Acta Technica 62 (2017), No. 6B, 1399–1408

A precision evaluation index system based on soft fuzzy rough sets

Baohua Gao[1], yan Wu[1]

Abstract. In order to improve evaluation precision of multiple English teaching and writing capacities, a kind of evaluation precision of multiple English teaching and writing capacities based on soft fuzzy rough set was proposed in the thesis. Firstly, factors which influence English writing capacities of undergraduates was analyzed; an evaluation indicator system which reflects English writing capacities of undergraduates is eventually established; then, based on rough set theory, soft fuzzy rough set model was used and improved so as to make them have capacities of English writing label problem; Next, this kind of model shall be used to analyze evaluation of English writing capacities. Finally, effectiveness of the algorithm is testified through simulation experiment.

Key words. Soft fuzzy, Rough set, English writing, Capability evaluation.

1. Introduction

Correct classification for English writing capacities of undergraduates not only can make objective evaluation on writing skills of students but can also provide important references for immediate adjustment of writing teaching scheme for teachers in the process of grouping teaching. Simple linear model classification is used in traditional evaluation method; however, English writing capacities of undergraduates are influenced by many factors, making evaluation for writing capacities of undergraduates have high-dimensional and non-linear characteristics. Evaluation result of traditional methods has great differences with actual condition, which can not meet actual requirements.

In recent years, with rapid development of artificial neural network technique, data mining method based on neural network provides a new channel for English writing capacities of undergraduates. At present, propagation Neural Networks and soft fuzzy rough set among numerous artificial neural network types is a kind of the most representative network models which are the most widely used. Automatic

[1]School of Foreign Languages, Northwestern Polytechnical University, Xi'an Shaanxi, 710072, China

evaluation system based on soft fuzzy rough set is used in Literature [1–3] to conduct evaluation for English writing capacities of students, which to some extent improves precision rate of classification. However, network topology structure of soft fuzzy rough set is hard to be determined with slow convergence rate, and it is easy to be involved in local minimum in the process of learning, and other defects.

As rough set has strong associative memory capacity, it is widely used in the field of pattern recognition; fairly successful experience is obtained. Therefore, soft fuzzy rough set is used in the thesis to make evaluation for English writing capacities of undergraduates so as to solve the problem of not high classification precision of traditional methods.

2. Construction of writing evaluation index system

English writing capacities of undergraduates are influenced by many factors; indicators are further detailed in the thesis based on current researches; evaluation indicator system which reflects English writing capacities of undergraduates is finally established. The system has 12 indicators, which are oral (X), listening (Xz), vocabulary quantity (X3), grammar (X4), reading comprehension (X5), translational level (X6), learning motivation (X7), learning interest (X8), cross-cultural communicative competence (X9), writing strategy (X10), discourse knowledge (X11), British and American culture knowledge (X12). Data collection process is as follows: evaluation form shall be made according to 10 scores for all indicators as full scores; relevant rating standards shall be given. 25 English teachers shall mark English writing of students (60 people in total) who are no English major in 2 natural teaching classes of a university in the form of interview and written examination; after obtaining original rating data of all teachers to all indicators, original data shall be subject to the following pre-treatment: tendentious data shall be eliminated; effective data shall be kept; the lowest three marks and the highest three marks of all indicators shall be eliminated so as to obtain mean value; score value of 12 indictors for 60 students can be successively obtained so as to avoid influences of objective factors in the evaluation process as much as possible. Soft fuzzy rough set is used in the thesis to figure out weight of all evaluation indicators. Linear weighting and S of all indicators is final writing score. Classification method of evaluation result is as follows: $9 \le S \le 10$ is class 1 (excellent); $8 \le S \le 9$ is class 2 (good); $7 \le S \le 8$ is class 3 (medium); $6 \le S \le 7$ is class 4 (pass); $0 \le S \le 6$ is class 5 (fail). Part of obtained original data about English writing evaluation of students is shown in Table 1.

Table 1. Part of original data about writing evaluation

No.	X1	X2	X3	X4	...	X9	X10	X11	X12	Grade
1	9.35	9.57	9.66	9.23	...	9.19	8.75	9.33	9.08	1
2	9.54	9.48	9.75	9.61	...	9.11	8.89	9.14	9.22	1
3	9.30	9.66	9.82	9.57	...	9.23	8.63	9.30	8.98	1
:	:	:	:	:	...	:	:	:	:	:
48	6.91	7.54	6.45	5.98	...	6.06	6.55	6.05	7.16	5
49	6.99	6.96	6.08	5.42	...	5.69	6.31	5.96	7.14	5
50	7.08	6.60	6.85	6.08	...	6.02	6.17	6.20	6.97	5

3. Soft fuzzy rough set model

3.1. Soft fuzzy rough set

A thought of selecting soft threshold in soft margin (SVM) is introduced from soft fuzzy rough set theory to fuzzy-rough set theory. A kind of concept which is different from soft distance of the minimum distance method for original calculation sample is proposed.

