder therapy is an iterative process. Mined knowledge should enter the loop of the system and steer and enhance the whole process.

While prediction is the strongest goal, knowledge discovery is the weaker approach and usually prior to prediction. For example, in a medical application for a disease recognition, which belongs to predictive data mining, we must extract the data base for a set of rules that describes the diagnosis knowledge. Then this knowledge is used for the prediction of the disease when a new patient comes in.

Data classification is a two step process. In the first step, a model is built relating a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes.

Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label attribute. In the context of classification, data tuples are also referred to as samples, examples, or objects. The data tuples analyzed to build the model collectively form the training data set. The individual tuples making up the training set are referred to as training samples and are randomly selected from the sample population.

Since the class label of each training sample is provided, this step is also known as supervised learning (i.e., the learning of the model is 'supervised' in that it is told to which class each training sample belongs). It contrasts with unsupervised learning (or clustering), in which the class labels of the training samples are not known, and the number or set of classes to be learned may not be known in advance.
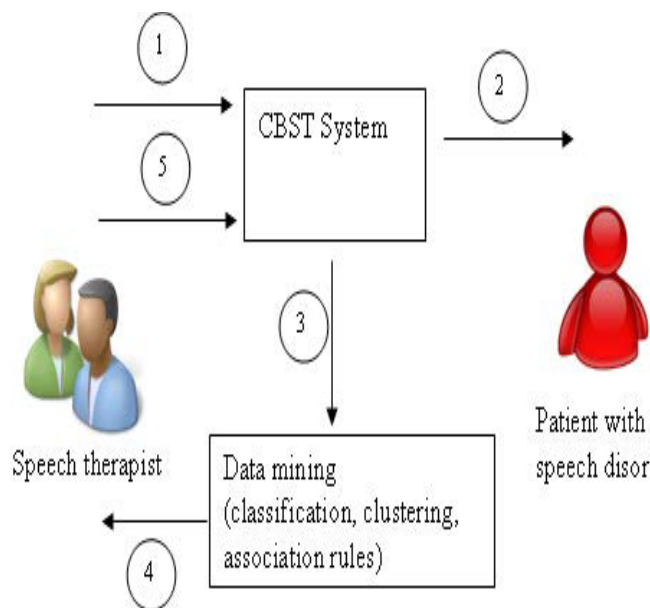
Figure 2: Speech therapy Process

III. COMPARISON BETWEEN CRISP-DM, LOGO-DM AND ASUM-DM METHODS

3.1. *Cross Industry Standard Process for Data Mining (CRISP-DM)*

Cross Industry Standard Process for Data Mining (CRISP-DM) commonly known by its acronym CRISP-DM, was a data mining process model that describes normally used approaches that data mining experts use to tackle problems.

CRISP-DM breaks the process of data mining into six major phases.

The succession of the phases is not strict and moving back and forth between different phases is always required. The arrows in the process diagram point to the most important and frequent dependencies between phases. The outer circle in the diagram symbolizes the cyclic nature of data mining itself. A data mining process continues after a solution has been deployed.

a) Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

b) Data Understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

c) Data Preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools.

d) Modeling

In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.

e) Evaluation

At this stage in the project you have built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

f) Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that is useful to the customer.

Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data scoring (e.g. segment allocation) or data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. Even if the analyst deploys the model it is important for the customer to understand up front the actions which will need to be carried out in order to actually make use of the created models.

IV. LOGO -DM - A DATA MINING SYSTEM FOR SPEECH THERAPY OPTIMIZATION

4.1 System Objectives

The sustainable development, in which special attention is given to all aspects of health care and the need to respond to the high efficiency requirements have led to the need for handling information "what is the predicted final condition for a child or what will be his/her state at the end of various stages of therapy, which are the best exercises for each case, and how patients can focus on their effort to effectively solve these exercises, or how the family receptivity - that is an important factor in the success of the therapy - is associated with other aspects of family and personal anamnesis"[5].

For all of these, the answer may be obtained by applying data mining techniques on data collected by TERAPERS. It is also interesting to try to enrich the knowledge base of expert system embedded in TERAPERS, with knowledge discovered in data mining process. In order to achieve these goals, we have proposed the development of Logo-DM system. Essentially, its objectives aim to perform an analysis of available data collected from children assisted by TERAPERS system and to prepare them in order to assure a proper quality for data mining algorithms, to try to select only those features that donate to the model building by removing those that are irrelevant or redundant, to choose the most appropriate methods and algorithms for data mining, to find models that can help to solve problems raised in speech disorders therapy, and to validate these models on new cases.

It is worth mentioning that the patterns, represented as rules, provided by Logo-DM, could, after some post- processing operations, be used to enrich the knowledge base of the embedded expert system in TERAPERS. Although, market claims many systems that allow data mining implementation, such as Weka and RapidMiner, their use implies IT skills. Our system is designed so that patterns can be easily discovered by speech therapists. They process a real dataset obtained by integrating data collected by all speech therapists that use TERAPERS. [6]

B. General Architecture

The graphical user interface allows the successive operations required by the knowledge discovery process. The preprocessing module prepares data for data mining algorithms and performs data transformation and feature selection for patterns building. These operations can be made both in centralized, distributed, or parallel ways. In order to achieve the proposed goals, the data mining kernel performs classification and association rules mining. Finally, the extracted models are evaluated by experts. If they meet the requirements of novelty and utility, they are considered knowledge.

V. ASUM-DM

In 2015, IBM Corporation released a new methodology called Analytics Solutions Unified Method for Data Mining/Predictive Analytics (also known as ASUM-DM) which refines and extends CRISP-DM.

IBM Analytics Services have released a new implementation method for Data Mining/Predictive Analytics projects called Analytics Solutions Unified Method for Data Mining/Predictive Analytics (also known as ASUM-DM) which is a refined and extended CRISP-DM.

VI. CONCLUSION

Already models available with CRISP-DM, LOGO-DM to treat the speech disabled children. The future study can take ASUM-DM for disabled children in order to improve their speech.

REFERENCES

[1] Castellani, J., Jeffs, T., Emerging reading and writing strategies using technology, Teaching Exceptional Children, 33(5), 60-67, 2007

[2] M. Danubianu, T. Socaciu Does   Data Mining Techniques Optimize   the

Personalized Therapy of Speech Disorders? Journal of Applied Computer Science and Mathematics, ISSN:1843-1046, 2009, pag 15-19

[3]   Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996)The KDD process pentru extracting useful knowledge from volumes of data. Communications of the ACM, 39(11):, pag.27-34, 1996

[4]   Wirth, R. and Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pages 29-39, Manchester, UK., 2000

[5]   Danubianu, S.G. Pentiuc, I. Tobolcea, and T. Socaciu, "Model of a data mining system for personalized therapy of speech disorder", Journal of Applied Computer Science, no. 6(3), 2009, pp. 28-32.

[6]   M. Danubianu, V. Popa, and I. Tobolcea, "Unsupervised Information-Based Feature Selection for Speech Therapy Optimization by Data Mining Techniques", Proceedings of ICCGI2012, June 2012, pp. 206-211.(LOGO DM)