Feature Extraction with TF-IDF and Game-Theoretic Shadowed Sets

Yan Zhang[1(✉)], Yue Zhou[1], and JingTao Yao[2]

[1] School of Computer Science and Engineering, California State University,
San Bernardino, CA, USA
{Yan.Zhang,Yue.Zhou}@csusb.edu
[2] Department of Computer Science, University of Regina, Regina, SK, Canada
jtyao@cs.uregina.ca

Abstract. TF-IDF is one of the most commonly used weighting metrics for measuring the relationship of words to documents. It is widely used for word feature extraction. In many research and applications, the thresholds of TF-IDF for selecting relevant words are only based on trial or experiences. Some cut-off strategies have been proposed in which the thresholds are selected based on Zipf's law or feedbacks from model performances. However, the existing approaches are restricted in specific domains or tasks, and they ignore the imbalance of the number of representative words in different categories of documents. To address these issues, we apply game-theoretic shadowed set model to select the word features given TF-IDF information. Game-theoretic shadowed sets determine the thresholds of TF-IDF using game theory and repetition learning mechanism. Experimental results on real world news category dataset show that our model not only outperforms all baseline cut-off approaches, but also speeds up the classification algorithms.

Keywords: Feature extraction · TF-IDF · Text classification · Game-theoretic shadowed sets

1 Introduction

Term Frequency-Inverse Document Frequency, or TF-IDF, is one of the most commonly used weighting metrics for measuring the relationship of words and documents. It has been applied to word feature extraction for text categorization or other NLP tasks. The words with higher TF-IDF weights are regarded as more representative and are kept while the ones with lower weights are less representative and are discarded. An appropriate selection of word features is able to speed up the information retrieval process while preserving the model performance. However, for many works, the cutoff values or the thresholds of TF-IDF for selecting relevant words is only based on guess or experience [6,11,12]. Zipf's law is used to select words whose IDF exceeds a certain value [10]; Lopes et al. [9] proposed a cut-off policy by balancing the precision and recall from the model performance.

Despite their success, those cut-off policies have certain issues. The cut-off policy that the number of words to keep is determined by looking at the precision and recall score of the model can be restricted in specific domains or tasks. In addition, the number of relevant words may vary in different categories of documents in certain domains. For instance, there exists an imbalance between the number of representative positive words and negative words in many sentiment classification tasks. Thus, a cut-off policy that is able to capture such imbalance is needed.

To address these issues, we employ game-theoretic shadowed sets (GTSS) to determine the thresholds for feature extraction. GTSS, proposed by Yao and Zhang, is a recent promising model for decision making in the shadowed set context [22]. We calculate the difference of TF-IDF for each word between documents as the measurement of relevance, and then use GTSS to derive the asymmetric thresholds for word extraction. GTSS model aims to determine and explain the thresholds from a tradeoff perspective. The words with the difference of TF-IDF less than β or greater than α are selected. We regard the words whose difference of TF-IDF are between α and β as neutral. These words can be safely removed since they can not contribute much in text classification.

The results of our experiments on a real world news category dataset show that our model achieves significant improvement as compared with different TF-IDF based cut-off policies. In addition, we show our model can achieve comparable performance as compared to the model using all words' TF-IDF as features, while greatly speed up the classification algorithms.

2   Related Work

TF-IDF is the most commonly used weighting metrics for measuring the relationship of words and documents. By considering the word or term frequency (TF) in the document as well as how unique or infrequent (IDF) a word in the whole corpus, TF-IDF assigns higher values to topic representative words while devalues the common words. There are many variations of TF-IDF [19,20]. In our experiments, we use the basic form of TF-IDF and follow the notation given in [7]. The TF-IDF weighted value $w_{t,d}$ for the word t in the document d is thus defined as:

$$w_{t,d} = tf_{t,d} \times \log_{10}(\frac{N}{df_t}) \tag{1}$$

where $tf_{t,d}$ is the frequency of word t in the document d, N is the total number of documents in the collection, and df_t is the number of documents where word t occurs in.

