Mathematical Treatment of Uncertainty in the Speech Recognition Process

Hesdras Oliveira Viana, Diogo Pereira Silva de Novais, Roque Mendes Prado Trindade

Abstract—One of the main difficulties in the speech recognition process is the treatment of the imprecisions around it. They have origin in the differences between the articulatory system of each person and the physical properties of the sound propagation. Moreover, the circuits involved in the sound storing and analysis works with a degree of uncertainty and also adds some imprecision in the process. This paper discuss the applications of mathematical methods used to treat uncertainty in the computational speech recognition and the possibility of using interval analysis for this propose.

Keywords—Speech Recognition, Mathematical Treatment, Uncertainty, Fuzzy, Hidden Markov Model, Interval Analysis.

I. Introduction

ACCORDING to Ladefoged [1], one of the main difficulties in the acoustic analysis of speech is the lack of possibility of analyzing the original sound. It occurs because when the sound is stored through analogical or digital devices what is analyzed is not the produced sound, but the captured one instead.

Even if the sound is captured in a soundproof booth it brings with it a series of uncertainty because of the characteristics of the human speech natures and the intrinsic imprecisions of the circuits used in the process. Moreover, the conversion of analogical audio into digital audio for computational analysis involves two discretizations, sampling, in time domain, and quantization, in the amplitude domain [2]. As well as any substitution of a infinite model for a finite one, the sampling and the quantization produce errors [3] which reflect directly in the accuracy of the speech recognition results.

The mathematics provides some resources to treat the uncertainty and the lack of accuracy in the modeling of a event. Neural Networks, Hidden Markov Models, Fuzzy Logic and Probabilistic Models, such as Bayesian Networks, are largely used for the decision-making with uncertainty. Interval Arithmetics are used to improve the accuracy of numeric calculations but are used also in the representation of uncertainty.

This paper presents and discuss Fuzzy Sets, Hidden Markov Models and Neural Networks characteristics and applications in Speech Recognition and discuss about the possibility of using of interval analysis in a similar way of these other methods in the computational speech recognition.

Hesdras Oliveira Viana is getting bachelor's degree in Computer Science in the Universidade Estadual do Sudoeste da Bahia

Diogo Pereira Silva de Novais is getting bachelor's degree in Computer Science in the Universidade Estadual do Sudoeste da Bahia

Roque Mendes Prado Trindade is with the Department of Exact Sciences in the Universidade Estadual do Sudoeste da Bahia

II. Fuzzy Sets

The classical set theory can represent efficiently problems in which is always possible determine if a given element is or not is a member of a set according to it's properties. The problem is that some problems doesn't provides informations to defines clearly the boundaries of a set according to the characteristics of it's members. That's difficult, for example, define a boundary between the warm and the hot or between good and bad.

In this context, the Fuzzy Sets Theory was created to modeling more efficiently problems inherently uncertain, where is very difficult or even impossible to determine with completely security if an element is or isn't a member of a set.

According to Klir [4], a crisp (or a classical) set A can be defined as:

$$A = \{x | P(x)\}, \tag{1}$$

where $P(x)$ evaluates if x is a member of the set A. Alternatively, sets can be represented by a characteristic function X_A defined as:

$$X_A = \begin{cases} 0 & \text{if} \quad x \in A \\ 1 & \text{if} \quad x \ni A \end{cases} \tag{2}$$

The function X_A in the crisp sets relates members of A with members of $\{0, 1\}$ and can be formally expressed by:

$$X_A : X \to \{0, 1\}. \tag{3}$$

Whereas in the fuzzy sets, the equivalent function X_{fA} can be expressed by:

$$X_{fA} : X \to [0, 1], \tag{4}$$

where the value returned by the function X_{fA} is a number in the interval $[0, 1]$ and represents the degree in which x is a member of A.

The main difference between crisp sets and fuzzy sets is the cardinality of the codomain of the characteristic function. The fuzzy set has an infinite cardinality or the biggest cardinality supported by the system in the case of discrete systems, whereas the crisp set has cardinality equals to 2. This characteristic of the fuzzy sets make them interesting for problems involving uncertainty, in these cases, the degree in which an element is member of a set, represents the degree of correctness in which it meets some characteristic.

