

УДК 004.93

Topic Categorization Based on Collectives of Term Weighting Methods for Natural Language Call Routing

Roman B. Sergienko*

Muhammad Shan†

Wolfgang Minker‡

Institute of Telecommunication Engineering

Ulm University

Albert-Einstein-Allee, 43, Ulm, 89081

Germany

Eugene S. Semenin§

Informatics and Telecommunications Institute

Siberian State Aerospace University

Krasnoyarskiy Rabochiy, 31, Krasnoyarsk, 660037

Russia

Received 26.12.2015, received in revised form 11.01.2016, accepted 20.02.2016

Natural language call routing is an important data analysis problem which can be applied in different domains including airspace industry. This paper presents the investigation of collectives of term weighting methods for natural language call routing based on text classification. The main idea is that collectives of different term weighting methods can provide classification effectiveness improvement with the same classification algorithm. Seven different unsupervised and supervised term weighting methods were tested and compared with each other for classification with k -NN. After that different combinations of term weighting methods were formed as collectives. Two approaches for the handling of the collectives were considered: the meta-classifier based on the rule induction and the majority vote procedure. The numerical experiments have shown that the best result is provided with the vote of all seven different term weighting methods. This combination provides a significant increasing of classification effectiveness in comparison with the most effective term weighting methods.

Keywords: natural language call routing, text classification, term weighting.

DOI: 10.17516/1997-1397-2016-9-2-235-245.

Introduction

Natural language call routing is an important problem in the design of modern automatic call services and the solving of this problem could lead to improvement of the call service [1]. Generally natural language call routing can be considered as two different problems. The first one is speech recognition of calls and the second one is topic categorization of users' utterances for further routing. Topic categorization of users' utterances can be also useful for multi-domain

*roman.sergienko@uni-ulm.de

†muhammad.shan@uni-ulm.de

‡wolfgang.minker@uni-ulm.de

§eugenesemenkin@yandex.ru

spoken dialogue system design [2]. In this work we treat call routing as an example of a text classification application

In the vector space model [3] text classification is considered as a machine learning problem. The complexity of text categorization with a vector space model is compounded by the need to extract the numerical data from text information before applying machine learning algorithms. Therefore, text classification consists of two parts: text preprocessing and classification algorithm application using the obtained numerical data.

Some text preprocessing methods are based on the idea that the category of the document depends on the words or phrases from this document. One of the most popular models for document representation is the "bag-of-words" model, in which the word order is ignored. The simplest approach for the "bag-of-words" model application is to take each word of the document as a binary coordinate and the dimensionality of the feature space will be the number of words in our dictionary. More advanced approaches are term weighting methods. There exist different unsupervised and supervised term weighting methods. The most well-known unsupervised term weighting method is TF-IDF [4]. The following supervised term weighting methods are also considered in the paper: Gain Ratio (GR) [5], Confident Weights (CW) [6], Term Second Moment (TM2) [7], Relevance Frequency (RF) [8], Term Relevance Ratio (TRR) [9], and Novel Term Weighting (NTW) [10]; these methods involve information about the classes of the documents.

After text preprocessing, machine learning algorithms are applied for the classification, such as k -NN, support vector machine (SVM) [11], Rocchio classifier [12] etc. There exist a lot of approaches based on collectives of different classification algorithms, such as majority vote, bagging [13], and boosting [14]. The collectives of classification algorithms can demonstrate better classification effectiveness than the best algorithm in the collective; it was also demonstrated for text classification [15]. Therefore, meta-classification is very popular in the field of machine learning. In our paper we propose an idea that collectives of different term weighting methods can also provide text classification effectiveness improvement even with the same classification algorithm. Two approaches for the handling of the collectives are considered: the meta-classifier based on the rule induction [16] and the majority vote procedure.

This paper is organized as follows: In Section 1, we describe the problem and the database. Section 2 describes the considered term weighting methods. The results of feature selection and the comparison of the term weighting methods are presented in Section 3. Section 4 reports on the experimental results of the collectives of term weighting methods. Finally, we provide concluding remarks and directions for future investigations in Section 5.

1. Corpus description

The data for testing and evaluation consists of 292,156 user utterances recorded in English language from caller interactions with commercial automated agents. Utterances are short and contain only one phrase for further routing. The database contains calls in textual format after speech recognition. The database is provided by the company *Speech Cycle* (New York, USA). Utterances from this database are manually labelled by experts and divided into 20 classes (such as appointments, operator, bill, internet, phone and technical support). One of them is a special class TE-NOMATCH which includes utterances that cannot be put into another class or can be put into more than one class.