Definition 1: a sample example x and a sample entity set $Y = \{y_1, y_2, \ldots, y_n\}$ shall be given; soft distance between x and Y is defined as

$$SD(x, Y) = \arg\max_i\{d(x, y_i) - C \times m_i\}, y_i \in Y, i = 1, 2, \ldots, n, \tag{1}$$

Where $d(x, y_j)$ is distance function between x and y_j; C is penalty factor; m_i is sample size, which meets the condition of $d(x, y_j) < d(x, y_i), j = 1, 2, ..., n$.

An example to determine soft distance is given in Fig. 1. In case sample x belongs to class 1, while other samples belong to class 2, Y shall be used to indicate the sample set. If $y1$ is considered as a noise sample and is ignored, $d(x, y_j)$ shall be $d2$. Therefore, there is required a penalty item to determine how many noise samples are required to be ignored. If a sample is ignored, $d(x, y_j)$ shall deduct C. In terms of all candidate distance $d(x, y_j)$, $d(x, y_k) = \arg\max_i\{d(x, y_i) - C \times m_i\}$ shall be taken as soft distance between x and Y, which means that distance $d(x, y_j)$ is the maximum value after punishing all ignored samples.

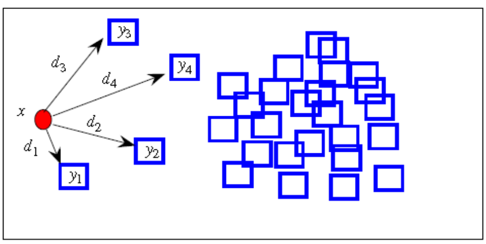

Fig. 1. Schematic diagram of soft distance

On the basis of soft distance, definition of soft fuzzy rough set is as follows:

Definition 2 U shall be considered as a nonempty domain; R is a fuzzy equivalence relation on U; F (U) is fuzzy power set of U. Upper and lower approximation of soft fuzziness for $F \in F(U)$ is defined as

$$\begin{cases} \underline{R}^S F(x) = 1 - R\left(x, \arg_y \sup_{F(y) \le F(y_L)} \{1 - R(x,y) - C \times m\}\right), \\ \overline{R}^S F(x) = R\left(x, \arg_y \sup_{F(y) \ge F(y_U)} \{1 - R(x,y) + C \times n\}\right). \end{cases} \tag{2}$$

Where,

$$\begin{cases} y_L = \arg_y \inf_{y \in U} \max\{1 - R(x,y), F(y)\}, \\ y_U = \arg_y \sup_{y \in U} \min\{R(x,y), F(y)\}, \end{cases} \tag{3}$$

Where C is a penalty factor; m is an ignored sample size at the time of calculating $\underline{R}^S F(x)$; n is an ignored sample size at the time of calculating $\overline{R}^S F(x)$.

If aggregation A is a clear set, lower approximate membership degree for soft fuzziness of sample x to A can be expressed as

$$\underline{R}^S A(x) = 1 - R(x, y_{AL}). \tag{4}$$

Where,

$$\begin{aligned} y_{AL} &= \arg_y \sup_{A(y)=0} \{1 - R(x,y) - C \times m\} = \arg_y \sup_{A(y)=0} \{d(x,y) - C \times m\} \\ &= \arg_y SD(x, U - A). \end{aligned} \tag{5}$$

Obviously, $\underline{R}^S A(x)$ is soft distance from sample x to $U - A$.

3.2. Soft fuzzy-rough classifier

On the basis of approximate definition under above-mentioned soft fuzziness, Hu Qinghua, et al. designs a robust classifier which can be used to solve the problem of classification for single label. Its principle can be summarized as: to calculate the value of lower approximate membership degree for soft fuzziness of a sample to be classified to all classes. A training sample set with k classes and a sample x to be classified shall be given. Firstly, in case x belongs to all classes, the value of lower approximate membership degree for soft fuzziness of x to k classes shall be calculated. Then, x shall be divided into a class with the largest membership degree. Its equation is:

$$class_i(x) = \arg \max_{1 \le j \le k} \{\underline{R}^S class_j(x)\}. \tag{6}$$

Where, $\underline{R}^S class_i(x)$ is lower approximate membership degree for soft fuzziness of x to $class_i$.