An example of application of fuzzy theory in the treatment of uncertainty in speech recognition is the master degree work of Mills [5], in which was done a isolated digits recognizer using fuzzy versions of some known algorithms used in speech recognition.

The same word spoken by different people has, between

ISBN: 978-960-474-252-3

other variations, variations in the loudness and in the velocity. To treat these variations, Mills [5] used a fuzzy version of the SDP (Symmetric Dynamic Programming) algorithm, to approximate two sounds in waveform in time domain. Once the SDP has a complexity $O(n^2)$, in Mills [5], was not done any analysis in frequency domain to avoid the excessive hardware consuming, because it would be necessary to use Fourrier Transform or LPC (Linear Predictive Coding) to analyse sounds in the frequency domain.

During the tests in Mills [5], the fuzzy version of the algorithms had obtained in the worst case the same efficiency of the crisp algorithms.

III. Hidden Markov Model

A Hidden Markov Model (HMM) is a stochastic model of state machine which has basically two states, the observable states and the hidden one's. Commonly, the hidden states are used to modeling physical characteristics of the problem [6]. The transitions between the observable states are predefined, whereas in the hidden states, they are defined through a probability matrix generated by some specific algorithm based on random events.

According to Rabbiner [7], an HMM can be defined by:

$$A = \left\{a_{ij} | a_{ij} = P(q^{n+1} = j | q^n = i) = P(q_j^{n+1} | q_j^n)\right\} \quad (5)$$

$$B = \{b_j(x_i) | b_j(x_i) = P(x_i | q = j) = P(x_i | q_i)\} \quad (6)$$

$$\pi = \left\{\pi_i | \pi_i = P(q^1 = i) = P(q_i^1)\right\}. \quad (7)$$

The expression (5) distributes the probability of transition from a state i to the state j, always having the same probability. In the expression (6), for each pair $P(xi|qi)$ there is a correspondent probability distribution of return in the case of a discrete HMM, or a probability distribution function for a continuous HMM. The term $b_j(x_i)$ refers to probality (in the discrete case) or verisimilitude (in the continuous case) of the simbol x_i generated by the state q_j. Finally, the expression (7) refers to probability distribution of the first state. In the left-right models, normally is assumed $\pi_1 = 1$ and $\pi_i = 0$ for each $i \neq 1$.

An important characteristic of the HMM's is the ability of model events dynamics in time domain. Thus, they can be implemented for pattern recognition in the speech signal as in the temporal analysis as in acoustic analysis or both [8].

Some important works in speech recognition with HMM are:

- Saadeq et al [9] proposes a new ASR (Automated Speech Recognition) applied to electromyographic and vibrocervigraphic and with HMM temporal modeling, to reduce the noise level;
- Kim et al [10] using HMM's to Korean digits recognition through the orthogonal mother Wavelets. They have obtained a recognition rate higher than the LPC descriptors and the MFCC (Mel-Frequency Cepstral Coefficient);
- Hwang [11] propose the shared distribution model to substitute generalized triphone models for independent ASR with HMM's. He have obtained the reducing of redundant states, getting a better recognition rate;
- Other works in speech recognition with HMM's are Tolba [12], Tan et al [13], Sarikaya et al [14], Zhu et al [15], Alcaim et al [16], Chan et al [17], Hosom [18] and Jiang et al [19].

ASR's with HMM's for isolated words recognition, have obtained an efficiency higher than 95% for some vocabularies with around one thousand words, for different speakers.

IV. Neural Networks

Many algorithms and other mathematical models are inspired in process and nature agents. A lot of abstract structures, for example, stacks, queues and trees were influenced for concrete entities. As this models, the neural networks were inspired in a nature structure, in this case, the organization and communication of human neurons.