The database contains 45 unclassified calls and they were removed. The database contains also 23,561 empty calls without any words. These calls were placed in the class TE-NOMATCH

automatically and they were also removed from the database. As a rule, the calls are short in the database; many of them contain only one or two words. So there are a lot of duplicated utterances in the database and utterance duplicates were removed. After that the database contains 24,458 unique non-empty classified calls. The average length of an utterance is 4.66 words, the maximal length is 19 words. The corpus is unbalanced. The largest class contains 27.05% and the smallest one contains 0.04% of the unique calls. The classes and their distribution are presented in Tab. 1.

Table 1. The distribution of the classes

Class	Percentage
TE-NOMATCH	13.95
serviceVague	1.06
appointments	3.60
none	2.05
cancelService	0.40
idk	0.20
orders	6.32
UpForDiscussion_ Complaint	0.04
operator	8.15
techSupport	24.87
bill	27.05
internet	1.91
phone	1.00
techSupport_internet	0.24
techSupport_phone	0.24
techSupport_video	0.76
video	1.96
changeService	3.79
UpForDiscussion_no_audio	2.32
UpForDiscussion_AskedToCall	0.09

For statistical analysis we performed 20 different separations of the database into training and test samples randomly. The train samples contain 90% of the calls and the test samples contain 10% of the calls. For each training sample we have designed a dictionary of unique words which appear in the training sample. The size of the dictionary varies from 3,275 to 3,329 words for different separations. Due to the appropriate size of the dictionary and short utterances we did not perform stopwords filtering and stemming.

2. Term weighting methods

As a rule, term weighting is a multiplication of two parts: the part based on the term frequency in a document (TF) and the part based on the term frequency in the whole training database. The TF-part is fixed for all considered term weighting methods and is calculated as following:

$$TF_{ij} = \log(tf_{ij} + 1); \quad tf_{ij} = \frac{n_{ij}}{N_j},$$

where n_{ij} is the number of times the i^{th} word occurs in the j^{th} document, N_j is the document size (number of words in the document).

The second part of the term weighting is calculated once for each word from the dictionary and does not depend on an utterance for classification. We consider seven different methods for the calculation of the second part of term weighting.

2.1. Inverse document frequency (IDF)

IDF is a well-known unsupervised term weighting method which was proposed in [4]. There are some modifications of IDF and we use the most popular one:

$$idf_i = \log \frac{|D|}{n_i},$$

where $|D|$ is the number of documents in the training set and n_i is the number of documents that have the i^{th} word.

2.2. Gain ratio (GR)

Gain Ratio (GR) is mainly used in term selection [17], but in [5] it was shown that it could also be used for weighting terms. The definition of GR is as follows:

$$GR(t_i, c_j) = \frac{\sum_{c \in \{c_j, \bar{c}_j\}} \sum_{t \in \{t_j, \bar{t}_j\}} M(t, c)}{-\sum_{c \in \{c_j, \bar{c}_j\}} P(c) \cdot \log P(c)},$$

$$M(t, c) = P(t, c) \cdot \log \frac{P(t, c)}{P(t) \cdot P(c)},$$

where $P(t, c)$ is the relative frequency that a document contains the term t and belongs to the category c ; $P(t)$ is the relative frequency that a document contains the term t and $P(c)$ is the relative frequency that a document belongs to category c . Then, the weight of the term t_i is the max value between all categories as follows:

$$GR(t_i) = \max_{c_j \in C} GR(t_i, c_j),$$

where C is a set of all classes.

2.3. Confident weights (CW)

This supervised term weighting approach has been proposed in [6]. Firstly, the proportion of documents containing term t is defined as the Wilson proportion estimate $p(x, n)$ by the following equation:

$$p(x, n) = \frac{x + 0.5z_{\alpha/2}^2}{n + z_{\alpha/2}^2},$$

where x is the number of documents containing the term t in the given corpus, n is the number of documents in the corpus and $\Phi(z_{\alpha/2}) = \alpha/2$, where Φ is the t -distribution (Student's law) when $n < 30$ and the normal distribution when $n \geq 30$.

