2019 9th International Conference on Social Science and Education Research (SSER 2019)

Cultural Background Differences in Legal Translation

Suwen Li

Zhengzhou University, No.100 Science Avenue, Zhengzhou City, Henan Province, P.R.China

Abstract: In the process of legal translation, it is necessary to master and master the legal system and legal culture of the relevant countries. Therefore, how to eliminate the barriers of communication between the original and the translated versions of the law due to cultural differences is a necessary prerequisite and a new challenge in the process of legal translation. Based on the cultural differences in the process of translation, this paper proposes a multi-label-based feature selection algorithm, which uses the classification interval and the number of sample classification intervals to classify the translation samples, and uses the same evaluation index and data set to compare the three mature algorithms. The experimental results show that the multi-label-based feature selection algorithm proposed in this paper is average accurate. The rate is the best, which shows that the algorithm has good classification performance.

Keywords: Legal Translation, Cultural Background Differences, Multi-Tag Based Feature Selection Algorithm

1. Introduction

In the past two decades, the translation industry has undergone a major transformation, usually requiring a high level of thematic expertise. With the increasing demand for legal translation, specialization and task complexity are particularly important for legal translation professionals [1]. For a long time, legal translation has been regarded as a way of communicating orderly legal orders, and the focus of attention has shifted to the emergence of multilingual laws in the legal world characterized by multiple interrelated levels. Legal communication (which must meet the needs of experts and respect the rights of citizens) is the result of a complex process of language communication, where legal translation is a specialized activity that requires expertise and involves the choice of highly specific institutional environments [2]. Legal translation plays a very important role in the communication between different nationalities and cultures in history, and plays a more important role in the increasingly globalized world [3]. Legal translation shows people's source and target languages in the legal, cultural and linguistic fields [4]. The Untranslatability of legal terms, especially the long-term translation between initially unrelated languages, has always been a real challenge in legal translation. It originates from the inconsistency of concepts between legal terms in different legal languages, which originate from different legal cultures and legal systems [5]. The concept of translation is no longer confined to the process of translating text from one language to another, but also describes the process of adapting to foreign knowledge, value or practice [6]. Studies of culture and legal transfer reveal significant similarities and differences between cultures [7].

In recent years, the task of text classification, scene automatic annotation, gene function prediction and other multi-label fields has aroused great interest [8]. Like traditional single label classification, feature selection plays an important role in multi-label classification [9]. Multi-label learning is mainly used to process data associated with a set of labels at the same time. The high dimensionality of data is the stumbling block of multi-label learning [10].

In this paper, data analysis and data mining theory in big data technology are introduced into the analysis of cultural background differences in legal translation process, and a classification model based on multi-label feature selection algorithm is established in view of cultural characteristics differences in translation process.

2. Method

2.1 Cultural Background Differences in Legal Translation

Different languages represent different cultures of different countries, and there are great differences and differences in different cultural backgrounds. Different cultural backgrounds influence people's language expression and understanding of different things. Language activities that correctly and completely re-express the thinking content of

Copyright © (2019) Francis Academic Press, UK **DOI: 10.25236/sser.2019.067**

another language with the help of one language are called translation. Translation is a kind of cross-linguistic and cross-cultural communication, which mainly transforms the meaning carried by one language culture into another. Languages from different cultural backgrounds are quite different in grammar. Some languages attach importance to understanding and vividness, focusing on how to grasp comprehension and spiritual world at a macro level, while others are characterized by rationality and attach importance to grasping object accuracy and formal argument. Different cultural connotations will produce different ways of language cognition, which will further affect the effect of translation and translation strategies. And culture has its unique connotation. There are great cultural differences among different countries, nationalities and societies, so culture has national, national and social characteristics. The connotation and characteristics of culture are formed step by step in certain social changes, historical conditions and natural environment.

The process of legal translation is the transformation of the unique connotations of national politics, culture, legal thinking and language. Therefore, as a part of culture, there are also great differences in legal culture. Only by comprehensively and scientifically analyzing the influencing factors in the process of legal translation and adopting effective means and methods can the process of legal translation truly achieve the goal of integration and coherence. Cultural differences often restrict and affect the quality of legal language conversion, so the inconsistency between Chinese and English legal words caused by cultural differences is one of the biggest difficulties in legal translation. Therefore, it is necessary to fully recognize the different cultural backgrounds and related legal and cultural knowledge, and to strengthen the research on the differences of legal knowledge and legal and cultural characteristics between China and other countries, so that legal translators can carry out their work accurately and effectively.