A neural network is formed by a set of nodes or units, which are connected by links. Each link has a weight associated to it. Each unit has input links from other units and output links for other units. The main idea is the possibility of each unit works depending exclusively on it's inputs, without be necessary it have a global knowledge of the network [20].

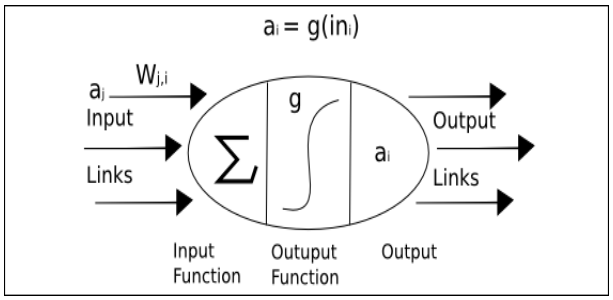

Figure 1. A unit of a neural network [20].

The figure 1 shows a unit of a neural network, where $w_{i,j}$ is the weight of the link for the unit i, a_j the input value of the link, $g(in_i)$ the input value applied to the activation function and a_i the output value of the unit i. A neural network, as soon as it's created, have it's weights adjusted in a random way and it's learning occurs through the adjustment of it's weights with a training algorithm. This algorithm works in the follow way:

- a input value is attributed for the network and the output value is observed;
- if the output value is right, nothing is done;
- if it's wrong, the error is calculated and the links which contributed for the decision have it's weights adjusted proportionally to the error.

This technique is known as error backpropagation. Neural networks are commonly used to pattern recognition in environments statics in time domain or in a located segment of time [21]. Because of this characteristics, it hardly can be used to make temporal analysis of speech signal [21]. Because this problem, neural networks are usually combined with other models, doing just the acoustic analysis of speech. In his work, Tebelskis [21], used a hybrid model, with neural networks doing acoustic analysis and HMM's doing temporal analysis.

Other application for speech recognition are proposing

the using of neural networks of the SVM (Support Vector Machine) type as classifiers.

In this context, some important works are:

- Daveport and Garudadri [22] used the database TIMIT for neural networks training. He have obtained a recognition rate of 84%;
- Yuk [23] have used speech recognizer in different environments and observed that neural networks works well with noise environments;
- Picone [24] have used a hybrid model with SVM and he have obtained a upper performance than just HMM as classifier;
- Juneja [25] observed a better efficient in SVM classifiers than HMM's ones for phonemes classification;
- Abdulla [26] proposed the use of SVM for phonemes separation and got a upper efficience than HMM in intra-word phonemes detection.
- Other importants works in speech recognition with neural networks and SVM are Abe and Inoue [27], Clarkson and Moreno [28], Mporas et al [29], Melin et al [30] and Lin and Lee[31].

The exposed works show that neural networks and SVM can work well in environments with noise, and presented goods results too in phoneme detection.

V. Interval Analysis

Between two real numbers there are infinite numbers. Using digital computers to solve problems involving real numbers is a very discussed problem, because digital computers are discrete and are not capable of representing this continuous space.

Programming Languages commonly use float point arithmetic to represent a subset of the real numbers in computers. These numbers are constructed, some times, through rounding or truncating of real numbers. Thus, using float point arithmetic, the obtained solution is a approximation of the real solution, but the size of the imprecision is not represented.

Interval analysis proposes the representation of a interval of real numbers as a new numeric type formed for two real numbers, one representing the lower bound and the other representing the upper bound [32].

A interval can be defined as a closed bounded set of real numbers, and can be represent as follows [32]:

$$[a, b] = \{x : a \leq x \leq b\} \tag{8}$$

The main advantage of using intervals is that the imprecision can be considered and represented during all the process, whereas in the float point notation, immediately later the rounding or truncating, the information about the imprecision is lost. However, it's importante to remember that computers can not represent all real intervals, once the bounds of the interval will be represented by a discrete number, usually, a float point number.