TF-IDF measures how relevant a word to a certain category of documents, and it is widely used to extract the most representative words as features for text classification or other NLP tasks. The extraction is often done by selecting top n words with the largest TF-IDF scores or setting a threshold below which the words are regarded as irrelevant and discarded. But an issue arises

about how to choose such cut-off point or threshold so as to preserve the most relevant words. Many works choose such threshold only based on trial or experience [6,11,12,23]. On the other hand, some approaches address the issue. Zipf's law is used to select the words whose IDF exceeds a certain value in order to speed up information retrieval algorithms [4,10]. Lopes et al. [9] proposed a cut-off policy which determines the number of words to keep by balancing precision and recall in downstream tasks. However, such cut-off points should not be backward induced by the performance of downstream task; rather, the thresholds should be derived before feeding the extracted words to the classifier to speed up the model without reducing the performance. In addition, for certain domains, the number of relevant words may vary in different categories of documents. For instance, the number of words relevant to positive articles and the number of words relevant to negative articles are often imbalanced in many sentiment analysis tasks. Therefore, the cut-off points or thresholds may also vary in different categories. By observing these drawbacks, we attempt to find asymmetric thresholds of TF-IDF for feature extraction by using game-theoretic shadowed sets.

3   Methodology

In this section, we will introduce our model in details. Our model aims to find an approach of extracting relevant words and discarding less relevant words based on TF-IDF information so as to speed up learning algorithms while preserving the model performance. We first calculate the difference of TF-IDF for each word between documents as one single score to measure the degree of relevance, and then use game-theoretic shadowed sets to derive the asymmetric thresholds for words extraction.

3.1   TF-IDF Difference as Relevance Measurement

Consider a binary text classification task with a set of two document classes $C = \{c_1, c_2\}$. For each word t, we calculate the difference of TF-IDF weighted value between document c_1 and c_2 as:

$$DW_t = w_{t,c_1} - w_{t,c_2} = (tf_{t,c_1} - tf_{t,c_2}) \times \log_{10}(\frac{N}{df_t}). \qquad (2)$$

We use DW_t to measure how relevant or representative the word t is to the document classes. The greater the magnitude of DW_t, the more representative the word t is to distinguish the document categories. A large positive value of DW_t indicates that a word t is not common word and more relevant to the document c_1, while a significant negative value shows the word t is representative to document c_2. If DW_t is closed to zero, then we regard the word t as neutral. In the

next section, we will choose the cut-off thresholds for selecting the most representative word features by using the Game-theoretic Shadowed Sets method. For convenience, we here normalize the DW_t with min-max linear transformation.

3.2 Shadowed Sets

A shadowed set S in the universe U maps the membership grades of the objects in U to a set $\{0, [0,1], 1\}$, i.e., $S : U \rightarrow \{0, [0,1], 1\}$ a pair of thresholds (α, β) while $0 \leq \beta < \alpha \leq 1$ [13]. Shadowed sets are viewed as three-valued constructs induced by fuzzy sets, in which three values are interpreted as full membership, full exclusion, and uncertain membership [15]. Shadowed sets can capture the essence of fuzzy sets at the same time reducing the uncertainty from the unit interval to a shadowed region [15]. The shadowed set based three-value approximations are defined as a mapping from the universe U to a three-value set $\{0, \sigma, 1\}$, if a single value σ ($0 \leq \sigma \leq 1$) is chosen to replace the unit interval $[0,1]$ in the shadowed sets, that is [3],

$$T_{(\alpha,\beta)}(\mu_A(x)) = \begin{cases} 1, & \mu_A(x) \geq \alpha, \\ 0, & \mu_A(x) \leq \beta, \\ \sigma, & \beta < \mu_A(x) < \alpha. \end{cases} \tag{3}$$

The membership grade $\mu_A(x)$ of an object x indicates the degree of the object x belonging to the concept A or the degree of the concept A applicable to x [21].

