Application of Intelligent Fuzzy Decision Tree Algorithm in English Translation Education

Bing Chen(✉)

Chengdu Polytechnic, Chengdu 610000, Sichuan, China
applechen123@126.com

Abstract. Fuzziness is the basic characteristic and conventional phenomenon of the real world. As a language describing the objective world, it is bound to have fuzziness. Fuzziness, as one of the basic characteristics of language, widely exists in any language in the world and frequently appears in people's daily communication and literary works. It can express infinite meaning with limited language means. The accurate use of fuzzy language can enhance the flexibility of language and improve the expression effect of language. Fuzzy language is used at all levels of language use, resulting in a humorous, funny and vivid language charm. Although fuzziness is common to all languages, due to different life and cultural backgrounds, English and Chinese have great differences in Fuzziness and fuzzy beauty. These differences have brought many challenges and difficulties to translation.

Keywords: Fuzziness · English translation · Decision tree algorithm

1 Introduction

Fuzzy information is a common phenomenon in natural language, which mainly includes lexical fuzziness and semantic fuzziness. Lexical vagueness includes time words, age words, color words, temperature words, taste words, etc. because the boundary of the meaning it refers to is difficult to be clearly defined, it forms fuzzy information. For example, there is generally no clear definition of how old "young people" refer to. Semantic vagueness can also be reflected in some words used in sentences, such as I wonder, I propose, it is estimated, it sees, which can mean subjective speculation or reporting others' views; A bit, adequately, approximately, around and other words can adjust the degree of authenticity or scope of change of discourse [1].

With the development of information processing technology, a large number of intelligent translation software continue to appear. Machine translation gradually replaces manual translation and becomes an important role in English translation in the future. The accuracy and adaptability of machine translation become an important topic in the research of translation software in the future. The optimization of machine translation is based on semantic selection and feature extraction. By extracting the text semantic

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 310–315, 2023.
https://doi.org/10.1007/978-3-031-23950-2_33

information features of English context, combined with intelligent analysis and semantic information retrieval technology, ontology mapping and adaptive adjoint tracking of semantic information are realized, and English intelligent translation is realized combined with pattern recognition method, The research on fuzzy semantic selection technology in English semantic machine translation is of great significance in optimizing machine translation software design and improving translation accuracy and artificial intelligence. The traditional methods mainly include topic tree feature matching method, support vector machine algorithm, particle swarm optimization evolution method, etc. language evaluation set model construction and feature combination matching are carried out for the semantic features of English translation to improve the accuracy of translation, and some research results have been achieved. However, the traditional methods have some problems, such as weak suppression ability of context interference, easy coupling interference and so on.

2 Overview of Relevant Technologies

2.1 Overview of Fuzziness

Classical logic is a binary logic that satisfies the law of contradiction and exclusion, that is, it can only be judged according to binary logic such as “yes and no”, “right and wrong”, “0 and 1”, and there is no middle zone. However, as the main body of understanding the world and transforming the objective world, the reflection of human brain on natural phenomena is usually vague. This fuzziness is reflected in that it is not a clear and accurate reflection of “either this or that”. It more reflects the uncertainty between the attributes of things and the attribute division caused by the transition of differences, which makes the concept unclear and appears “this or that” and “seems to be rather than”. In 1965, Professor L.A. Zadeh of the University of California published a famous paper, in which he first put forward an important concept to express the fuzziness of things: membership function, which broke through Cantor’s classical set theory at the end of the 19th century and laid the foundation of fuzzy theory. It not only broadens the classical mathematics, but also builds a bridge between man and machine in theory, making the intelligence of machine possible [2].

Definition: fuzziness refers to the indistinct extension of the concept of things in the objective world, and the same and different attribution of things to the concept.

$$U = \{u_1, u_2, \ldots u_n\} \tag{1}$$

$$V = \{v_1, v_2, \ldots v_m\} \tag{2}$$

$$A = \{a_1, a_2, \ldots a_n\} \tag{3}$$

$$R_i = \{r_{i1}, r_{i2}, \ldots r_{in}\} \tag{4}$$

2.2 Decision Tree Algorithm

As a common classification algorithm in data mining, decision tree algorithm has been widely used in many fields. It has many advantages, such as small amount of calculation, convenient rule acquisition, high classification accuracy and so on. From the initial CLS algorithm to today's various decision tree classification algorithms, the representative ones are ID3, C4.5 algorithm and so on. The following is a brief description of these common decision tree classification algorithms.

