ISSN 1392 – 1215 ELECTRONICS AND ELECTRICAL ENGINEERING 2011. No. 4(110)
ELEKTRONIKA IR ELEKTROTECHNIKA

T 121 SIGNAL TECHNOLOGY
SIGNALŲ TECHNOLOGIJA

Quality Estimation Methodology of Speech Recognition Features

R. Lileikyte
Recognition Process Department, Institute of Mathematics and Informatics,Vilnius University,
A. Goštauto str. 12, LT-01108 Vilnius, Lithuania, e-mail: rasalileikyte@gmail.com
L. Telksnys
Recognition Process Department, Institute of Mathematics and Informatics,Vilnius University,
A. Goštauto str. 12, LT-01108 Vilnius, Lithuania, e-mail: telksnys@mii.vu.lt

Introduction

The concept of quality may be defined as the degree of element goodness given a certain criterion [1–18]. Thus, it can be determined by comparing a set of inherent characteristics with a set of requirements. If inherent characteristics are met by requirements, then high quality is achieved [6]. Therefore, more quality descriprions are represented in [2, 3, 18].

Selection of quality feature set is essential in speech recognition system. The aim is to reduce amount of data, discard irrelevant information and enhance those that contribute the most discriminate information [7]. Furthermore, classification error is significantly reduced if it's operated with quality features. Two approaches of feature quality estimation are provided by other authors: 1) estimate metrics that focus on data complexity measures [1, 13, 14, 16]; 2) assess feature quality by classification error [1, 5, 16]. However, it's simpler to compute metric instead of constantly executing classification process. Finally, the studies don't suggest the way for choosing the most suitable quality metric. And there is no investigation made for speech features.

We propose quality estimation methodology in this paper. The methodology combines quality metrics with classifier similarity measure results. Investigation is performed for Lithuanian context phonemes, applying three speech feature systems: Linear Frequency Cepstral Coefficients (LFCC), Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP). We choose Dynamic Time Warping (DTW) classifier.

The paper is organized as follows: the selected set of quality measures are reviewed firstly (more details can be found in [1, 4, 5, 8, 9, 15]). Measures are grouped into following goups: overlap of individual features, separability of classes, measures of geometry and topology. Then, quality estimation methodology is formulated. Afterwards, experimental results are given. Finally, conclusions are made.

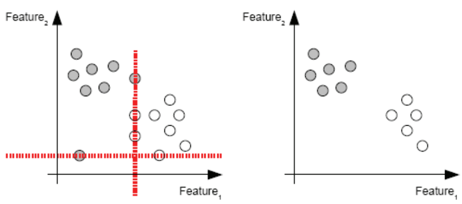
Overlap of individual features

Fisher's discriminant ratio (F1). Fisher's discriminant ratio is calculated for the problem. It's based on the ratio of between-class variance to within-class variance.

Maximum Fisher's discriminant ratio (F1m). Fisher's discriminant ratio is calculated. For multidimensional problem, it's not necessary all features to contribute to class discrimination. The problem will be easy if there exists at least one discriminating feature element. It is considered only the maximum over all features.

Volume of overlap region (F2). The measure is the overlap of the tails of classes – conditional distributions. We can measure it by finding for each feature the maximum and minimum values of each class, and then calculating the length of the overlap region normalized by the range of values spanned by both classes.

Feature efficiency (F3). The measure describes how much each feature contributes to the separation of classes. If there is an overlap in the feature values of classes, we consider the classes ambiguity in that region along that dimension. The ambiguity between classes can be avoided by separating only those points that lie outside the overlapping region in each chosen dimension. The efficiency of each feature is defined as the fraction of all remaining points separable by that feature (Fig. 1).



a) b)

Fig. 1. Removing overlapping points: a – classes with overlapping points; b – classes after removing overlapping points (adapted from [11])

Overstep boundary error rate (F4). Suppose that every class is a super sphere which center is class - mean and radius is equal to the farthest distance from its class - mean to its samples. Overstep boundary error occurs if there exists a sample that belongs to the other classes and comes to the super sphere of this class. The measure is normalized by the total number of points.

Separability of classes

Error rate of linear classifier (L1). The measure corresponds to the error rate of linear classifier.