Given a concept A and an element x in the universe U, if the membership grade of this element $\mu_A(x)$ is greater than or equal to α, the element x would be considered to belong to the concept A. An elevation operation elevates the membership grade $\mu_A(x)$ to 1 which represents a full membership grade [14]. If the membership grade $\mu_A(x)$ is less than or equal to β, the element x would not be considered to belong to the concept A. An reduction operation reduces the membership grade $\mu_A(x)$ to 0 which represents a null membership grade [14]. If the membership grade $\mu_A(x)$ is between α and β, the element x would be put in a shadowed area, which means it is hard to determine if x belongs to concept A. $\mu_A(x)$ is mapped to σ which represents the highest uncertainty, that is we are far more confident about including an element or excluding an element in the concept A. The membership grades between α and σ are reduced to σ; The membership grades between σ and β are elevated to σ. We get two elevated areas, $E_1(\mu_A)$ and $E_\sigma(\mu_A)$, and two reduced areas, $R_0(\mu_A)$ and $R_\sigma(\mu_A)$ shown as the dotted areas and lined areas in Fig. 1 (a). Figure 1 (b) shows the shadowed set based three-value approximation after applying the elevation and reduction operations on all membership grades.

The vagueness is localized in the shadowed area as opposed to fuzzy sets where the vagueness is spread across the entire universe [5,16].
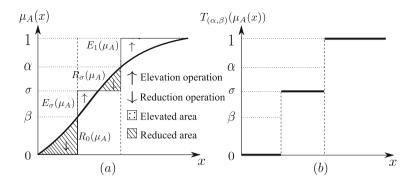
Fig. 1. A shadowed set based three-value approximation

3.3   Error Analysis

Shadowed set based three-value approximations use two operations, the elevation and reduction operations, to approximate the membership grades $\mu_A(x)$ to a three-value set $\{0, \sigma, 1\}$. Given an element x with the membership grade $\mu_A(x)$, the elevation operation changes the membership grade $\mu_A(x)$ to 1 or σ. The reduction operation changes the membership grade $\mu_A(x)$ to 0 or σ. These two operations change the original membership grades and produce the elevated and reduced areas which show the difference between the original membership grades and the mapped values 1, σ, and 0, as shown in Fig. 1(a). These areas can be viewed as the elevation and the reduction errors, respectively. The elevation operation produces two elevation errors $E_1(\mu_A)$ and $E_\sigma(\mu_A)$; the reduction operation produces two reduction errors $R_0(\mu_A)$ and $R_\sigma(\mu_A)$, that is

- The elevation error E_1 is produced when the membership grade $\mu_A(x)$ is greater than or equal to α (i.e., $\mu_A(x) \geq \alpha$), and the elevation operation elevates $\mu_A(x)$ to 1. We have $E_1(\mu_A(x)) = 1 - \mu_A(x)$.
- The elevation error E_σ is produced when $\beta < \mu_A(x) < \sigma$, and the elevation operation elevates $\mu_A(x)$ to σ. We have $E_\sigma(\mu_A(x)) = \sigma - \mu_A(x)$.
- The reduction error R_0 is produced when $\mu_A(x) \leq \beta$, and the reduction operation reduces $\mu_A(x)$ to 0. We have $R_0(\mu_A(x)) = \mu_A(x)$.
- The reduction error R_σ is produced when $\sigma < \mu_A(x) < \alpha$, and the reduction operation reduces $\mu_A(x)$ to σ. We have $R_\sigma(\mu_A(x)) = \mu_A(x) - \sigma$.

The elevation errors $E_{(\alpha,\beta)}(\mu_A)$ is the sum of two elevation errors produced by elevation operation. The total reduction errors $R_{(\alpha,\beta)}(\mu_A)$ is the sum of two reduction errors produced by reduction operation. For discrete universe of discourse, we have a collection of membership values. The total elevation and reduction errors are calculated as [22],

$$E_{(\alpha,\beta)}(\mu_A) = E_1(\mu_A) + E_\sigma(\mu_A)$$

$$= \sum_{\mu_A(x) \geq \alpha} E_1(\mu_A(x)) + \sum_{\beta < \mu_A(x) < \sigma} E_\sigma(\mu_A(x))$$

$$= \sum_{\mu_A(x) \geq \alpha} (1 - \mu_A(x)) + \sum_{\beta < \mu_A(x) < \sigma} (\sigma - \mu_A(x)), \qquad (4)$$

$$R_{(\alpha,\beta)}(\mu_A) = R_0(\mu_A) + R_\sigma(\mu_A)$$

$$= \sum_{\mu_A(x) \leq \beta} R_0(\mu_A(x)) + \sum_{\sigma < \mu_A(x) < \alpha} R_\sigma(\mu_A(x))$$

$$= \sum_{\mu_A(x) \leq \beta} \mu_A(x) + \sum_{\sigma < \mu_A(x) < \alpha} (\mu_A(x) - \sigma). \qquad (5)$$

Given a fixed σ, the elevation and reduction errors change when the thresholds (α, β) change. No matter which threshold changes and how they change, the elevation and reduction errors always change in opposite directions [22]. The decrease of one type of errors inevitably brings the increase of the other type of errors. The balanced shadowed set based three-value approximations are expected to represent a tradeoff between the elevation and reduction errors.