Algorithm description is as follows:

Input: training sample set $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ and testing sample set $X' = x'_1, x'_2, \ldots, x'_m$;

Output: *classi* of all testing sample *xi'.*

Step1: Calculate class No.;

Step2: All testing samples *xi'*$\in X'$ shall be subject to the following treatment:

(1) In terms of classes *classj*$\in Y(Y = \{y_1, y_2, ..., y_k\})$, distance from xi' and all samples in other classes shall be calculated so as to obtain candidate distance.

(2) Obtained candidate distance shall be sorted; corresponding soft distance of *classj* shall be calculated according to Equation (3).

(3) It is known from equation (4) to (6) that obtained soft distance value from xi' to samples of other classes is equal to its corresponding value of lower approximate membership degree so as to obtain lower approximate membership degree for soft fuzziness of xi' to all classes.

(4) Corresponding class label *class* shall be selected at the time of taking the maximum value for membership degree; then, it shall be returned as so to obtain class of sample xi'.

Step3: Repeat step 2 until class labels of all testing samples are obtained.

3.3. Parameter setting

Value of penalty factor $C$for soft fuzzy rough concentration is of great significance to its robustness. In terms of parameter setting, a method is given in Literature [8].

In case taking sample x as an example, credibility f for a soft hypersphere taking the sample as the center of sphere shall be given. When x is taken as the center of the sphere so as to calculate credibility of soft hypersphere, if its value is larger than or is equal to f, when the credibility is equal to f, radius difference between the soft hypersphere and the hard hypersphere shall be compared with No. of several different classes of samples in last soft hypersphere; the specific value is obtained value of C taking sample x as center of sphere. Meanwhile, lower approximate credibility for soft fuzziness is ensured. In terms of dataset with n samples, all samples shall be taken as center of sphere so as to calculate mean value of C, which can obtain value of parameter C for the database.

In terms of multiple English writing capacity dataset, parameters of all classes can be selected through translate multiple English writing capacity dataset to multiple two-category dataset. BR methods have different parameter values for different classes, which can be obtained through Equation (9). Algorithm which is subject to SFRC transformation shall take weighted mean of all parameters as value of its penalty factor C; weight is the proportion of label No. in all classes among all labels, which can be obtained in Equation (10).

Calculation equation of parameter C is as follows:

$$C_i = \frac{SD(x, Y) - HD(x, Y)}{m}, \tag{7}$$

$$C = \sum_{i=1}^{L} w_i \times C_i \,. \tag{8}$$

Where, L indicates total No. of label; wi indicates weight of class i.

Lower approximate credibility for soft fuzziness selected in the experiment in the thesis is larger than or equal to 95%, which indicates that sample error rate in soft hypersphere is smaller than 5%.

4. Experimental analysis

Experimental results are obtained by using SVM and SFRC multiple English writing capacity classifiers established by taking advantage of BR method and by using ML_SFRC classifier established by taking advantage of algorithm adaptation method through experiment in the section. All experimental results shall take mean value of four times of cross-over experiments so as to ensure stability of dataset to classification performance. On the basis of experimental results, two types of classifiers based on transformation of soft fuzzy rough set model shall be subject to comparative analysis.

4.1. Influence of No. of characteristic items on classification result

Selection of characteristic words has an impact on classification result in multiple English writing capacity evaluation. In order to obtain better classification effect, top 400, top 600, top 800, and top 1000 of characteristic words in document with high frequency are separately selected as characteristic items in the experiment. Two types of ML_SFRC_Mean classifiers in BR_SFRC and ML_SFRC which takes expectation value as limit are used so as to obtain experimental results. Classification performance can only take exact match, hamming loss, and F-measure as reference.

Table 2. Classification result of br_sfrc classifier at the time of selecting different no. of characteristic words

Characteristic No. / Indicator	400	600	800	1000
Exact Match	0.6060	0.5350	0.4780	0.4220
Hamming Loss	0.1006	0.1228	0.1387	0.1563
F-measure	0.7605	0.7058	0.6664	0.6215

Table 3. Classification result of ML_SFRC_mean classifier at the time of selecting different no. of characteristic words

Indicator \ Characteristic No.	400	600	800	1000
Exact Match	0.6000	0.5240	0.4710	0.4200
Hamming Loss	0.1003	0.1217	0.1381	0.1563
F-measure	0.7583	0.7067	0.6678	0.6223

It is shown in results in Table 2 and Table 3, best result can be obtained when No. of characteristic words are 400. Therefore, we shall used top 400 characteristic words in document frequency statistical results in subsequent experiment as characteristic item of vector space modal.