One of the advantage of intervals is the possibility of considering the uncertainty during the calculations. To make it possible, Moore [32] defined a specific arithmetic for intervals, where the usual operations are defined as:

$$X * Y = \{x * y : x \in X, y \in Y\}, * \in \{+, -, \times, \div\}. \tag{9}$$

Beyond the basic arithmetics, a series of other mathematical definitons already have a interval version, such as the Integral [32] and the Line Integral [33] and others used in digital signal processing.

Before discuss a interval version of mathematical "tools" for digital signal processing, it's important to define a digital interval signal. Based on the conceps of Oppenheim and Schafer [34], Chen [35] and Lyra [36], we deduced that a digital interval signal can be seen as a signal represented by a sequence of digital intervals, where the lenght of each term represents the degree of imprecision of the signal. In this scope, digital interval signals can represent the uncertainty of circuits and activities involved in the speech recognition process, the microphone and the discretizations, for example.

The discrete convolution is the most important operation of digital signal processing. It is a linear system. If a linear system is fully specified by impulse response then it satisfies all mathematical convolution conditions. It can also be used for average move filter implementation. The concept of convolution is strongly correlated with the concept of mobile average. The output of a linear system can be given by the convolution of input with the impulse response of the system. In statistics, the density function of the probability of the sum of two independent variables X and Y is given by the convolution of the respective probability density function. In the multiplication of polynomials, the coefficients of the product is given by the convolution of the coefficients of input polynomials [37].

Other mathematical "tools" already has a interval version, such as the fourier transform used to analyse a signal originaly in time domain in the frequency domain; the *Z*-transform [38], used to make analysis on the linear system used in the signal processing; and the K-means interval algorithm used in the separation of patterns [39].

Similarly to the work of Mills [5] with fuzzy logic, the substitution of classical methods for interval methods in the speech recognition can improve the results, once the uncertainty will be considered during the analysis.

Other important question related to the use of interval analysis in speech recognition is the concept of the distance between two intervals. We see the use of distances in several areas, such as cluster algorithms for automatic classification of data of high dimensionality in the work of Fu and Huang [40], and in problems of segmentation for audio by Sundaram and Narayanan [41]. The concepts of acoustic distance and phonemic distance are presented in Lin and Lee [31]. Although Moore proposed a metric to measure the distance between two intervals, it can not represent efficiently the uncertainty, once the distance of two intervals of uncertainty is a real number in his definition. Trindade [42] proposed and defined an strict interval metric, that generalizes the Euclidean distance, where the distance between two intervals is a interval. The proposed metric accomplishes the numeric aspect if one takes an interval as an approximation of a real number and the logic aspects if one takes an interval as a fuzzy information.

ISBN: 978-960-474-252-3

VI. Conclusion

This paper presented some of the main mathematical resources used to treat the uncertainty in the computational speech recognition, it's characteristics, applications and results.

Among the most widely used method for speech recognition, we see that the Hidden Markov Model has been extensively used in conjunction with neural networks, presenting a high rate of recognition, especially when done through the Support Vector Machine classifier.

The fuzzy logic begins to be applied to speech recognition. The use of fuzzy version of classical algorithms have permitted the representation of the uncertain nature of the speech recognition problem and have provided better results.

Another contribution of this papers is the discussion about use of interval analysis in the digital signal processing and the possibility of applications in speech recognition. It presented the existence of important tools signal processing and an strict interval metric that can represent the uncertain in the notion of distance of two intervals, and accomplish with the vision of intervals like a approximation of a real number and logically as a fuzzy information.

This paper can serves as basis for understanding and comparisons between the discussed models and their applications in speech recognition. As future works are experiments with interval analysis in speech recognition, other combinations of these methods, and the possible definition of a interval acoustic or phonetical distance.

Acknowledgment

The authors would like to thank the members of the research group in speech recognition formed by members of Universidade Estadual do Sudoeste da Bahia-UESB, Instituto Federal da Bahia-IFBA, Faculdade Independente do Nordeste-FAINOR and Faculdade de Tecnologia e Ciencias-FTC.

References

[1] LADEFOGED, Peter. *"Elements of acoustic phonetics"*. 2rd ed. The University of Chicago Press. Chicago. 1996.