In this work $\alpha = 0.95$ and $0.5z_{\alpha/2}^2 = 1.96$ (as recommended by the authors of the method). For each term t and each class c two functions $p_{pos}(x, n)$ and $p_{neg}(x, n)$ are calculated. For $p_{pos}(x, n)$ x is the number of documents which belong to the class c and have term t ; n is the number of documents which belong to the class c . For $p_{neg}(x, n)$ x is the number of documents which have the term t but do not belong to the class c ; n is the number of documents which do not belong to the class c .

The confidence interval (p^-, p^+) at 0.95 is calculated using the following equation:

$$M = 0,5z_{\alpha/2}^2 \sqrt{\frac{p(1-p)}{n + z_{\alpha/2}^2}}; \quad p^- = p - M; \quad p^+ = p + M.$$

The strength of the term t in the category c is defined as the follows:

$$str(t, c) = \begin{cases} \log_2 \frac{2p_{pos}^-}{p_{pos}^- + p_{neg}^+}, & \text{if } p_{pos}^- > p_{neg}^+, \\ 0, & \text{otherwise.} \end{cases}$$

The maximum strength (Maxstr) of the term t_i is calculated as follows:

$$Maxstr(t_i) = \max_{c_j \in C} str(t_i, c_j)^2.$$

2.4. Term second moment (TM2)

This supervised term weighting method was proposed in [7]. Let $P(c_j|t)$ be the empirical estimation of the probability that a document belongs to the category c_j with the condition that the document contains the term t ; $P(c_j)$ is the empirical estimation of the probability that a document belongs to the category c_j without any conditions. The idea is the following: the more $P(c_j|t)$ is different from $P(c_j)$, the more important the term t_i is. Therefore, we can calculate the term weight as the following:

$$TM2(t_i) = \sum_{j=1}^{|C|} (P(c_j|t) - P(c_j))^2,$$

where C is a set of all classes.

2.5. Relevance frequency (RF)

The RF term weighting method was proposed in [8] and is calculated as the following:

$$rf(t_i, c_j) = \log_2 \left(2 + \frac{a_j}{\max\{1, \bar{a}_j\}} \right), \quad rf(t_i) = \max_{c_j \in C} rf(t_i, c_j),$$

where a_j is the number of documents of the category c_j which contain the term t_i and \bar{a}_j is the number of documents of all the other categories which also contain this term.

2.6. Term relevance ratio (TRR)

The TRR method [9] uses tf weights and it is calculated as the following:

$$TRR(t_i, c_j) = \log_2 \left(2 + \frac{P(t_i|c_j)}{P(t_i|\bar{c}_j)} \right),$$

$$P(t_i|c) = \frac{\sum_{k=1}^{|T_c|} tf_{ik}}{\sum_{l=1}^{|V|} \sum_{k=1}^{|T_c|} tf_{lk}},$$

$$TRR(t_i) = \max_{c_j \in C} TRR(t_i, c_j),$$

where c_j is a class of the document, \bar{c}_j is all of the other classes of c_j , V is the vocabulary of the training data and T_c is the document set of the class c .

2.7. Novel term weighting (NTW)

This method was proposed in [10]. The details of the procedure are the following. Let L be the number of classes; n_i is the number of documents which belong to the i_{th} class; N_{ij} is the number of occurrences of the j_{th} word in all documents from the i_{th} class. $T_{ij} = N_{ij}/n_i$ is the relative frequency of occurrences of the j_{th} word in the i_{th} class; $R_j = \max_i T_{ij}$; $S_j = \arg \max_i T_{ij}$ is the class which we assign to the j_{th} word. The term relevance C_j is calculated by the following:

$$C_j = \frac{1}{\sum_{i=1}^L T_{ij}} \cdot \left(R_j - \frac{1}{L-1} \cdot \sum_{i=1, i \neq S_j}^L T_{ij} \right).$$

3. Feature selection and comparison of term weighting methods

K -NN with distance weighting was applied as the classification algorithm with all seven term weighting methods. Some investigations [18, 19] have shown effectiveness of the k -NN algorithm for text classification. We have varied k from 1 to 15. The classification criterion is the macro F -score [20].

Term weighting methods provide a natural feature selection method; it is possible to ignore terms with the lowest weights. Therefore, at first we performed the investigation of the feature selection based on term weights. For RF, TM2, and TRR methods we decreased the dictionary size from 100% to 10% with the interval equals 10. It means deleting the corresponding number of the terms with the lowest weights. IDF and NTW provide getting a lot of terms with the equal highest value. For IDF the highest weight means that the term occurs only in one document from the training sample, for NTW it means that the term occurs only in documents of one class. Therefore, for these two methods we used different constraints for the value of weights; the predefined percentage of the dictionary size is not appropriate for NTW and IDF. CW and GR provide getting a lot of terms with zero weights; it means that these two methods provide feature selection automatically. For our problem we have 43.5% of the dictionary as terms with non-zero weights for GR and 20.4% for CW on the average. We also decreased the size of the dictionary for CW and GR.