Experimental effect of multiple English writing capacity evaluation is related to selection of classification algorithms. SVM is a common classification model. SVM, BR_SFRC multiple English writing capacity classification model, and ML_SFRC multiple English writing capacity classification model established in Section 5.2 are used in the experiment in the thesis so as to classify multiple English writing capacity text.

Table 3. Classification result of ML_SFRC_mean classifier at the time of selecting different no. of characteristic words

Indicator \ Algorithm	BR	
	SVM	SFRC
ExactMatch	0.6280	0.6060
Hamming Loss	0.0537	0.1006
Accuracy	0.7957	0.7173
Precision	0.9229	0.7756
Recall	0.8154	0.7563
F-measure	0.8475	0.7605

It is shown in Table 4 that various indicators of classification result obtained by using SVM classifier is superior to that of SFRC in the way of BR method.

Table 5. Classification result of ML _SFRC multiple english writing capacity classifier at the time of selecting different threshold values under algorithm adaption method

Indicator \ Algorithm	ML _SFRC					
	90%	92%	94%	96%	98%	Mean
Exact Match	0.2990	0.3510	0.3870	0.3870	0.3090	0.6000
Hamming Loss	0.5128	0.4373	0.3588	0.2760	0.1988	0.1003
Accuracy	0.4630	0.5141	0.5577	0.5842	0.5606	0.7143
Precision	0.4811	0.5489	0.6234	0.7047	0.7702	0.7806
Recall	0.9694	0.9392	0.8924	0.8174	0.6827	0.7507
F-measure	0.5589	0.6033	0.6425	0.6704	0.6574	0.7583

It is shown in Table 5 that in ML_SFRC obtained on the basis of algorithm adaption, classification result obtained at the time of selection expectation value for limit is superior to result at the time of being given a fixed threshold value.

4.2. Generalization capacity inspection of model

In order to contrast and to explain effectiveness of classification model for soft fuzzy rough set, data in Table 1 is regarded as training sample set for soft fuzzy rough set after being subject to normalization processing. Typical single hidden layer structure is applicable to soft fuzzy rough set. No. of nodes in input layer is the same as characteristic vector dimensions of sample, which equals to 12. No. of nodes in input layer is identical to No. of classification results, which equals to 5. According to empirical equation and through several experiments, network performance is best when No. of nods in hidden layer is 15. Topology structure of soft fuzzy rough set is eventually determined to be 12-15-5. Training function traingd of standard gradient descent algorithm is used to train network. Transfer function in hidden layer is set as tansig function. Transfer function in input layer is set as purelin function. Data in Table 2 is regarded as training sample set for soft fuzzy rough set after being subject to normalization processing. Simulation result shows that classification model of soft fuzzy rough set can not correctly identify third class of test sample and its classification precision is 80%.

Classification results of samples for soft fuzzy rough set to be classified obtained through associative learning are shown in Fig. 2. It is observed that classification results are completely consistent with actual results, which shows that established classification model of soft fuzzy rough set has strong classification capacity and generalization capacity. Meanwhile, it is found in simulation experiment that if No. of evaluation indicator decreases, complexity of network model will be reduced, which will lead to decrease of classification precision and bad generalization capacity of model. Therefore, precision of model can be further improved by adding No. of evaluation indicator. As input value of soft fuzzy rough set is comparative result of single indicator value and linear weighted sum of various indicators at the time of conducting associative learning, thus evaluation standard is not only quantitative evaluation indicator but also includes qualitative evaluation indicator, which means that established soft fuzzy rough set can not only realize qualitative evaluation for writing capacity but also can realize quantitative evaluation. Compared with single soft fuzzy rough set, soft fuzzy rough set makes writing evaluation more visualized and requires 20 times of iterations for network convergence, while soft fuzzy rough set requires 2,534 times of iteration for convergence, thus soft fuzzy rough set has advantages of speed and precision in the process of writing evaluation process.