3.4   Game-Theoretic Shadowed Sets

Game-theoretic shadowed sets (GTSS) use game theory to determine the thresholds in the shadowed set context. The obtained thresholds represent a tradeoff between two different types of errors [22]. GTSS use a game mechanism to formulate games between the elevation and reduction errors. The strategies performed by two players are the changes of thresholds. Two game players compete with each other to maximize their own payoffs. A repetition learning mechanism is adopted to approach a compromise between two players by modifying game formulations repeatedly. The resulting thresholds are determined based on the game equilibria analysis and selected stopping criteria.

Game Formulation. Three elements should be considered when formulating a game G, i.e., game player set O, strategy profile set S, and utility functions u, $G = (O, S, u)$ [8,17]. The game players are the total elevation and reduction errors which are denoted by E and R, i.e., $O = \{E, R\}$.

The strategy profile set is $S = S_E \times S_R$, where $S_E = \{s_1, s_2, ..., s_{k1}\}$ is a set of possible strategies for player E, and $S_R = \{t_1, t_2, ..., t_{k2}\}$ is a set of possible strategies for player R. We select (σ, σ) as the initial threshold values, which represent that we do not have any uncertainty on all membership grades and we have the smallest shadowed area. Starting from (σ, σ), we gradually make α and β further to each other and increase the shadowed area. c_E and c_R are two constant change steps, denoting the quantities that two players E and R use to change the thresholds, respectively. For example, we set the initial threshold values $(\alpha, \beta) = (0.5, 0.5)$. The player E performs increasing α and the player R performs decreasing β. When we set $c_E = 0.01$ and

$c_R = 0.02$, we have $S_E = \{\alpha \text{ no change}, \alpha \text{ increases } 0.01, \alpha \text{ increases } 0.02\}$, and $S_R = \{\beta \text{ no change}, \beta \text{ decreases } 0.02, \beta \text{ decreases } 0.02\}$.

The payoffs of players are $u = (u_E, u_R)$, and u_E and u_R denote the payoff functions of players E and R, respectively. The payoff functions $u_E(\alpha, \beta)$ and $u_R(\alpha, \beta)$ are defined by the elevation and reduction errors, respectively, that is,

$$u_E(\alpha, \beta) = E_{(\alpha, \beta)}(\mu_A),$$
$$u_R(\alpha, \beta) = R_{(\alpha, \beta)}(\mu_A), \tag{6}$$

where $E_{(\alpha, \beta)}(\mu_A)$ and $R_{(\alpha, \beta)}(\mu_A)$ are defined in Eqs. (4) and (5). We try to minimize the elevation and reduction errors, so both players try to minimize their payoff values.

We use payoff tables to represent two-player games. Table 1 shows a payoff table example in which both players have 3 strategies.

Table 1. An example of a payoff table

		R		
		β	$\beta \searrow c_R$	$\beta \searrow 2c_R$
E	α	$\langle u_E(\alpha, \beta), u_R(\alpha, \beta) \rangle$	$\langle u_E(\alpha, \beta - c_R),$ $u_R(\alpha, \beta - c_R) \rangle$	$\langle u_E(\alpha, \beta - 2c_R),$ $u_R(\alpha, \beta - 2c_R) \rangle$
	$\alpha \nearrow c_E$	$\langle u_E(\alpha + c_E, \beta),$ $u_R(\alpha + c_E, \beta) \rangle$	$\langle u_E(\alpha + c_E, \beta - c_R),$ $u_R(\alpha + c_E, \beta - c_R) \rangle$	$\langle u_E(\alpha + c_E, \beta - 2c_R),$ $u_R(\alpha + c_E, \beta - 2c_R) \rangle$
	$\alpha \nearrow 2c_E$	$\langle u_E(\alpha + 2c_E, \beta),$ $u_R(\alpha + 2c_E, \beta) \rangle$	$\langle u_E(\alpha + 2c_E, \beta - c_R),$ $u_R(\alpha + 2c_E, \beta - c_R) \rangle$	$\langle u_E(\alpha + 2c_E, \beta - 2c_R),$ $u_R(\alpha + 2c_E, \beta - 2c_R) \rangle$