At first we performed averaging by 20 different test samples for each value of k and after that we chose the best F -score by k . The common results of the feature selection investigations are presented in Fig. 1.

The main conclusion of the investigation is that the best F -score for each term weighting method provides with using the maximal number of the terms. The reason of such results is that the database contains short calls. The maximal length of a call is not more than 20 words. In this situation every word in a call can be useful for classification. Feature selection can increase classification effectiveness in the case of large documents with the large or redundant dictionary [21, 22].

Tab. 2 demonstrates ranking of the term weighting methods with using the maximal number of the terms. Statistical significance of the ranking is provided with t -test for 20 test samples. The confidence probability equals 0.99.

Therefore, we have three best term weighting methods (TRR, TM2, and NTW) which have no significant difference between each other. sectionCollectives of TermWeighting Methods The

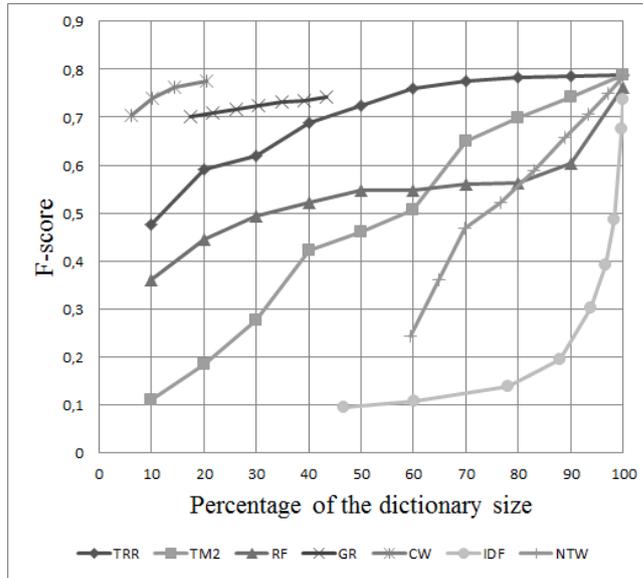


Fig. 1. The results of feature selection for the term weighting methods

Table 2. Ranking of the term weighting methods

Rank	Method	Max F -score by k	The best k
1-3	TRR	0.788	2.45
1-3	TM2	0.788	2.55
1-3	NTW	0.787	1.9
4	CW	0.774	2.25
5	RF	0.763	4.15
6	GR	0.743	1.4
7	IDF	0.737	3.2

next stage is the investigation of the collectives of term weighting methods. We have designed the collectives with different numbers of included methods from 7 to 2, with consistent exception of the worst methods. Therefore, the collective 7 contains all 7 term weighting methods; the collective 6 contains TRR, TM2, NTW, CW, RF, and GR; the collective 5 contains TRR, TM2, NTW, CW, and RF; the collective 4 contains TRR, TM2, NTW, and CW; the collective 3 contains TRR, TM2, and NTW; the collective 2 contains TRR and TM2 despite the fact that there is no statistically significant difference between TRR, TM2 and NTW. For the collective design we use the results which are obtained with the same value k of the k -NN algorithm. It means that any feature of the collectives depends on only cooperation between term weighting methods. This allows us to verify the hypothesis that the collectives of the term weighting methods can improve classification effectiveness even with the same classification algorithm with the same settings. Therefore, we designed $15 \times 6 = 90$ collectives for each division of the database.

We propose two different approaches for the handling of the collectives. The first one is a meta-classifier based on the rule induction method [16]. After classification with k -NN was

performed with all term weighting methods we put the classification results (the predicted class by each term weighting method) as categorical features for the rule induction. We propose two schemas for rule induction learning. With the first way we perform rule induction learning for the whole training sample but in this case samples for k -NN learning and rule induction learning are the same. With the second way we divided the training sample into two parts in the proportion 7:2. The first division is used for k -NN learning and the second one (validating sample) is used for rule induction learning. *RapidMiner* was used as software for rule induction application.

Tabs. 3–5 contain the results for different approaches for the collective handling. The scheme of getting final results is the same as in the previous stage of the investigation. The t -test values for the comparison with three best term weighting methods are also presented.

