ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2012. No. 2(118)
ELEKTRONIKA IR ELEKTROTECHNIKA

SIGNAL TECHNOLOGY
T 121
SIGNALŲ TECHNOLOGIJA

Quality Measurement of Speech Recognition Features in Context of Nearest Neighbour Classifier

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

crossref http://dx.doi.org/10.5755/j01.eee.118.2.1165

Introduction

The selection of the quality feature system is the key of successful speech recognition system. Therefore, the inquiry can be stated - how to choose the quality feature system? The concept of quality can be defined by comparing a set of inherent characteristics with a set of requirements. If these subjects are met, then high quality is achieved [16]. Also, more quality descriptions are represented in [6, 7, 18, 26]. The choice of quality features is the essential as low classification error can be achieved if quality features are used. On the contrary high classification error is achieved for not quality feature system. A variety of speech feature systems exists. Accordingly, currently quality of features is used to estimate by calculating the classification error. However, this method is limited as it causes running classification experiments with each explored feature system.

The major issue of current research is to propose the method for quality estimation of speech recognition feature system with the approach that doesn‘t require performing classification experiments. Moreover, the method is based on metrics: feature volume of class boundary, nearest neighbour distances ratio of classes, overstep of class boundary.

The paper is structured as follows. Firstly, the proposed method for quality estimation of speech recognition feature system is presented. Then, metrics for quality estimation of speech recognition features are displayed. Next, the results of experimental researches are represented. Finally, conclusions are made.

Method of quality measurement for speech recognition features

The quality of features has been estimated using classification error. Therefore, this method requires classification experiments running with each feature system. Let’s suppose that we have five feature systems S_1, S_2, S_3, S_4, S_5 under exploration. In order to choose the quality feature system, the classification process must be run five times. Consequently, we propose a new method of quality estimation for speech recognition features. Quality feature system is defined using metrics instead of making classification (Fig. 1). Let’s suppose that quality feature system was established S_4.

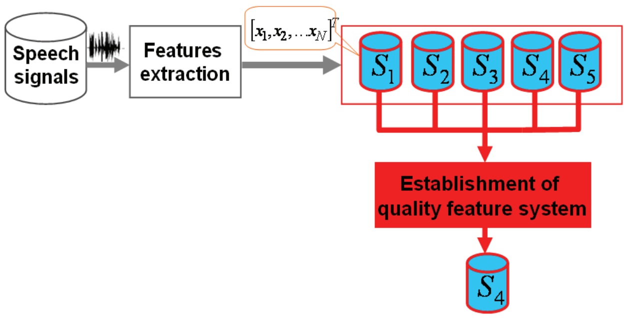

Fig. 1. Features quality estimation using the proposed method

The detailed scheme of the method is presented in Fig. 2 and steps of the method are displayed below.

Selection of feature system. The metrics are calculated for investigated feature systems.

Calculation of metrics for classes. The metrics are calculated for instances combinations that belong to certain classes combination

$$V_{nm}^k(i,j) = \sigma^k(h_n^i, h_m^j)\,, \tag{1}$$

where $\sigma^k(h_n^i, h_m^j)$ - k-th metric calculated for h_n^i n-th instance of i-th class and h_m^j is m-th instance of j-th class, $k=1,\ldots,K$, $n=1,\ldots,NH$, $m=1,\ldots,MH$, $i, j=1,\ldots C$, $i \neq j$, K - number of metrics, NH, MH - instances numbers of i-th and j-th classes, C - number of classes. Low value of metrics identifies high quality feature system. The average of metrics is calculated for each combination

$$VG^k(i,j) = \frac{\sum_{n=1}^{M}\sum_{m=1}^{M} V_{nm}^k(i,j)}{PBsk}, \quad (2)$$

where $PBsk$ is the number of classes combinations.

Calculation of *quality for the feature systems*. After the averages of metrics are calculated for each class combination, the index of feature system is calculated for each classes combination by „voting"

$$FG^p(i,j) = \arg\max_p \sum_{k=1}^{K} \upsilon(S^p = f(\arg\min VG^k(i,j))), \quad (3)$$

where $f(\arg\min VG^k(i,j))$ defines that S^p is the p-th feature system that was established to have the lowest value of k-th metric for i-th and j-th class combination, P – number of feature systems, $p = 1,...,P$, $\upsilon(\cdot)$ returns 1 if the equality is satisfied, otherwise it gives 0. Then „the quality indexes of feature systems" are calculated

$$RFG^p = \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} o(FG^p(i,j))}{KBsk}, \quad (4)$$

where $o(\cdot)$ function gives 1 if p-th feature system is identified as the best one in case of i-th and j-th class combination. Otherwise, 0 is given, $KBsk$ – the number of class combinations. It defines the quality of feature system in percentages, where 0% identifies low quality and 100% high quality

Decision taking about the quality feature system. The quality feature system is established with the highest quality index

$$SFG = \arg\max_p RFG^p. \quad (5)$$

Validation of the adequateness of the proposed method. To validate the adequateness of the proposed method classification error has to be calculated for the explored feature systems.