[2] STRANNEBY, Dag. *"Digital Signal Processing: DSP and Applications"*. Oxford. 2001

[3] CLAUDIO, D.; MARINS, J. *"Calculo numerico computacional: teoria e pratica"*. 2rd ed. Editora Atlas. Sao Paulo. 1994.

[4] KLIR, G. J.; YUAN, Bo. *"Fuzzy sets and fuzzy logic: theory and applications"*. Prentice Hall. New Jersey. 1995

[5] MILLS, P. M. *"Fuzzy speech recognition"*. University of South Carolina. South Carolina. 1996.

[6] CHING, Wai-Ki; NG, Michael K.. *"Markov Chains: models, algorithms and applications"*. Springer. New York. 2006.

[7] RABINER, Lawrence R.; JUANG, Biing-Hwang. *"Fundamentals of Speech Recognition"*. New Jersey, USA. Ed. Prentice Hall, 1993.

[8] JUANG, B. H.; RABINER, L. R.. *"Hidden Markov models for speech recognition"*. Technometrics, Vol. 33. 1991.

[9] SAADEQ, R. M.; ALI Akbar Khazaei. *"A novel model characteristics for noise-robust Automatic Speech Recognition based on HMM"*. International Conference of the IEEE. Wireless Communication, Network and Information Security (WCNIS). 2010.

[10] KIM, Kidae; YOUN, Dae Hee; LEE Chulhee. *"Evaluation of wavelet filters for speech recognition"*. IEEE International Conference on Systems, Man, and Cybernetics. 2000.

[11] HWANG, M.; HUANG, X. *"Shared-Distribution Hidden Markov Models for Speech Recognition"*. Carnegie Mellon University. 1991.

[12] TOLBA, H.. *"Comparative Experiments to Evaluate the Use of Syllables for the Improvement of Automatic Recognition of Dysarthric Speech"* IEEE 16th International Conference on Systems, Signals and Image Processing. 2009

[13] TAN, B. T.; FU, Minyue; SPRAY, A.; DERMODY, Philipy. *"The use of wavelet transforms in phoneme recognition"*. Fourth International Conference, on Spoken Language, ICSLP 96. 1996.

[14] SARIKAYA, Ruhi.; GOWDY, John N. *"Subband based classification of speech under stress"*. Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '98. Vol. 01. 1998.

[15] ZHU Q.; ALWAN, A. *"On the use of variable frame rate analysis in speech recognition"*. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '00. 2000.

[16] ALCAIM, Abraham; SANTOS, Sidney. C. B. dos. *"Silabas como unidades foneticas para o reconhecimento de voz em portugues"*. SBA Controle & Automacao. 2001.

[17] CHAN, C. P.; CHING, P. C.; LEE, Tan. *"Noisy speech recognition using de-noised multiresolution analysis acoustic features"*. Journal Acoustical Society of America. 2001.

[18] HOSOM, John P. *"Automatic Phoneme Alignment Based on Acoustic-Phonetic Modeling"*. International Conference on Spoken Language Processing. ICSLP 02. 2002.

[19] JIANG, Hai. Er.; MENG Joo.; GAO, Yang. *"Feature extraction using wavelet packets strategy"*. Proceedings 42rd IEEE Conference on Decision and Control. Vol.5. 2003.

[20] RUSSEL, Stuart J.; NORVIG, Peter. *"Artificial intelligence: A modern approach"*. Prentice Hall. New Jersey. 1995.

[21] TEBELSKIS, Joe. *"Speech recognition using neural networks"*. Canegie Mellon University. Pittsburgh. 1995.

[22] DAVEPORT, Michael R.; GARUDADRI, Harinath. *"A Neural Net Acoustic Phonetic Feature Extractor Based on Wavelets"*. IEEE Pacific Rim Conference on Communication, Computers and Signal Processing. 1991.

[23] YUK, Doung-Suk; CHE, Chi Wei; JIN, Limin; LIN, Qiguang. *"Environment-independent continuous speech recognition"*.

