STABLE DISTRIBUTIONS FOR FEATURE EXTRACTION FROM SPEECH SIGNALS

Zdeněk Mžourek
Doctoral Degree Programme (1), FEEC BUT
E-mail: xmzour01@stud.feec.vutbr.cz

Supervised by: Zdeněk Smékal
E-mail: smekal@feec.vutbr.cz

Abstract: The aim of this paper is to introduce class of stable distributions as a potentional tool for statistical modelling of features extracted from speech signals. Alpha-stable distributions are generalization of the Gaussian distribution therefore they can be used in modeling of more variety of different problems. It is described why can stable distributions be useful in speech processing and potential useful applications are proposed for feature extractions and reduction.

Keywords: stable distribution, feature extraction, feature reduction, speech processing

1 INTRODUCTION

The most often used probability distribution for describing statistical properties of data in signal processig is the Gaussian distribution. Its usage is justified by the central limit theorem and empirical measurements show many physical processes can be well described by the Gaussian distribution or its mixtures.

However, the probability density of many physical phenomena has tails that are heavier that tails of the Gaussian density, therefore its use can lead to performance degradation. If a random process has heavier tails than the Gaussian density and we can assume that the process satisfies stability property then the class of stable distributions can used as a useful model which can improve performance and robustness.

2 STABLE DISTRIBUTIONS

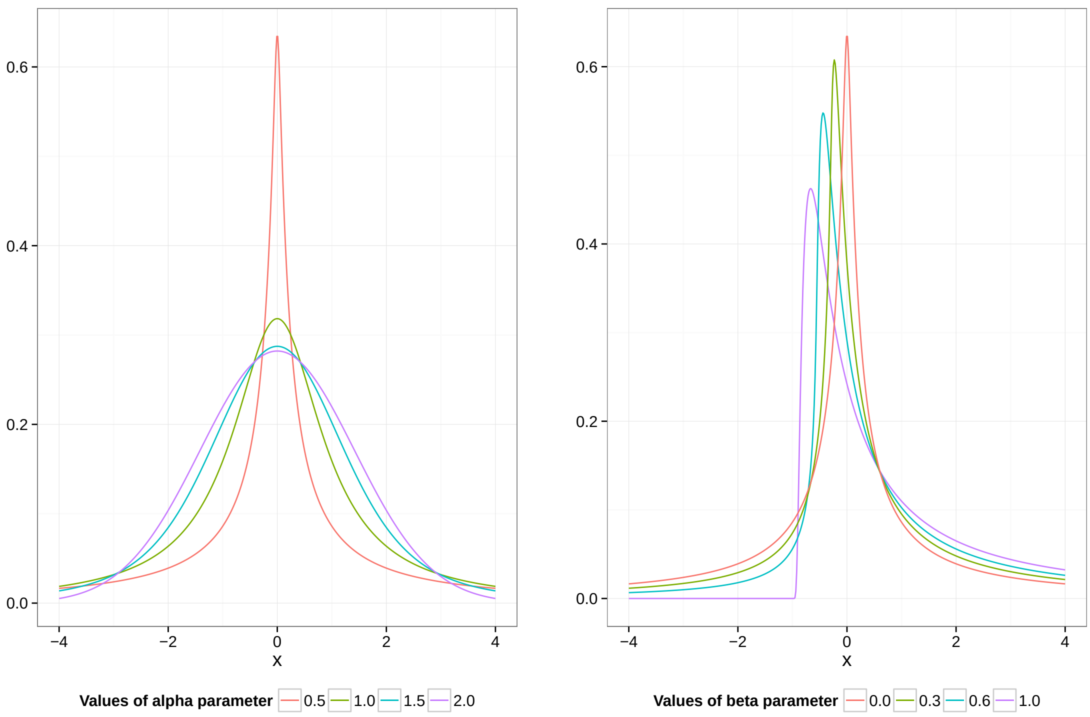
Stable distributions can be percieved as a class of distributions which can generalize the Gaussian distribution because Gaussian distribution is a special case of stable distributions [6]. Main difference between the Gaussian distribution and the stable distribution is following: Tails of the stable density are heavier than those of the Gaussian density therefore more “extreme values” can be observed. This can be seen on Figure 1a. When comparing line for $\alpha = 2.0$, which represents Gaussian density, with any other line the difference can be cleary seen.

As can be seen on the Figure 1b stable distributions can also represent class of asymmetrical densities.

2.1 PARAMETRIZATION

Stable distributions in general do not have a closed form of their density function. It means a density cannot be expressed in terms of mathematical functions such as a logarithm. This is very limiting for any manipulation therefore instead of distribution function its characteristic function is often used in such cases.