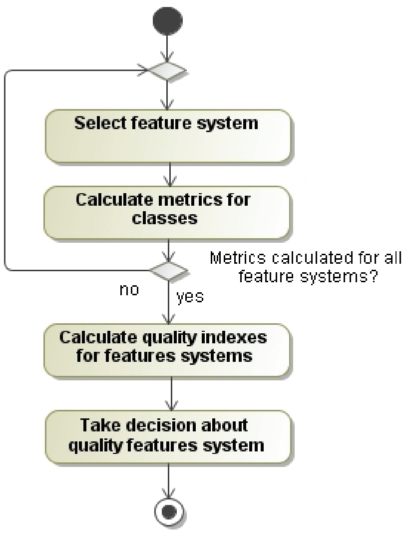

Fig. 2. Scheme of the method for establishing the quality feature system

Quality metrics for speech recognition features

Length of class boundary (G1). The metric counts the number of samples connected to the opposite class by the edges in Minimum Spanning Tree (MST) [1, 14, 20]. Let $L = \{l_1, l_2, \ldots, l_K\}$ is the vertex set of MST [2, 3, 9]. These vertexes (samples) belong to the edges (Euclidean distance between samples), that connect vertexes of different classes, K – the number of the vertexes. The metric is defined as the ratio of such vertexes number to the all vertexes number

$$G1 = \frac{K}{N}. \quad (6)$$

Ratio of intra/inter class nearest neighbour distances (G2). The metric is calculated as the ratio of nearest neighbours distances sum from the same class over the sum from opposite class [13, 14, 24, 25]

$$G2 = \frac{\sum_{i=1}^{C}\sum_{n=1}^{N_i} \min_k d(x_n^i, x_k^i)}{\sum_{i=1, j=1, i\neq j}^{C}\sum_{n=1}^{N_i} \min_m d(x_n^i, x_m^j)}, \quad (7)$$

where $\min_k d(x_n^i, x_k^i)$ is minimal Euclidean distance between x_n^i - n-th sample of i-th class and x_k^i - k-th sample of i-th class, N_i - number of samples in i-th class.

Overstep boundary (G3). Let us suppose that every class is represented by the sphere. The radius of the sphere is defined as the distance from centre to the farthest sample of the class, where the centre is the mean of the class [19]. The metric is calculated:

$$G3 = Q[d(\mu^i, x_n^j) \le r^i] \cdot \frac{1}{C-1} \cdot \frac{1}{N}, \quad (8)$$

$$r^i = \max_n d(\mu^i, x_n^i), \quad (9)$$

$$d(\mu^i, x_n^j) \le r^i, \quad (10)$$

where $d(\mu^i, x_n^i)$ is Euclidean distance between μ^i - centre of the i-th class and x_n^i - n-th sample of i-th class, r^i - radius of the sphere of i-th class, $Q[d(\mu^i, x_n^j) \le r^i]$ is the number of samples satisfying the overstep condition (10).

Experimental results and discussion

The experimental researches were made with 14 sets of different phonemes. Each set consisting of 100 instances of single phoneme. Most frequent Lithuanian phonemes were selected as the target for this experimental study [22]: [a], [e], [i], [j], [k], [m], [n'], [o:], [r], [r'], [s], [s'], [t], [t'] (n', t', s' are soft consonants and o: is long vowel). Data was used from University of VDU, VDU-TRI4 repository [23]. Experimental researches were made in context of NN classifier [4, 5, 10, 12, 15]. We employed two feature systems for the experimental researches: 12th order Linear

`Frequency Cepstral Coefficients (LFCC) [8, 21] and 12th order Perceptual Linear Prediction (PLP) [11, 17]. Experiments were made with three data sets: Data set1, Data set2, Data set3. Each data set consisted of 14 classes, each class having 100 instances. Calculations were made for 91 pairs of classes, including 10000 instances combinations for each pair of classes.

The experimental results of the proposed method are presented by providing quality indexes of feature systems. The results of the experiments with estimated quality indexes of feature systems are showed in Fig. 3.