2.2 Application of Data Analysis in Translation Process

With the rapid development of the information age, legal translation can be perfected by means of big data analysis technology through various legal sentences (samples) and different meanings (features) represented by their vocabulary in the process of translation, ultimately translated into relevant meanings (markers). However, the meanings of some words are not directly related to the whole sentence or have nothing to do with it at all. Therefore, in order to help us understand the meaning of the whole sentence, we need to find some obvious features and delete some irrelevant features. This is the concept of feature selection in large data analysis in the process of legal translation.

In order to make the translation of legal translation process more accurate and accurate, the complexity of translation storage can be reduced by feature dimension reduction, but the process of feature dimension reduction will destroy the feature space of the original data to a certain extent, so the problems encountered in feature extraction can be avoided by feature selection. That is to say, a set of evaluation criteria can be selected from the process of legal translation to accurately judge the characteristics of the translation, and new concrete meanings can be obtained, which can represent the original legal translation, thus reducing the complexity of the large data mining process of translation. Feature selection, also known as attribute reduction or feature subset selection, refers to the selection of feature subsets that can represent the physical meaning of the original feature space from existing features. It is the processing step of key data in pattern recognition. In order to remove some features with low performance and high cost in time and space, feature space dimension reduction is the process.

Different cultural backgrounds show different meanings (features) in translation and the same language and sentence have very complex expressions, which makes the feature space of data have high dimensionality. Therefore, data in general legal translation process often presents the characteristics of high dimensionality and multi-dimensionality.

2.3 Sample Differences in Legal Translation

Some features in the process of multi-label learning determine the diversity of samples. It is necessary to recognize such features to improve the classification performance. Each sample is not only a label belonging to a certain category in the multi-label learning process, but the semantics of the sample object may need to be represented by multiple category labels at the same time. At the same time, not all the sample features in the multi-label data set have the same importance, so clustering technology can be used to sample the multi-label data set to form a new multi-label decision-making system.

Suppose there is a multi-labeled data set $D = \{(x_i, l_i)|1 \leq i \leq n\}$, and the feature vectors of the d-dimensional sample x_i are composed of $[(x_{i_1}, x_{i_2}, \ldots x_{i_d})^T$, where $x_i \in L$. l_k is used to represent all marked sets of sample x_i. For markers $l_k \in L$, the positive class sample set can be expressed as:

$P_k = \{x_i|(x_i, l_i) \in D, l_k \in L\},$ (1)

the negative class sample set can be expressed as: $N_k = \{x_i|(x_i, l_i) \in D, l_k \notin L\}$ (2)

P_k denotes a set of samples belonging to category l_k and N_k denotes a set of samples not belonging to category label l_k.

In order to effectively represent data and analyze the intrinsic properties of samples, this paper uses K-means to cluster positive and negative samples. The cluster centers $m_k{}^+$ of the set of samplesP_k are $\{p_1{}^k, p_2{}^k, \ldots, p_{m_k{}^+}{}^k\}$ .The cluster centers $m_k{}^-$ of negative sample set N_k are $\{n_1{}^k, n_2{}^k, \ldots, n_{m_k{}^-}{}^k\}$. If the number of samples of positive and negative classes is unbalanced, the number of clusters of P_k and N_k is equal, that is, $m_k = m_k{}^- = m_k{}^+$, and the number of clusters of samples on positive and negative classes set P_k and N_k is set to $m_k = [r \cdot \min(|P_k|, |N_k|)]$ (3).

$r \in [0,1]$ is set in order to limit the number of clustering samples.

The multi-label data set D can be transformed into a multi-label decision system composed of representative samples < U, F, L> by formula (1) and formula (2) (3), where the sample set $U = \{x_1, x_2, \ldots, x_n\}$, a set of features $F = \{f_1, f_2, \ldots, f_n\}$, a set of tags $L = \{l_1, l_2, \ldots, l_n\}$.