[24] PICONE, J.; GANAPATHIRAJU, A. *"Support Vector Machines For Automatic Data Cleanup"*. Proceedings of the International Conference of Spoken Language Processing, Vol. 42. Beijing, China. 2000.

[25] JUNEJA, A.; ESPY-WILSON, C. *"Speech segmentation using probabilistic phonetic feature hierarchy and support vector machines"*. Proceedings of International Joint Conference on Neural Networks. Portland, Oregon, 2003.

[26] ABDULLA, W.; KECMAN, V.; KASABOV, N. *"Speech-background classification by using SVM technique"*. In Proc. of Artificial Neural Networks and Neural Information Processing ICANN/ICONIP 2003 International Conference. Istanbul, Turkey. 2003.

[27] ABE, Shigeo; INOUE, Takuya. *"Fuzzy Support Vector Machines for Multiclass Problems"*. European Symposium on Artificial Neural Networks, ESANN'02-Bruges (Belgium). 2002.

[28] CLARKSON, P.; MORENO, P.J. *"On the use of support vector machines for phonetic classification"*. In Proc. ICASSP 99, Phoenix, AZ USA, Vol. 2. 1999.

[29] MPORAS, I.; GANCHEV, T.; ZERVAS, P.; FAKOTAKIS. F. *"Recognition of Greek Phonemes using Support Vector Machines"*. SETN 2006, Crete, Greece. 2006.

[30] MELIN, P.; URIAS, J.; SOLANO, D.; SOTO, M.; LOPEZ, M.; CASTILLO O. *"Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms"*. Journal of Engineering Letters, Vol. 13. 2006.

[31] LIN, Che-Kuang; LEE Lin-Shan. *"Pronunciation Modeling for Spontaneous Speech Recognition using Latent Pronunciation Analysis (LPA) and Prior Knowledge"*. IEEE International Conference on Acoustic, Signal and Processing. ICASSP 2007. 2007

[32] MOORE, R. M. *"Methods and applications of interval analysis"*. Siam. Philadelphia. 1979.

[33] CALLEJAS-Bedregal, Roberto; BREDEGAL, Benjamn Ren Callejas. *"A generalization of the moore and yang integral approach"*. Preprint submitted to Computational & Applied Mathematics. 2005

[34] OPPENHEIM, A. V.; SCHAFER, R. W. *"Discrete-Time Signal Processing"*. Prentice Hall. 1989.

[35] CHEN, CHI-TSONG. *"Linear System Theory and Design"*. Oxford University Press. 1999.

[36] LYRA, A. *"Uma fundamentacao matematica para o processamento de imagens digitais intervalares"*. PhD thesis, Programa de Pos-graduacao em Engenharia Eletrica - Laboratorio de Engenharia de Computacao e Automacao - Universidade Federal do Rio Grande do Norte-UFRN. 2003.

[37] TRINDADE, Roque Mendes Prado; BREDEGAL, B. R. C.; DORIA, A. D. N. *"Basic Concepts of Interval Digital Signal Processing"* International Journal of Electrical, Computer, and Systems Engineering. 2010.
[38] TRINDADE, Roque Mendes Prado. *"Uma fundamentacao matematica para processamento digital de sinais intervalares"*. PhD thesis. Universidade Federal do Rio Grande do Norte-UFRN. Natal. 2009.
[39] CRUZ, M. M. C.; DORIA, A. D. N. TRINDADE, R. M. P. *"O Algoritmo K-means aplicado a Quantizacao intervalar para imagens digitais"* Escola Regional de Matematica Aplicada e Computacional-ERMAC. Recife-PE. 2004.
[40] FU, Y.; HUANG, T. S. *"Unsupervised locally embedded clustering for automatic high-dimensional data labeling"*. ICASSP2007. 2007.
[41] SUNDARAM, S.; NARAYANAN, S. *"Analysis of audio clustering using word descriptions"*. ICASSP2007. 2007.
[42] TRINDADE, Roque, M. P.; DORIA, A. D. N.; ACIOLY, B. M. *"An Interval Metric"*. New Advanced. 2010.