Characteristic function is in simple terms Fourier transform of distribution function and it has several very useful properties. For example, characteristic function of distribution function always exists



(a) $\beta = 0, \gamma = 1, \delta = 0$ (b) $\alpha = 0.5, \gamma = 1, \delta = 0$

Figure 1: Graphs of alpha stable densities for given parameter values

and it is unique for given distribution therefore it can be used instead of the distribution function (or density function) it represents.

There exist several different parameterizations of stable distributions based on the characteristic function. All of them involve different specifications of the characteristic function. A useful parameterization for practical applications [2] is following:

$$E\exp(\mathrm{j}tX) = \begin{cases} \exp\{-\gamma^{\alpha}|t|^{\alpha}[1+\mathrm{j}\beta\,(\tan\frac{\pi\alpha}{2})(\operatorname{sign}t)((\gamma|t|)^{1-\alpha}-1)]+\mathrm{j}\delta t\} & \alpha \neq 1 \\ \exp\{-\gamma|t|[1+\mathrm{j}\beta\frac{2}{\pi}(\operatorname{sign}t)(\ln|t|+\ln\gamma)]+\mathrm{j}\delta t\} & \alpha = 1, \end{cases} \tag{1}$$

where $X \sim \mathbf{S}(\alpha,\beta,\gamma,\delta)$ and $E\exp(\mathrm{j}tX)$ is the characteristic function of random variable X.

Parameters $\alpha,\beta,\gamma,\delta$ can be intepreterted in the following way:

α is stability parameter determines the impulsivness, $\alpha \in (0,2]$,

β is skewness parameter represents the symmetry, $\beta \in [-1,1]$,

γ is scale parameter, $\gamma \in (0,\infty)$,

δ is location parameter, $\delta \in (-\infty,\infty)$.

Scale parameter γ and location parameter δ are often interpreted as variance and mean respectfully [3], however, this is true only for some values of parameters α and β [2]. An example of inappropriate use of this terminology is the Cauchy distribution which has no mean or variance defined. Since the Cauchy distribution is a special case of stable distributions using this interpretation does not make any sense.

Different parametrizations have different “properties”. For example, a parametrization (1) allows to reduce correlation between estimated parameters α, β, γ, and δ. [2].

2.2 Motivation for using stable distributions

There exist several good reasons why consider the use of stable distributions as a model for some phenomena. The main argument is the generalized central limit theorem which states that sum of infinitely many i.i.d. random variables with or without variance converges to a stable distribution. If a signal can be thought of as a sum of large number of independent and identically distributed effects then a stable model may be appropriate [4]. An example can be noise in communication systems.

Another argument which can support this claim is empirical. Many large data sets exhibit heavy tails and skewness. Stable distributions can also describe impulisive nature of signals because they do not have to have a finite variance.

Another reason for use of stable distribution is following. Statistical properties of a signal can be describe by different statistics, most often used are mean, variance, median, percentiles etc. This statistics are used to estimate the statistical behavior of a signal. However, if the distribution of the random process which represents the signal is known, there is no need for individual statistics. Parameters of this distribution can be estimated and used instead of individual statistics beacuse it provides more “complete” information about random process.

2.3 Parameter estimation

There are several different approaches for parameter estimation. Any procedure for estimating stable parameters will find a “best fit” based on its criteria.

The maximum likelihood approach maximizes the likehood function numericaly. Methods based on quantiles try to match quantiles of those of stable distributions. There also exist methods based on fitting appropriate empirical characteristic function and methods based on fractional ordered momements [2, 3].

All methods will give estimated values of parameters that do not have to be same, e.g. the data is multi-modal. Therefore it is important to have some means of verifing that the estimated model is reasonable for given application.

It is important to verify the estimated model because there can exist many “false-positive” cases which can’t be interpreted in any meaningful way [2], therefore algorithms using this model can be performing poorly.

3 PROPOSED USE OF STABLE DISTRIBUTIONS FOR FEATURE EXTRACTION

In the following part two proposed methods are described as a way to use stable distributions for a feature extraction from speech signals.

3.1 Model based on spectral analysis methods

One of the most common approaches in feature extraction from speech signals is to use tools such as short-time Fourier transform and wavelets. Features often used in classification tasks of speech signal are for example spectral centroid or mel frequency cepstral coefficients (MFFCs) [3].

As a new feature based on spectral analysis the parameters of stable distributions can be used as follows. Once a spectral analysis method is used on a signal, a coefficients describing the distribution of frequencies of which is the signal composed. If those coefficients are thought of as a random process, then it is possible to estimate its distribution function.

In case of stable distribution which can be describe by 4 parameters, exactly 4 features are recieved independently of the signal length. Similar approach [3] was taken for classification of musical instruments. In that case new features based on stable distribution performed as well as MFFCs and in some cases they even outperformed them [3].