Table 3. Rule induction only with the training sample

Combination	F -score	t -test		
		TRR	TM2	NTW
Coll.2	0.793	0.000	0.000	0.003
Coll.3	0.793	0.001	0.000	0.000
Coll.4	0.792	0.002	0.006	0.074
Coll.5	0.789	0.320	0.541	0.381
Coll.6	0.787	0.815	0.774	0.998
Coll.7	0.789	0.169	0.370	0.124

Table 4. Rule induction with the training and validating samples

Combination	F -score	t -test		
		TRR	TM2	NTW
Coll.2	0.781	0.081	0.055	0.105
Coll.3	0.783	0.154	0.113	0.238
Coll.4	0.785	0.382	0.333	0.437
Coll.5	0.791	0.026	0.047	0.035
Coll.6	0.791	0.080	0.182	0.128
Coll.7	0.788	0.811	0.910	0.760

Table 5. Majority vote

Combination	F -score	t -test		
		TRR	TM2	NTW
Coll.3	0.800	0.000	0.002	0.000
Coll.4	0.800	0.000	0.000	0.000
Coll.5	0.803	0.000	0.000	0.000
Coll.6	0.803	0.000	0.000	0.000
Coll.7	0.805	0.000	0.000	0.000

The results of the numerical experiments provide the following conclusions. Rule induction

only with the training sample provides a statistically significant increment of F -score with the collective 2 and the collective 3 in comparison with the best term weighting methods (confidence probability 0.99). The increment equals 0.005 on the average. Rule induction with the training and validating samples provides worse results. Only for the collective 5 we can observe a significant increment of F -score with the confidence probability 0.95.

The best results are obtained with the majority vote. All combinations of the term weighting methods provide a significant increment of the F -score with the confidence probability 0.99 (without the collective 3 we can put the confidence probability as 0.999). The best result is achieved with the collective which contains all seven term weighting methods; the average increment of the F -score equals 0.017. The best collective for the majority vote exceeds the best collectives for the rule induction with the confidence probability 0.999. The comparison of the best collective with other combinations for the majority vote is illustrated in Tab. 6.

Table 6. The comparison of the best combination with others for the majority vote

	Coll.3	Coll.4	Coll.5	Coll.6
t-test	0.000	0.000	0.024	0.111

Tab. 6 demonstrates that the statistically significant difference of the best collective increases with the decrement of the collective size.

The best result is obtained with an average $k = 2.35$.

Conclusions

In this work a text classification problem for natural language call routing was considered. Seven different term weighting methods (IDF, GR, CW, TM2, RF, TRR, and NTW) were applied for text preprocessing. K -NN with distance weighting was applied as the classification algorithm. The numerical experiments showed that the best results are obtained by using the maximal number of the terms for all methods. Three best term weighting methods (TRR, TM2, and NTW) were determined with the comparison of the methods. After that two approaches for the term weighting method collective handling were proposed. The first one is a meta-classifier based on the rule induction and the second one is the majority vote. The best result is obtained with the majority vote with all seven term weighting methods. The increment of F -score equals 0.017. Therefore, we confirmed the hypothesis that collectives of different term weighting methods provide text classification effectiveness improvement with the same classification algorithm with the same settings.

As future directions we propose the following:

- Numerical experiments for different text classification problems.
- Application of different classification algorithms, such as Rocchio classifier and SVM.
- Optimization of the method weights for the vote procedure.

References

- [1] B.Suhm, J.Bers, D.McCarthy, B.Freeman, D.Getty, K.Godfrey, P.Peterson, A Comparative Study of Speech in the Call Center: Natural Language Call Routing vs. Touch-Tone Menus, Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2002, 283–290.