1) ID3 Decision Tree Algorithm

ID3 decision tree algorithm, which mainly aims at the problem of selecting attributes during the construction of decision tree. It is the most influential and classic algorithm in decision tree. ID3 decision tree algorithm takes information theory as the theoretical basis, measures the criteria of attribute selection through information entropy and information gain in information theory, and realizes the best selection of candidate attributes. The algorithm calculates the information gain of each candidate attribute, and then selects the candidate attribute that can maximize the information gain as the test attribute to construct the node of the decision tree, so as to construct the decision tree.

ID3 decision tree algorithm has clear theory and simple algorithm structure by using information gain in attribute selection, but ID3 algorithm depends on attributes with a large number of attribute values when calculating information gain, and ID3 algorithm is sensitive to noise and can not deal with continuous attributes well.

$$\mathrm{SplitInfo}\,(A) = -\sum_{i=1}^{m} \frac{|S_j|}{|S|} \times \log_2 \frac{|S_j|}{|S|} \tag{5}$$

2) C4.5 Decision Tree Algorithm

Aiming at some shortcomings of ID3 algorithm, C4.5 decision tree algorithm uses information gain rate to select the best candidate attribute as the test attribute, which can eliminate the dependence of ID3 algorithm on multivalued attributes, and C4.5 algorithm can have good processing effect on continuous attributes and can process incomplete data sets.

Due to the complexity of fuzziness itself, it is difficult to fit with a unified form of function in practice. At present, the selection and establishment methods of membership function mainly include fuzzy based method, comparative ranking method, expert evaluation method and so on. As shown in Fig. 1, triangular membership function diagram.

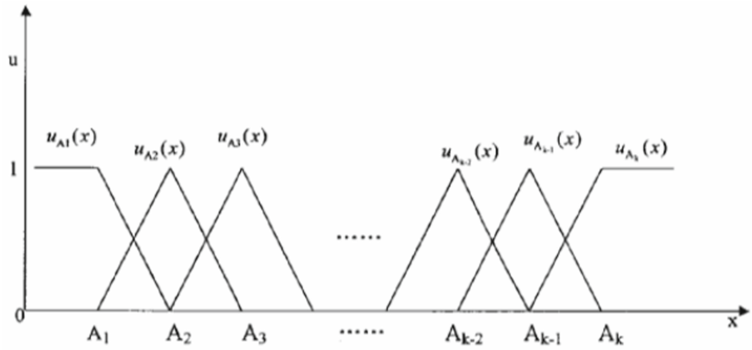

Fig. 1. Triangular membership function graph

3 Application of Intelligent Fuzzy Decision Tree Algorithm in English Translation

Nida also pointed out: "although absolute communication between people is impossible, highly effective communication between people is still possible regardless of the same language region or between different language regions, because people's ideas, physical reactions, cultural experiences and the ability to adjust other people's behavior are similar."

The expression of the target language (or target language) is the ultimate goal of translation. Expression is to reproduce the ideological content and style of the original text to the greatest extent, reproduce the attitude, feelings and writing techniques of the original author, and conform to the expression habits of the target language. Understanding is an extremely complex process with some fuzziness; Expression is not easy, so we need to use various translation skills in order to make the translation conform to the expression habits of the target language.

Proper use of vague language in target language expression can not only improve the accuracy of language expression, but also make language expression euphemistic, implicit and appropriate. To a certain extent, appropriateness in language use is as important as accuracy. We can also say that appropriateness itself is communicative accuracy. To achieve this - point, in the process of expression, we need to comprehensively consider the commonness and individuality in the expression of the source language and the target language, the differences in the understanding of the source language and the target language, the mutual social and cultural background and differences embodied in the source language and the target language, and so on [3].