The relationship between samples in markup space of multi-labeled data sets is uncertain. The classes divided may not be identical under different category markers, that is mean that they may be considered to be similar under one label but different under another label. Therefore, measuring the importance of features only from the interval of samples has some limitations in multi-label decision-making system.

For multi-label decision systems < U, F, L>, there exists $l \in L$. Given sample x, the classification interval of sample x under feature f is defined as: $m_f^l(x) = \triangle\left(x, NM_f^l(x)\right) - \triangle\,(x, H_f^l(x))$ (4)

$NH_f^l(x)$ among formula (4) denotes the same sample nearest to the sample x under the category label, and $NM_f^l(x)$ denotes the different sample nearest to the sample x under the category label. $\triangle\left(x, NM_f^l(x)\right)$ denotes the distance from sample point x to M_f^l, and $\triangle\left(x, NH_f^l(x)\right)$ denotes the distance from sample point x to $H_f^l(x)$.

According to the idea of Relief algorithm, the weight of features can be judged by the classification interval of samples. The larger the classification interval between the samples, the stronger the discriminating ability of the features; otherwise, the weaker the discriminating ability of the features. In a feature space, the classification interval is used to measure the discriminative ability of samples in multi-label decision-making system. That is to say, given the multi-label decision-making system $< U, F, L >$, the classification interval of feature$f \in F$ in the sample space is $m_f^l(x)$ For the whole label space L, the classification interval $m_f^l(x)$ of sample x under feature f is defined as:

$m_f^l(x) = \sum_{l \in L} m_f^l(x)$ (5)

When $m_f^l(x) > 0$, the distance from sample x to different class labels for label L is greater than that from the same class label sample for label l, that is to say, the feature is separable for sample x; otherwise it is inseparable. For convenience of calculation, $m_f^l(x) < 0$ is set to $m_f^l(x) = 0$

In the sample space, the set of samples whose classification interval is greater than zero is denoted as follows:

$N_f^l(x) = \{x_i|m_f^l(x_i) > 0, x_i \in U\}$ (6)

Which the number of samples whose classification interval is greater than zero is$|N_f^l(x)|$.

In the multi-label decision system$< U, F, L >$, for all$x \in U$, i f $m_f^l(x_i) > 0$ and $\left|N_f^l(x)\right| > 0$ under the characteristic f condition f, the samples are not identical, \and the samples are different.

W as features for the weight vectors with $x \in U$, the evaluation function of the feature subset is

$e(f) = \sum_{l \in L} N_f^l \cdot \sum_{l \in L} m_f^l(x)$ (7)

By calculating the weights of the features measured by the classification interval of sample x under feature f, the formula for calculating the weight w_f of feature f is as follows.

$$w_f = \sum_{l=1}^{t} N^l \cdot \sum_{i=1}^{n}(d_f(x_i, NM^l(x_i)) - d_f(x_i, NH^l(x_i))) \qquad (8)$$

$(d_f(x_i, NM^l(x_i))$ among them represents the sample nearest to and most different from the sample $x_i$under the feature F and class label l, and$d_f(x_i, NH^l(x_i))$ represents the distance from the sample nearest to the sample with the same label. Distance $d_f(x, y)$ is defined as:

$$d_f(x, y) = \frac{|x(f) - y(f)|}{\max(f) - \min(f)} \qquad (9)$$

For the data analysis of the legal translation process, it is necessary to calculate the interval of the samples in the vocabulary (feature space) under each translation result (marker), and also to fully consider the influence of the difference of different sentences (samples) on the feature weight learning. In this paper, we first use the clustering method to obtain representative sentences (samples) in legal sentences (multi-label systems) in different contexts; then define the classification intervals and differences of legal sentences (samples) that need to be translated in multi-marking decision systems. The calculation formula of the feature weight and the corresponding algorithm are designed in this foundation. Finally, the effectiveness of the proposed algorithm is verified by experiments.

3. Experimental simulation

3.1 Algorithmic Design

This paper designs a multi-label feature selection algorithm (MFSD) based on the sample differences in legal translation. The specific algorithm steps are as follows:

1) Enter a multi-label data set D;

2) According to formula (1) (2) (3), a multi-label decision-making system consisting of representative samples $< \mathrm{U, F, L} >$ is obtained.