Repetition Learning Mechanism. The involved players are trying to maximize their own payoffs in the formulated games. But one player's payoff is effected by the strategies performed by the other player. The balanced solution or game equilibrium is a strategy profile from which both players benefit. This game equilibrium represents both players reach a compromise or tradeoff on the conflict. The strategy profile (s_i, t_j) is a pure strategy Nash equilibrium, if for players E and R, s_i and t_j are the best responses to each other [17], this is,

$$\forall s_k \in S_E, \quad u_E(s_i, t_j) \geqslant u_E(s_k, t_j), \quad \text{where } s_i, s_k \in S_E \text{ and } k \neq i, t_j \in S_R,$$
$$\forall t_l \in S_R, \quad u_R(s_i, t_j) \geqslant u_R(s_i, t_l), \quad \text{where } t_j, t_l \in S_R \text{ and } l \neq j, s_i \in S_E. \tag{7}$$

The above equations can be interpreted as a strategy profile such that no player would like to change his/her strategy or they would loss benefit if deriving from this strategy profile, provided this player has the knowledge of other player's strategies.

The equilibrium of the current formulated game means the threshold pair corresponding this equilibrium are the best choices within the current strategy sets. We have to check if there are some threshold pairs near the current equilibrium that are better than the current ones. Thus we repeat the games with the updated initial thresholds. We may be able to find more suitable thresholds with repetition of thresholds modification.

We define the stopping criteria so that the iterations of games can stop at a proper time. There are many possible stopping criteria. For example, the payoff of each player is beyond a specific value; the thresholds (α, β) violate the constraint $0 \leq \beta \leq \sigma \leq \alpha \leq 1$; the current game equilibrium does not improve the payoffs gained by both players under the initial thresholds; no equilibrium exists. In this research, we compare the payoffs of both players under the initial thresholds and the thresholds corresponding to the current equilibrium. We set the stopping criteria as one of the players increases its payoff values, or there does not exist a pure strategy Nash equilibrium in the current game.

3.5   Feature Extraction

We now select the most representative words by applying the thresholds (α, β) derived in previous sections. The words with normalized DW_t being greater than the upper threshold α and less than the lower threshold β will be kept as our word features for text classification while the rest words are discarded.

4   Experiments

4.1   Dataset and Evaluation Metrics

We evaluate our approach on the HuffPost news category dataset [18]. This dataset consists of 200,853 news headlines with short descriptions from HuffPost website during the year 2012 to 2018. It contains 31 categories of news such as politics, entertainment, business, healthy living, art, and so forth. We use the largest two categories, the 32,739 politics news and 14,257 entertainment news, as the binary text classification data in our experiments. The news text is obtained by concatenating the news headline and the corresponding short description. We extract 381449 words from these selected news. We use 80% data for training and 20% for testing, and adopt accuracy and F1 scores as metrics for model evaluation.

4.2   Deriving Thresholds with GTSS

We first normalize DW_t using min-max normalization linear transformation. The distribution of normalized DW_t is shown in Fig. 2 Almost 80% of words have the normalized DW_t 0.548054 so we set $\sigma = 0.5481$ aiming to minimize the errors produced by mapping all DW_t values to three values $\{0, \sigma, 1\}$ via game-theoretic shadowed set model. If we set σ as other value instead of 0.548054, mapping the large amount of DW_t 0.548054 to σ definitely will produce more errors.
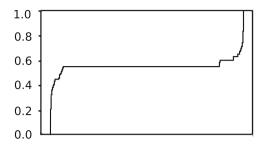
Fig. 2. Normalized DW_t information