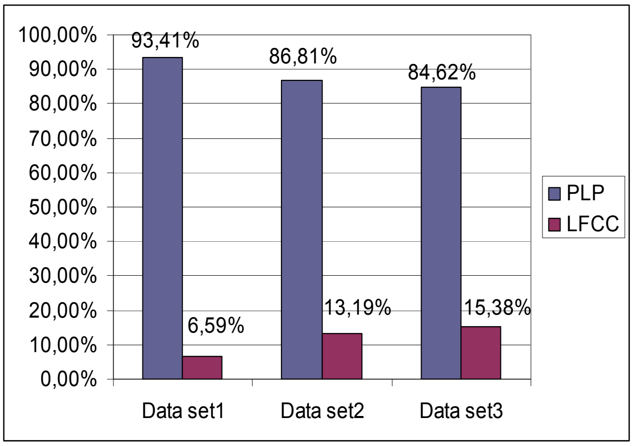

Fig. 3. Experimental results of quality indexes of feature system

Also, in order to validate the adequateness of the proposed method, NN classification errors were calculated and are provided in Table 1.

Table 1. Nearest neighbour classification error

Data set	Features System	Classification error
Data set1	PLP	9,14 ± 1,11%
	LFCC	13,48 ± 1,51%
Data set2	PLP	8,12 ± 0,93%
	LFCC	13,12 ± 1,50%
Data set3	PLP	10,09 ± 1,18%
	LFCC	13,66 ± 1,55%

The experimental results of Data set1. The results of quality indexes showed that PLP feature system gained the highest quality index. It identifies PLP system to be the highest quality. As well as this, the results of NN classification error showed that the lowest classification error was achieved for PLP.

The experimental results of Data set2. The results of quality indexes identified that PLP gained the highest quality index. Additionally, the results of NN classification error showed that the lowest classification error was achieved in case of PLP feature system.

The experimental results of Data set3. The results of quality indexes also showed that PLP gained the highest quality index. As well as this, the results of NN classification error identified that the lowest classification error was achieved for PLP.

As a result, the experimental results of feature quality estimation using the proposed method coincided with the results of quality estimation using the classification error. In both cases, with all three data sets it was established that PLP is the quality feature system. The results of the experimental researches approbated the correctness of the proposed method.

Conclusions

The paper attributes to the issue of quality estimation method of speech recognition features. The new method for quality estimation of speech recognition features is proposed that doesn't require executing classification experiments. The method is based on the usage of metrics: feature volume of class boundary, nearest neighbour distances ratio of classes and as well as this - overstep of class boundary.

The experiments were performed with feature systems of PLP and LFCC, composing three data sets. The experimental results of feature quality estimation using the proposed method showed that PLP is the quality feature system, as it gained the highest quality index. As well as this, Nearest neighbour classification error was calculated to validate the adequateness of the method. The lowest error was achieved for PLP feature system. Therefore, the highest quality index identifies the feature system with the lowest classification error

In conclusion, the experimental results of feature quality estimation using the proposed method coincided with the results of quality estimation using the classification error. Therefore, the results of the experimental researches approbated the adequateness of the proposed method of quality estimation of speech recognition features.

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. **Bock F**. An algorithm to construct a minimum direct spanning tree in a direct network // Developments in Operations Research. – Gordon and Breach, New York, 1971. – P. 29–44.
3. **Cheriton D., Tarjan R. E**. Finding minimum spanning trees // In SIAM Journal on Computing, 1976. – Vol. 5. – P. 724–741.
4. **Cover T. M., Hart P. E**. Nearest neigbour pattern classification // IEEE Transactions on Information Theory IT, 1967. –No. 13(1). – P. 21–27.
5. **Dinuls R., Lorencs A., Mednieks I.** Performance Comparison of Methods for Tree Species Classification in Multispectral Images // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 5(111). – P. 119–122.
6. **Eidukas D., Kalnius R**. Models Quality of Electronics Products // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 29–34.
7. **Eidukas D., Kalnius R.** Stochastic models of quality level mechatronic products // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 3(83). – P. 43–48.