Length of class boundary (N1). This metric defines the percentage of points in the dataset that lie near the class boundary. It counts the number of points connected to the opposite class by an edge in the Minimum Spanning Tree (MST). These points are considered to be close to the class boundary. The count is normalized by number of total edges in MST (Fig. 2).

Fig. 2. A minimum spanning tree connecting points of two classes (adapted from [5])

Intra/inter class nearest neighbour distances (N2). The measure assess the dispersion of points within classes relative to the separability between classes. It is the ratio of the average distance to intra-class nearest neighbor and the average distance to inter-class nearest neighbor (Fig. 3).

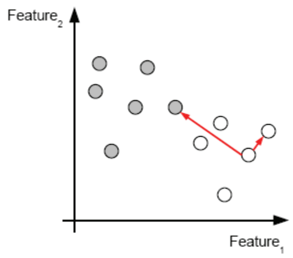


Fig. 3. Distance from each point to its nearest neighbor within the class and outside the class of two classes (adapted from [11])

Error rate of nearest neighbor classifier (N3). This measure refers to the estimated error rate of the nearest neighbor classifier using the leaving-one-out validation method. Nearest neighbor classifier is sensitive to the number of misplaced points, while linear classifier is sensitive to their location.

Measures of geometry and topology

Nonlinearity of nearest neighbour and linear classifiers (L2, N4). Given a training set, the method first creates a test set by linear interpolation. Then the error rate of the classifier (trained by the same training set) on this test set is measured. Nonlinearity of linear classifier and nearest neighbour classifier is considered.

Thickness of manifolds (T1). The measure is calculated using ε-topology. Adherence subsets are grown to the highest order such that it includes only points of the same class (Fig. 4). Such „balls“ number is normalised by total points number. As a result, all samples are grouped into hyper-spheres that contain samples of the same class. The measure provides interior description. Additionally, the number of small “balls” indicates the complexity of the classes boundaries.

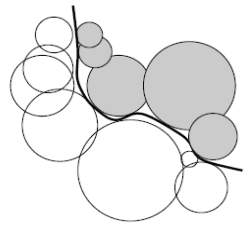


Fig. 4. Adherence subsets of two classes (adapted from [5])

Methodology of quality measure estimation

We will propose the methodology for feature quality measure estimation. The methodology focuses on combination of quality metric and chosen classifier similarity measure. As far as classification error depends on particular classifier [15], there can’t be the only best metric for all classifier types.

The quality measure estimation scheme is following. Firstly, quality metrics should be calculated for classes (Fig. 5). Secondly, similarity for classes based on classifier should be estimated. Then, correlation should be assessed for metrics and similarity results. As a result, metric with highest correlation should be chosen.

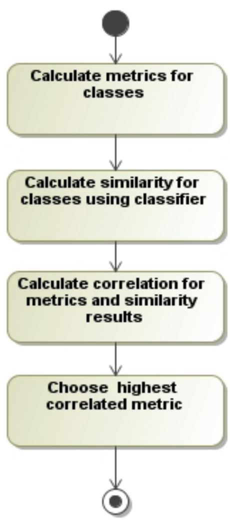


Fig. 5. Scheme of quality measure estimation

Experimental setup

We selected 11 measures (described above) for experiment. Exploration was made using collection of 9 sets of different phonemes representing different classes, each set consisting of 100 instances (pronounced by one female speaker). Most frequent Lithuanian phonemes were selected as target for this experimental study [12]: [a], [i], [e], [j], [k], [t], [t’] [s], [s’] (t’ and s’ are soft consonants).

18 two-classes combinations were calculated during experiment, for high overlap and small overlap cases: s - s', s' - t, t - k, t - s, t' - k, t' - t, a - s, a - e, k - s, a - i, a - k, e - i, e - k, e - s, e - t, i - s, s' - j, s' - k.

12th order LFCC, MFCC and PLP analyses were selected for experiment. MFCC and PLP feature sets are most frequently used for speech signals, and LFCC was chosen for comparison purposes.

DTW classifier was implemented [11, 17]. We use DTW distance to relate metrics with similarity of classes.

Experiment results

The metrics were calculated for chosen classes combinations. Then correlation of DTW classifier similarity and metrics was estimated. Three cases of highest correlation are given in Table 1, Table 2, Table 3. First rows of tables "1 result" identifies metric with the highest correlation.