We formulate a competitive game between the elevation and reduction errors to obtain the thresholds. The game players are elevation and reduction errors, i.e., $O = \{E, R\}$. The strategy profile set is $S = \{S_E \times S_R\}$. The game is being played with the initial thresholds $(\alpha, \beta) = (0.55, 0.54)$. Here, $\alpha = 0.55$ is the smallest value that greater than $\sigma = 0.5481$, and $\beta = 0.54$ is the largest value that less than $\sigma = 0.5481$. The player E and R try to increase α and decrease β with the change steps as 0.01 and 0.02, respectively. The strategy set of E is $S_E = \{\alpha$ no change, α increases 0.01, α increases 0.02$\}$. The corresponding α values are 0.55, 0.56, and 0.57. The strategy set of R is $S_R = \{\beta$ no change, β decreases 0.02, β decreases 0.4$\}$. The corresponding α values are 0.54, 0.52, and 0.5. The payoff functions are defined in Eqs. (4) and (5). Table 2 is the payoff table. The cell at the right bottom corner is the game equilibrium whose strategy profile is (α increases 0.02, β decreases 0.04). The payoffs of the players are (17689, 8742). We set the stopping criterion as one of players' payoff increases. When the thresholds change from (0.55, 0.54) to (0.57, 0.5), the elevation error is decreased from 17933 to 17689, and the reduction error is decreased from 10472 to 8742. We repeat the game by setting $(0.57, 0.5)$ as the initial thresholds.

Table 2. The payoff table

		R		
		β	$\beta \downarrow 0.02$	$\beta \downarrow 0.04$
E	α	<17933, 10472>	<17965, 9624>	<18034, 8733>
	$\alpha \uparrow 0.01$	<17690, 10476>	<17722, 9628>	<17792, 8738>
	$\alpha \uparrow 0.02$	<17588, 10480>	<17620, 9632>	**<17689, 8742>**

The competitive games are repeated four times. The result is shown in Table 3. In the fourth iteration, we find out that the payoff value of player E increases. The repetition of game is stopped and the final result is the initial thresholds of the fourth game $(\alpha, \beta) = (0.61, 0.42)$.

Table 3. The repetition of game

	Initial(α, β)	Result(α, β)	Payoffs	both decrease?
1	(0.55, 0.54)	(0.57, 0.5)	<17689, 8742>	$\checkmark$
2	(0.57, 0.5)	(0.59, 0.46)	<17686, 6516>	$\checkmark$
3	(0.59, 0.46)	(0.61, 0.42)	<7832, 3934>	$\checkmark$
4	(0.61, 0.42)	(0.63, 0.38)	<8221, 2693>	$\times$

We got $(\alpha, \beta) = (0.61, 0.42)$, which means we keep the words with DW_t greater than 0.61 and less then 0.42, and discard the words with DW_t between 0.61 and 0.42.

4.3   Baselines

We calculate the DW_t value for each single word and bi-gram, and then use Support Vector Machine (SVM) [1,2] as our unique classifier to compare our approach with: (1) **ALL**, in which we keep all words with no feature extraction; (2) **Sym-Cutoff**, in which the symmetrical cut-off values are drawn purely based on a simple observation of the statistical distribution of DW_t; (3) **Sym-N-Words**, where we select $2n$ words given n smallest DW_t and n largest DW_t such that $2n$ is approximately equal to the total number of words extracted with our approach.

4.4   Results

We show the model performance of different extraction approaches on the new category dataset in Table 4. Our model is named as "Asym-GTSS-TH". From the results, we can observe that: (1) Our approach achieves superior performance compared with Sym-Cutoff which is purely based on a guess given TF-IDF distribution. It verifies our claim that the GTSS can better capture the pattern of TF-IDF and provide a more robust range for selecting relevant words given TF-IDF for text classification; (2) The Sym-N-Words approach achieves close performance as ours, because it takes the advantage of the information of the number of words to keep derived with our thresholds. However, our approach still outperforms the Sym-N-words approach since it evenly selects relevant words for document c_1 and c_2. It indicates that there exists imbalance of representative words between different categories of documents which is better captured by our model; (3) Compared with using all words' TF-IDF score as input, our model discards more than 52% words and speed up the process of classification while preserving the performance.