8. **Furui S.** Digital Speech Processing, Synthesis, and Recognition. Second edition. – Signal Processing and Communications, 2001. – 452 p.
9. **Graham R. L., Hell P.** History of the Minimum Spainning Tree Problem // Annals of the History of Computing, 1985. – No. 7(1). – P. 43–57.
10. **Golipour L., O'Shaughnessy D**. Context–independent phoneme recognition using a K–Nearest Neighbour classification approach // IEEE International Conference on Acoustics Speech and Signal Processing, 2009. – P. 1341–1344.
11. **Hermansky H.** Perceptual linear predictive (PLP) analysis of speech // Journal of Acoustical Society of America, 1990. – No. 87(4). – P. 1738–1752.
12. **Henry J. L.** A k–Nearest Neighbour Method for Managing the Evolution of a Learning Base // Proceedings of the Fourth International Conference on Computational Intelligence and Multimedia Applications (ICMA'2001). – Yokusika City, Japan, 2001. – 357–361.
13. **Ho T. K**. Data complexity analysis for classifier combination // Springer Berlin / Heidelberg, Multiple classifier systems, 2001. – Vol. 2096(2001). – P. 53–67.
14. **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.
15. **Hua O., Ji A., He O.** Multiple Real–valued K Nearest Neighbor Classifiers System by Feature Grouping // IEEE International Conference on Systems Man and Cybernetic (SMC), 2010. – P. 3922– 3925.
16. **ISO 9000**, 9001, and 9004 Quality Management Definitions. Online: http://www.praxiom.com/iso–definition.htm.
17. **Junqua J. C., Haton J. P**. Robustness in automatic speech recognition: Fundamentals and applications. – Kluwer Academic Publishers Norwell, MA, USA, 1996. – 440 p. DOI: 10.1007/978-1-4613-1297-0.
18. **Lefkaditis D., Tsirigotis G.** Intelligent Optical Classification System for Electronic Components // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 10–14.
19. **Li Z., Ditang F**. The research on speech feature representation method and dis–tance measure method // Pattern recognition, 9th International Conferency, 1998. – Vol. 1. – P. 631–633.
20. **Lileikyte R., Telksnys L**. Quality Estimation Methodology of Speech Recognition Features // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 4(110). – P. 113–116.
21. **Noll A. M.** Cepstrum pitch estimation // Journal of the Acoustical Society of America, 1967. – Vol. 4. – P. 293–309. **Raškinis G., Kazlauskienė A**. Phone frequency in standard Lithuanian // Respectus philologicus, 2009. – No. 16(21). – P. 169–182.
22. **Raškinis A., Raškinis G., Kazlauskienė A.** Lietuvių bendrinės šnekamosios kal–bos garsų fondo kūrimo principai // Lituanistica, 2004. – No. 4 (60). – P. 53–62.
23. **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.
24. **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), 2005. – P. 77– 83.
25. **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 09 05
Accepted after revision 2011 11 16

R. Lileikyte, L. Telksnys. Quality Measurement of Speech Recognition Features in Context of Nearest Neighbour Classifier // Electronics and Electrical Engineering. – Kaunas: Technologija, 2012. – No. 2(118). – P. 9–12.

The quality feature set is a key of importance of successful speech recognition system. The quality of features is estimated by classification error. Yet, this method is limited as the classification experiments must be run with each feature system. The major issue of this paper is to propose the method for quality estimation of speech recognition features that is based on metrics and does not require classification experiments. Experimental researches were made in context of Nearest neighbour classifier usage. Within the proposed method PLP was established to have the higher quality comparing to LFCC. The adequateness of the method was validated by Nearest neighbour classification error. Ill. 3, bibl. 25, tabl. 1 (in English; abstracts in English and Lithuanian).

R. Lileikytė, L. Telksnys. Šnekos signalų atpažinimo požymių kokybės vertinimas, kai klasifikavimui naudojamas artimiausio kaimyno klasifikatorius // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2012. – Nr. 2(118). – P. 9–12.

Kokybiškos požymių sistemos parinkimas yra svarbus etapas šnekos atpažinimo sistemos projektavimo etapas. Požymių kokybė yra matuojama klasifikavimo klaida. Tačiau taikant šį metodą klasifikavimo eksperimentai turi būti atliekami su kiekviena tiriama požymių sistema. Šio darbo tikslas – pateikti metodą požymių kokybei vertinti, kuris būtų grindžiamas metrikomis ir nereikalautų atlikti eksperimentų. Darbe atliki eksperimentiniai tyrimai naudojant artimiausio kaimyno klasifikatorių. Taikant pateiktą metodą nustatyta, kad PLP yra kokybiška požymių sistema, palyginti su LFCC požymių sistema. Metodo teisingumas patikrintas apskaičiavus artimiausio kaimyno klasifikatoriaus klaidą. Il. 3, bibl. 25, lent. 1 (anglų kalba; santraukos anglų ir lietuvių k.).