3) The loop executes when $\mathrm{f} \in \mathrm{F}$.

4) Calculate the weight of each feature according to formula (8).

5) End of cycle

6) The eigenvalues are sorted in label_ feature according to the weight of the eigenvalues.

In order to verify the effectiveness of the proposed algorithm, MFSD algorithm, MDDMspc algorithm and MDDMproj algorithm are used as comparison algorithms. In this paper, we choose a set of features to rank, in which the number of features is set as the number of features in Relief algorithm. ML-KNN is used to measure feature selection, in which the smoothing parameter s is set to 1 and K is set to 10. *3.2 Experimental data*

This paper selects the legally relevant statement data in the Australian Sign Language signs, and the selected data are applied to different classifications of legal translation. The data related information is shown as Table 1 .

Table1 Related data sets in the experiment

Number of statements	Number of samples	Number of features	Number of Categories	Number of training samples	Number of test samples
DS1(5671)	5100	463	26	2100	3000
DS2(3461)	3200	231	14	1100	2100
DS3(537)	570	51	6	391	109
DS4(5723)	5200	791	37	2000	3200

3.3 Evaluation indicators

In order to verify the effectiveness of the algorithm, this paper evaluates through the following indicators.

1) In order to evaluate the sort order of the markers in the concept class in the sample, the experiment uses the average precision to indicate the average proportion of the markers ranked before a particular marker .Which is recorded as $\mathrm{AP(f)} = \frac{1}{m}\sum_{i=1}^{m}\frac{\{k|rank_f(x_i,k) \le rank_f(x_i,l), k \in R_i\}}{rank_f(x_i,l)}$

2) This paper uses Hamming loss rate to investigate the misclassification of samples on a single concept class, which is recorded as $\mathrm{HL(h)} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{L}\sum_{l=1}^{L}[h_l(x_i) \neq Y_{il}]$

3) In order to traverse all the tags associated with the sample, this article uses the coverage metric to average the number of steps each sample needs to find, which is record as $\mathrm{CV(f)} = \frac{1}{m}\sum_{i=1}^{m}\max_{l\in R_i} rank_f(x_i, l) - 1$, Where $rank_f(x, l),, k \in R_i$ represents the ordering function of the prediction function $f_l(x)$.

4. Experimental results and analysis

1) In all data sets, the average precision of MFSD algorithm is the first. Both MDDMspc algorithm and MDMDproj algorithm are lower than the average precision of the MLNB algorithm, as shown in Table 2.

Table 2 Comparison of the average precision of each algorithm

DataSet	MDDMspc	MDDM proj	MFSD	Original
DS1(5671)	0.5071	0.494	0.516	0.501
DS2(3461)	0.634	0.628	0.639	0.634
DS3(537)	0.773	0.76	0.784	0.781
DS4(5723)	0.471	0.551	0.581	0.457
Average	0.571	0.608	0.631	0.593

2) Under the Hamming loss evaluation index, the MFSD algorithm obtains the optimal value under the three data sets, only the result under the fourth data set is not optimal, and the sub-optimal value is obtained, which is shown as Table 3.

Table3 Comparison of the Hamming Loss of each algorithm

DataSet	MDDMspc	MDDM proj	MFSD	Original
DS1(5671)	0.0607	0.0614	0.0627	0.0601
DS2(3461)	0.243	0.245	0.252	0.214
DS3(537)	0.063	0.062	0.065	0.0681
DS4(5723)	0.031	0.0321	0.0315	0.0347
Average	0.094	0.1001	0.102	0.0931

3) For the coverage index, the optimal coverage value is obtained for MFSD algorithm on the given experimental data set, suboptimal results were obtained for MDDMspc algorithm, which is shown as Table 4.