Table 4. Summary of the performance of different feature extraction approaches

	Accuracy	F1-score
ALL	94.7	92.3
Sym-Cutoff	91.9	88.2
Sym-N-Words	94.5	91.7
Asym-GTSS-TH	94.6	91.9

5   Conclusion

In this paper, we propose a feature extraction approach based on TF-IDF and game-theoretic shadowed sets in which the asymmetric thresholds for selecting relevant words are derived by repetitive learning on the difference of TF-IDF for each word between documents. Our model can explore the pattern of TF-IDF distribution as well as capture the imbalance of the number of representative words in different categories. The experimental results on the news category dataset show that our model can achieve improvement as compared to other cut-off policies and speed up the information retrieval process. In the future, we will explore the consistency of our model performance on more real world datasets and test the generalization ability of our GTSS model on different metrics that measures the relevance of words, such as BNS and Chi-square.

References

1. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. J. Mach. Learn. Res. **2**(12), 125–137 (2001)
2. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
3. Deng, X.F., Yao, Y.Y.: Decision-theoretic three-way approximations of fuzzy sets. Inf. Sci. **279**, 702–715 (2014)
4. Grossman, D.A., Frieder, O.: Information Retrieval: Algorithms and Heuristics, vol. 15. Springer, Netherlands (2012). https://doi.org/10.1007/978-1-4020-3005-5
5. Grzegorzewski, P.: Fuzzy number approximation via shadowed sets. Inf. Sci. **225**, 35–46 (2013)
6. Jing, L.P., Huang, H.K., Shi, H.B.: Improved feature selection approach TFIDF in text mining. In: Proceedings of the International Conference on Machine Learning and Cybernetics, vol. 2, pp. 944–946. IEEE (2002)
7. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Education, Inc., New Jersey (2008)
8. Leyton-Brown, K., Shoham, Y.: Essentials of game theory: a concise multidisciplinary introduction. Synth. Lect. Artif. Intell. Mach. Learn. **2**(1), 1–88 (2008)
9. Lopes, L., Vieira, R.: Evaluation of cutoff policies for term extraction. J. Braz. Comput. Soc. **21**(1), 1–13 (2015). https://doi.org/10.1186/s13173-015-0025-0
10. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)

11. Milios, E., Zhang, Y., He, B., Dong, L.: Automatic term extraction and document similarity in special text corpora. In: Proceedings of the 6th Conference of the Pacific Association for Computational Linguistics, pp. 275–284. Citeseer (2003)
12. Özgür, A., Özgür, L., Güngör, T.: Text categorization with class-based and corpus-based keyword selection. In: Yolum, I., Güngör, T., Gürgen, F., Özturan, C. (eds.) ISCIS 2005. LNCS, vol. 3733, pp. 606–615. Springer, Heidelberg (2005). https://doi.org/10.1007/11569596_63
13. Pedrycz, W.: Shadowed sets: representing and processing fuzzy sets. IEEE Trans. Syst. Man Cybern. **28**(1), 103–109 (1998)
14. Pedrycz, W.: Shadowed sets: bridging fuzzy and rough sets. In: Rough Fuzzy Hybridization: A New Trend in Decision-Making, pp. 179–199 (1999)
15. Pedrycz, W.: From fuzzy sets to shadowed sets: interpretation and computing. Int. J. Intell. Syst. **24**(1), 48–61 (2009)
16. Pedrycz, W., Vukovich, G.: Granular computing with shadowed sets. Int. J. Intell. Syst. **17**(2), 173–197 (2002)
17. Rasmusen, E.: Games and Information: An Introduction to Game Theory. Blackwell, Oxford (1989)
18. Misra, R.: News category dataset from HuffPost website (2018). https://doi.org/10.13140/RG.2.2.20331.18729
19. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. **24**(5), 513–523 (1988)
20. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill Inc. (1986)
21. Zadeh, L.A.: Toward extended fuzzy logic - a first step. Fuzzy Sets Syst. **160**(21), 3175–3181 (2009)
22. Zhang, Y., Yao, J.T.: Game theoretic approach to shadowed sets: a three-way tradeoff perspective. Inf. Sci. **507**, 540–552 (2020)
23. Zuo, Z., Li, J., Anderson, P., Yang, L., Naik, N.: Grooming detection using fuzzy-rough feature selection and text classification. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8. IEEE (2018)