5. Conclusion

1) Firstly, the method of soft fuzzy rough set is used to establish evaluation indicator system for English writing capacity of undergraduates; then, classification evaluation model for English writing capacity of undergraduates based on soft fuzzy rough set is established so as to be compared with classification result of evaluation model for single soft fuzzy rough set. Experimental result shows that classification

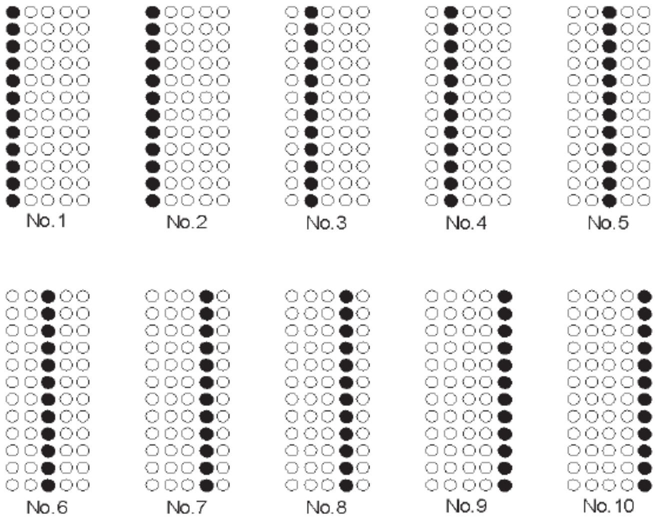

Fig. 2. Model simulation result

precision of evaluation model for soft fuzzy rough set is obviously higher than that of evaluation model for single oft fuzzy rough set; in addition, its classification results are consistent with actual results, which to a much greater extent avoid influences of subjective factors and make evaluation results more objective. Therefore, it is feasible to use soft fuzzy rough set for classification evaluation of writing capacity.

2) Associative mode is applied to soft fuzzy rough set, which makes evaluation of writing capacity more visualized with faster convergence speed. Designed model for soft fuzzy rough set is not only applicable to qualitative evaluation indicators but is also applicable to quantitative evaluation indicators of writing capacity. The more the evaluation indicators are, the reliable the evaluation results are. Training process of soft fuzzy rough set is simple without lots of training samples, which makes it has superiority under the condition that it is hard to obtain large-scale training samples.

References

[1] Ming-Qi X U: (2007) *Reflect on Culture Quality Education and College English Teaching*[J]. Journal of Heilongjiang College of Education.

[2] Pahlavan K, Krishnamurthy P and Geng Y: (2015) *Localization challenges for the emergence of the smart workd.* IEEE Access, 3(1), pp. 3058-3067

[3] Wang J, Chen S: (2009) *Language of Cross-culture English Teaching in College Students' Ideological and Moral Education*[J]. Asian Social Science, 2009(8):22.

[4] Chen J: (2011) *Cultural Teaching in College English Education in China*[J]. International Proceedings of Economics Development & Research.

[5] Lv, Z., Tek, A., Da Silva, F., Empereur-Mot, C., Chavent, M., & Baaden, M.: (2013). *Game on, science-how video game technology may help biologists tackle visualization challenges.* PloS one, 8(3), e57990.

[6] Chen M: (2003) *Cultural Education in the College English Teaching*[J]. Journal of Zhejiang Vocational & Technical Institute of Transportation.

[7] Dong X: (2012) *Culture Education in College English from the Perspective of Meme*[J]. Journal of Northeast Agricultural University.

[8] Liu Z Y, Chen Y: (2016) The Study of the Absence of Chinese Culture in College English Education[J]. Journal of Changchun Normal University.

[9] Zhang J J, Department F L, University A J: (2013) *The current situation of cross-culture education in college English and possible solutions*[J]. Journal of Jiamusi Education Institute, 6(6):261-270.

[10] Lv, Z., Halawani, A., Feng, S., Li, H., & Réhman, S. U.: (2014). *Multimodal hand and foot gesture interaction for handheld devices.* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 10.

[11] Wang F: (2008) *How to Keep Linguistic and Cultural Ecology in the Teaching of College English*[J]. Journal of Mudanjiang College of Education.

[12] Weisen Pan, Shizhan Chen, Zhiyong Feng: (2013) *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, 7(2): 675-681.

[13] Yingyue Zhang, Jennifer W. Chan, Alysha Moretti, and Kathryn E. Uhrich: (2015) *Designing Polymers with Sugar-based Advantages for Bioactive Delivery Applications*, Journal of Controlled Release, 219, 355-368.

[14] Yingyue Zhang, Qi Li, William J. Welsh, Prabhas V. Moghe, and Kathryn E. Uhrich: (2016) *Micellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity*, Biomaterials, 84, 230-240.

[15] Tennant S: (2013) *Teaching English as a Foreign Language by Exploring the Art and Culture of the Students' Own Country.*[J]. English Teacher An International Journal, 5:35-50.

[16] Shu-Xia M A: (2009) *General Discussion on Disposition of the Cultural Education in College English Teaching*[J]. Journal of North China Institute of Aerospace Engineering.

[17] Jennifer W. Chan, Yingyue Zhang, and Kathryn E. Uhrich: (2015) *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, 26(7), 1359-1369.

Received May 7, 2017