Table 1. Metrics correlation for LFCC

Corr.	F1	F1m	F4	N1	N2
1 result	0%	0%	11%	0%	**89%**
2 result	**78%**	0%	11%	0%	11%
3 result	6%	6%	**56%**	33%	0%

Table 2. Metrics correlation for MFCC

Corr.	L3	F1	F1m	F4	N1	N2
1 result	0%	17%	0%	16%	0%	**67%**
2 result	0%	**56%**	0%	11%	0%	33%
3 result	6%	6%	6%	**66%**	16%	0%

Table 3. Metrics correlation for PLP

Corr.	L2	F1	F1m	F4	N1	N2
1 result	0%	0%	0%	6%	0%	**94%**
2 result	0%	**67%**	0%	17%	10%	6%
3 result	22%	22%	17%	11%	**28%**	0%

The results showed that DTW-based distance had highest correlation with metric *N2*. It describes the ratio of the average distances of intra-class and inter-class nearest neighbors. Moreover, metric N2 took 89% of the highest correlation values for LFCC, 67% for MFCC and 94% for PLP analysis.

Second highest correlated metric was *F1*. It estimates Fisher's discriminant ratio and is refered to the ratio of between-class variance to within-class variance. Metric F1 covered 78% of the highest correlation values for LFCC, 56% for MFCC and 67% for PLP.

Afterward followed metric *F4* for LFCC and MFCC. The metric describes the overstep boundary error rate. It took 56% for LFCC and 66% for MFCC. Contrarily *N1* covered 28% of the highest correlation values for PLP analysis. It identifies length of class boundary counting the number of points connected to the opposite class by an edge in MST.

Results and conclusions

The paper proposes the methodology for feature quality estimation. The methodology consists of four steps: 1) calculate metrics for classes; 2) estimate similarity of classes using chosen classifier; 3) calculate correlation of metrics and similarity results; 4) choose the highest correlated metric. Due to the fact that classification error depends on particular classifier, there can't be the only best metric for all types of classifiers.

Exploration for speech features was made with 11 different metrics that were overviewed in the study. Metrics are grouped into three groups: 1) overlap of individual features; 2) separability of classes; 3) measures of geometry and topology. Three feature systems were tested: LFCC, MFCC, PLP analysis. Experiment was made for Lithuanian context phones, combining results with DTW distance. It was found that DTW classifier distance has the highest correlation with metric N2. As a result, N2 was constituted as the best quality metric for speech features evaluation. It describes the ratio of the average distances of intra-class and inter-class nearest neighbors. The metric gave the most accurate results for LFCC, MFCC, PLP feature systems: 89% of highest correlation for LFCC, 67% for MFCC and 94% for PLP.

The results of the experimental study points us to the direction of future research. The following study derivation could be quality estimation of speech feature sets using N2. The process could concentre on the principal N2 characteristic: N2 gains value near zero for small overlapped classes, on the other hand the bigger N2 identifies the higher classes overlap. With the result of this property, estimation process should encompass the following steps: 1) calculate N2 for classes with small overlap 2) calculate N2 for classes with high overlap 3) choose feature system that best agrees with principal N2 characteristic.

References

1. **Bernadó–Mansilla E., Ho T. K**. Domain of competence of XCS classifier system in complexity measurement space // IEEE Transactions on Evolutionary Computation, 2005. – Vol. 9(1). – P. 82–104.
2. **Eidukas D., Kalnius R**. Models Quality of Electronics Products // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 29–34.
3. **Eidukas D., Kalnius R.** Stochastic models of quality level mechatronic products // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 3(83). – P. 43–48.
4. **Ho T. K**. Data complexity analysis for classifier combination // Springer Berlin / Heidelberg, Multiple classifier systems, 2001. – Vol. 2096(2001). – P. 53–67.
5. **Ho T. K., Basu M**. Complexity measures of supervised classification problems // IEEE transactions on pattern analysis and machine intelligence, 2002. – Vol. 24(3). – P. 289–300.
6. ISO 9000, 9001, and 9004 Quality Management Definitions. Online: http://www.praxiom.com/iso–definition.htm.
7. **Lefkaditis D., Tsirigotis G.** Intelligent Optical Classification System for Electronic Components // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 10–14.
8. **Li Z., Ditang F**. The research on speech feature representation method and distance measure method // Pattern recognition, 9th International Conferency, 1998. – Vol. 1. –P. 631–633.