- [2] C.Lee, S.Jung, S.Kim, G.Lee, Example-Based Dialog Modeling for Practical Multi-Domain Dialog System, *Speech Communication*, **51**(2009), no. 5, 466–484.
- [3] F.Sebastiani, Machine Learning in Automated Text Categorization, *ACM computing surveys (CSUR)*, **34**(2002), no. 1, 1–47.
- [4] G.Salton, Ch.Buckley, Term-Weighting Approaches in Automatic Text Retrieval, *Information processing & management*, **24**(1988), no. 5, 513–523.
- [5] F.Debole, F.Sebastiani, Supervised Term Weighting for Automated Text Categorization, Text mining and its applications, 2004, 81–97
- [6] P.Soucy, G.Mineau, Beyond TFIDF Weighting for Text Categorization in the Vector Space Model, *IJCAI*, **5**(2005), 1130–1135.
- [7] H.Xu, Ch.Li, A Novel Term Weighting Scheme for Automated Text Categorization, Intelligent Systems Design and Applications, ISDA 2007, Seventh International Conference on, 2007, 759–764.
- [8] M.Lan, Ch.Tan, J.Su, Y.Lu, Supervised and Traditional Term Weighting Methods for Automatic Text Categorization, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **31**(2009), no. 4, 721–735.
- [9] Y.Ko, A Study of Term Weighting Schemes Using Class Information for Text Classification, Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012, 1029–1030.
- [10] T.Gasanova, R.Sergienko, E.Semenkin, W.Minker, Dimension reduction with coevolutionary genetic algorithm for text classification, ICINCO 2014, Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics, 2014, 215–222.
- [11] T.Joachims, Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms, 2002.
- [12] T.Joachims, A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization, 1996.
- [13] L.Breiman, Bagging Predictors, *Machine learning*, **24**(1996), no. 2, 123–140.
- [14] R.Schapire, Y.Singer, BoosTexter: A Boosting-Based System for Text Categorization, *Machine learning*, **39**(2000), no. 2–3, 135–168.
- [15] D.Morariu, L.Vintan, V.Tresp, Meta-Classification Using SVM Classifiers for Text Documents, *Intl. Jrnl. of Applied Mathematics and Computer Sciences*, **1**(2005), no. 1.
- [16] W.Cohen, Fast Effective Rule Induction, *Proceedings of the twelfth international conference on machine learning*, 1995, 115–123.
- [17] Y.Yang and J.Pedersen, A Comparative Study on Feature Selection in Text Categorization, *ICML*, **97**(1997), 412–420.
- [18] E.Han, G.Karypis, V.Kumar, Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification, 2001.

- [19] B.Baharudin, L.Lee, K.Khan, A Review of Machine Learning Algorithms for Text-Documents Classification, *Journal of advances in information technology*, 1(2010), no. 1, 4–20.
- [20] C.Goutte, E.Gaussier, A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation, Proceedings of the 27th European conference on Advances in Information Retrieval Research, Springer-Verlag Berlin, 2005, 345–359.
- [21] M.Rogati, Y.Yiming, High-Performing Feature Selection for Text Classification, Proceedings of the eleventh international conference on Information and knowledge management, 2002, 659–661.
- [22] E.Gabrilovich, Sh.Markovitch, Text Categorization with Many Redundant Features: Using Aggressive Feature Selection to Make SVMs Competitive with C4. 5, Proceedings of the twenty-first international conference on Machine learning, 2004, 41.

Определение темы для маршрутизации вызовов на естественном языке на основе коллективов методов взвешивания термов

**Роман Б. Сергиенко
Мухаммад Шан
Вольфганг Минкер**

Институт телекоммуникации и инжиниринга
Университет Ульма
Аллея Альберта Эйнштейна, 43, Ульм, 89081
Германия

Евгений С. Семенкин

Институт информатики и телекоммуникаций
Сибирский государственный аэрокосмический университет
Красноярский рабочий, 31, Красноярск, 660037
Россия

Маршрутизация вызовов на естественном языке – актуальная задача анализа данных, которая может найти применение в различных областях, включая аэрокосмическую индустрию. В статье представлено исследование коллективов методов взвешивания термов для маршрутизации вызовов на естественном языке на основе классификации текста. Основная идея предлагаемого подхода заключается в том, что коллективы методов взвешивания термов могут обеспечить повышение эффективности классификации при использовании одного и того же алгоритма классификации. Семь различных методов взвешивания термов были протестированы и сравнены между собой с использованием метода ближайших соседей в качестве алгоритма классификации. После этого были сформированы различные комбинации методов взвешивания термов для дальнейшего использования в коллективных решающих правилах. Рассмотрено два подхода для формирования коллективных решающих правил: мета-классификатор на основе индукции правил и голосование простым большинством. Численные исследования показали, что наилучший результат достигается при включении всех семи рассматриваемых методов взвешивания термов в коллективное решающее правило на основе голосования простым большинством. Такая комбинация обеспечивает статистически значимое улучшение эффективности классификации в сравнении с лучшим по эффективности отдельным методом взвешивания термов.

Ключевые слова: маршрутизация вызовов на естественном языке, классификация текста, взвешивание термов.