Table4 Comparison of the Coverage of each algorithm

DataSet	MDDMspc	MDDM proj	MFSD	Original
DS1(5671)	5.47	5.55	5.58	5.44
DS2(3461)	4.39	4.44	4.47	4.414
DS3(537)	1.94	2.04	2.07	1.9681
DS4(5723)	4.94	4.84	4.99	3.147
Average	4.18	4.21	4.26	3.67

From the analysis of the above experimental results, it can be seen that the MFSD algorithm proposed in this paper is not optimal in the fourth data set except the evaluation index HL in the four data sets, and the optimal values are obtained in other results, we can see that the chance of winning of the algorithm proposed in this paper to achieve 95% .The larger the value of the average precision, the better the classification performance. It can be seen from the relationship between the number of features and the average precision in Figure 1. As the number of features increases, the average precision of all comparison algorithms also shows an upward trend, but the precision of the MFSD algorithm in the four data sets is the best. At the same time, it is obvious that the proposed MFSD algorithm is effective in overall effect and has great advantages.

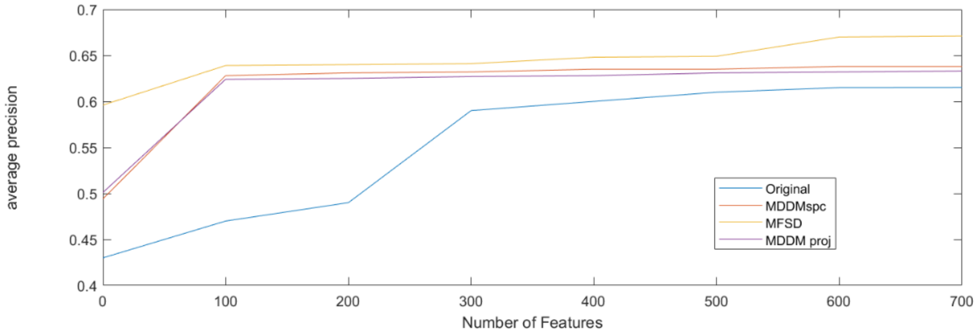

Figure1 The Relation Diagram between the Number of Features and the Average Precision Rate

5. Conclusion

Due to the differences in thinking styles in different cultural contexts, many factors should be considered in the translation process of the law. This paper proposes a multi-mark feature selection algorithm based on sample difference for data analysis and data mining of cultural background differences in legal translation process. Based on the definition of the sample difference in the translation process, this paper starts from the classification interval of the sample and the number of sample classification intervals, and uses the same data set and rating index to compare the proposed algorithm and the three algorithms. The result of the experiment shows that the multi-mark based feature selection algorithm has better classification performance.

References

[1]Rodríguez-Castro, M., & Sullivan, C. E. (2015). Rethinking the legal translation classroom: a course for legal translation professionals. *Interpreter & Translator Trainer, 9*(2), 205-228.

[2] Valérie, Dullion. (2017). Between specialised texts and institutional contexts – competence and choice in legal translation. Translation and Translanguaging in Multilingual Contexts, 3(1), 1-4.

[3] Loiacono, R., & Bertoli, L. (2018). The complexities of legal translation in the drafting of bilateral treaties between Italy and English-speaking countries. Perspectives, 26(5), 646-662.

[4] Stepanova, V. V. (2017). Translation strategies of legal texts: experimental approach ☆. Procedia - Social and Behavioral Sciences,237, 1201-1207.

[5] Alwazna, R. Y. (2019). Translation and legal terminology: Techniques for coping with the untranslatability of legal terms between Arabic and English. International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 32(1), 75-94.

[6] Foljanty, L. (2015). Legal Transfers as Processes of Cultural Translation: On the Consequences of a Metaphor. Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft, (2), 89-107.

[7]van der Burg, M. (2015). Cultural and legal transfer in Napoleonic Europe: codification of Dutch civil law as a cross-national process. Comparative Legal History, 3(1), 85-109.

[8] Lee, J., & Kim, D. W. (2013). Feature selection for multi-label classification using multivariate mutual information. *Pattern Recognition Letters*, *34*(3), 349-357.

[9] Lee J., Kim D. Mutual information-based multi-label feature selection using interaction information[J].Expert Systems with Applications, 2015, 42: 2013-2025.

[10] Lin, Y., Hu, Q., Liu, J., Chen, J., & Duan, J. (2016). Multi-label feature selection based on neighborhood mutual information. *Applied Soft Computing*, *38*, 244-256.