9. **Luengo, J., Herrera, F**. On the use of Measures of Separability of Classes to characterize the Domains of Competence of a Fuzzy Rule Based Classification System // Proceedings of the 2009 International Fuzzy Systems Association congress, European Society for Fuzzy Logic and Technology conference. – Lisbon, Portugal, 2009. – P. 1027–1032.
10. **Myers C., Rabiner L., Rosenberg A**. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition // IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980. – Vol. 28(6). – P. 623–625.
11. **Macia N., Bernado–Mansilla E.** Modeling problem transformations based on data complexity // CCIA 2007, Desè Congrés Internacional de l'Associació Catalana d'Intel·ligència Artificial. – Vol. 12.
12. **Raškinis G., Kazlauskienė A**. Phone frequency in standard Lithuanian // Respectus philologicus, 2009. – No. 16(21). – P. 169–182.
13. **Singh S**. Multiresolution Estimates of Classification Complexity // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. – Vol. 25(12). – P. 1534–1539.
14. **Singh S**. PRISM – A novel framework for pattern recognition // Pattern Analysis and Applications, 2003. – Vol. 6. – P. 131–149.
15. **Sohn S. Y**. Meta analysis of classification algorithms for pattern recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999. – Vol. 21(11). – P. 1137–1144.
16. **Sotoca J. M., Sanchez J. S., Mollineda R. A**. A review of data complexity measures and their applicability to pattern classification problems // Actas del III Taller Nacional de Mineria de Datos y Aprendizaje. – TAMIDA, 2005. – P. 77–83.
17. **Tamulevičius G., Arminas V., Ivanovas E., Navakauskas D.** Hardware accelerated FPGA implementation of Lithuanian isolated word recognition system // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 3(99). – P. 57–62.
18. **Targamadzė A., Petrauskienė R., Rubliauskas D**. Influence of technologies on quality of distance learning // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 6(102). – P. 131–134.

Received 2011 02 15

R. Lileikyte, L. Telksnys. Quality Estimation Methodology of Speech Recognition Features // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 4(110). – P. 113–116.

The best feature set selection is the key of successful speech recognition system. Quality measure is needed to characterize the chosen feature set. Variety of feature quality metrics are proposed by other authors. However, no guidance is given to choose the appropriate metric. Also no metrics investigations for speech features were made. In the paper the methodology for quality estimation of speech features is presented. Metrics have to be chosen on the ground of their correlation with classification results. Linear Frequency Cepstrum (LFCC), Mel Frequency Cepstrum (MFCC), Perceptual Linear Prediction (PLP) analyses were selected for experiment. The most proper metric was chosen in combination with Dynamic Time Warping (DTW) classifier. Experimental investigation results are presented. Ill. 5, bibl. 18, tabl. 3 (in English; abstracts in English and Lithuanian).

R. Lileikytė, L. Telksnys. Šnekos signalų atpažinimo požymių kokybės matas – literatūros apžvalga // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 4(110). – P. 113–116.

Geriausių požymių rinkinio sudarymas yra svarbus uždavinys kalbos atpažinimo sistemoje. Tam reikalingas kokybės matas, kuris leistų įvertinti požymių kokybę. Klasifikuojamiems požymiams įvertinti yra pasiūlyta nemaža metrikų. Tačiau nagrinėtuose darbuose nėra suformuluoto tinkamos metrikos pasirinkimo kriterijaus, nenagrinėtas metrikų pritaikymas kalbos signalo požymiams. Šiame darbe pateikiama metodologija kalbos signalo požymių kokybei nustatyti. Metrikos parenkamos atsižvelgiant į jų koreliaciją su klasifikatoriaus rezultatais. Eksperimentas vykdytas su kepstro (LFCC), melų skalės kepstro (MFCC) bei tiesinės suvokimo prognozės koeficientais. Tinkamiausia metrika pasiūlyta naudojant tiesinį laiko skalės kraipymo (DTW) klasifikatorių. Pateikiami ekperimento rezultatai. Il. 5, bibl. 18, lent. 3 (anglų kalba; santraukos anglų ir lietuvių k.).