However, each transformation decomposes the signal in a slighty different way. If the best performace of this approach is disered, then is necessary to choose an appropriate spectral analysis method and a model estimation method. This will be the subject of future reasearch.

3.2 Dimension reduction of hiearchical speech features

Speech features can be devided into two main categories; local features and high-level features. Local features are directly extracted from the speech signal. High-level features are derived from a vector or a matrix of local features which elementens are extracted from a segmented speech signal.

In a classification process there is a need for each feature to be a scalar value in order to apply a machine learning method such as the support vector machine (SVM), the random forest or the linear discrimination analysis. The most often used mapping is to use some kind of statistics such as mean, median, variance, percentiles, etc. This approach can significantly increase the dimension of feature space because in order to describe data, several statistics have to be used.

One of main problems with analysis of high dimensional data is so-called curse of dimensionality. This problem manifests itself in many applications. One example of this phenomena can be demonstrated on the volume of unit cube. With increasing dimensions its volume is growing so fast that the data become sparse. This is problem mainly for any method that requires statistical signifiance. In order to obtain a significant result the amount of data needs to grow exponentially with the dimension [5].

In order to prevent this phenomena to manifest two main techniques of data manipulation are often used. The first one is variable selection and the second one is dimension reduction. The variable selection selects few variables which describes most of the data variability. The dimension reduction method reduce the dimension by transforming the feature space to a lesser dimensional space with the same ability to describe variability in the data set. Each approach has its strengths and weaknesses. The following part is focused on the dimension reduction method.

One of the most known methods dimension reduction is the principal component analysis. In simple terms, it transforms feature space into different vector space which basis are linear combinations of original feature space basis, therefore it is often difficult to interpret the new features from this transormed feature space. It is one of the main drawbacks of this approach generally [5].

In this article, a similiar approach is proposed but with different consequences to interpretability. Instead of estimating many different statistics to describe properties of a random process, it is "only" needed to estimate parameters of a model for its distributions. If a stability of this random process is assumed it is possile to use a stable distribution model. Therefore instead of dozens new features (individual statistics) only 4 features (parameters of estimated distribution model) are retrieved. This way a new feature space is obtained with a lower dimension than feature space composed of individual statistics and, in theory, there is no significant information loss. Each feature also has a reasonable intepretation in terms of the estimated model.

This approach can also help to a better understanding of modelled phenomena because it can provide more insight into the process behavior.

4 CONCLUSION

In this paper a new approach for feature extraction and reduction in speech processing is proposed. Stable distributions are presented as a useful tool for modelling statistical properties of a random process which may in this case represent coeeficients of a spectral analysis method or used instead of statistics such as mean, variance, median, etc. Suitability of this method is supported by theoretical claims such as generalized central limit theorem and also some empirical evidence.

In the next section two novel methods are proposed for feature extraction from speech signals. First approach is based on modelling distribution of coefficients retrieved by spectral analysis. This approach is inspired by similiar method used for music instruments classification.

The main contribution of this article is the second proposed method for feature reduction of the high-level features build upon local parameters of speech signals. In this part of the article is explained why is the most common approach for retrieving highlevel features of speech signal suboptimal. This method can in theory provided better classification results but verification will be the subject of future research.

5 ACKNOWLEDGEMENT

Research described in this paper was financed by the National Sustainability Program under grant LO1401 and by projects NT13499 (Speech, its impairment and cognitive performance in Parkinson's disease), COST IC1206, project "CEITEC, Central European Institute of Technology": (CZ.1.05/1.1. 00/02.0068) from the European Regional Development Fund. For the research, infrastructure of the SIX Center was used.

REFERENCES

[1] Kidmose, P.: Alpha-Stable Distributions in Signal Processing of Audio Signals, 41st Conference on Simulation and Modelling, Scandinavian Simulation Society (SIMS), 2000, s. 87-94

[2] Nolan, J. P.: Fitting Data and Assessing Goodness-of-fit with Stable Distributions, 1999

[3] Özbek M. E., Çek M. E, Savaci, F. A.: Skewed alpha-stable distributions for modeling and classification of musical instruments, Turk J Elec Eng & Comp Sci, Vol. 20, No. 6, 2012

[4] Shao, M., Nikias, C.L.: Signal processing with fractional lower order moments: stable processes and their applications, Proceedings of the IEEE, Vol. 81, Issue 7, 1993, ISSN: 0018-9219

[5] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009, ISBN: 0387848576

[6] Uchaikin, V.V., Zolotarev, V. M.: Chance and Stability: Stable Distributions and their Applications, 1999, ISBN 3-110935-97-X