Language fuzziness has been especially favored by writers since ancient times. Therefore, in the treasure house of Chinese literature and English literature, literary fuzziness is a common means for writers to describe characters, scenery and character. For example, Song Yu's words used to describe beauty in the lustful Fu of dengtuzi are very popular because of the fuzziness of his language: "an increase of one cent is too long, and a decrease of one cent is too short. The powder is too white, and Shi Zhu is too red." after talking for a long time, I don't know how long, how short, how white and how red it is, but it is very vivid and thought-provoking. The expression of the target language should reproduce the thoughts, feelings, attitudes, positions, rhetorical devices and writing style of the source language author as “equivalent” as much as possible, but also reflect the social, cultural and contemporary characteristics contained in the source language to the greatest extent, and conform to the expression habits of the target language, so as to make a good translation.

Machine translation systems can be divided into different types according to different standards. According to the use environment of machine translation systems, machine translation systems can be divided into three categories: low-end machine translation systems, custom high-end machine translation systems and Internet-based machine translation systems. The target customers of low-end machine translation systems are individuals, while the target customers of custom high-end machine translation systems are companies and bases The machine translation system based on the Internet is a machine translation system used through the Internet.

Fuzziness of lexical data in English Translation

For continuous data, the commonly used discrete methods are: equidistant segmentation method, which divides the English translation data into n equidistant parts, and the equidistant parts have the same distance, but this method has a great impact on the collection of outliers and can not deal with inclined data well; equifrequency segmentation method, by dividing the data into equal parts, each equal part has the same amount of data, this algorithm It has good scalability and is suitable for data classification; the histogram method divides the data into thousands of buckets and represents each bucket with the average value or sum in the bucket; the cluster analysis method is based on “the maximum similarity within the class and the minimum similarity between classes” This paper uses the cluster center point to represent the objects contained in this class. This method is very effective and meets many actual data conditions.

4 Processing of Fuzzy Information in English Translation

Literal translation is the most common method to deal with fuzzy information in business translation, that is, using the fuzziness of one language to translate the fuzziness of another language, so as to retain the fuzzy information and realize the equivalence of fuzzy information [4].

The transformation syntax can be used for parsing, that is, the parser. Although the interpretation relationship expressed by transformation grammar simplifies the processing of subsequent stages, transformation rules are designed for deep structure, and it is very complex to use transformation rules to deal with surface structure. In short, there are at least three basic problems:

(1) For a given sentence, it can generate a set of syntactic trees. This set of syntax trees should include all surface trees to which the transformation syntax should be assigned;
(2) For a given tree that does not belong to the base, determine the transformation rules that can be used to generate the tree;
(3) You can judge the conversion rules one by one. If the result is the current tree, this rule will not be executed.

If we deal with these problems in the most direct way, we may have to try many wrong methods. Of course, for the first question, you can use context free syntax to generate transformation syntax. Syntax can be assigned to all surface trees of a sentence, and its number may be large. This syntax is also called overlapping syntax.

5 Conclusion

In the field of professional translation, machine translation technology is being used more and more widely. Therefore, we should change our ideas, improve the teaching conditions of machine translation, strengthen training, improve the level of teachers in machine translation, accumulate strength and compile relevant teaching materials. In addition, we should also cooperate closely with translation companies or enterprises to create conditions for machine translation teaching, actively carry out research and practice in this field and cultivate high professional quality, Talents who master various editing skills and word processing technology, and can use various translation software and modern technology to complete translation tasks quickly and efficiently, so as to serve social development and economic construction.

Acknowledgements. "Big Data Application Research Center for Higher vocational foreign language Education" (19kypt06), a scientific research platform project of Chengdu Polytechnic.

References

1. Jiangtao, L.: At present, the gap of translation talents in China is as high as 90%. World Educ. Inform. **4**, 5 (2006)
2. Yining, Y.: Translation technology and the cultivation of technical translation talents in China. Chin. Sci. Technol. Trans. **2**, 51–54 (2005)
3. Zhang, Y.: Computer aided translation. Shanghai Science and Technology Translation (1) (2002)
4. Murray, L.V.L.: Computer aided translation technology and translation teaching. Foreign Lang. Circ. **2007** (3), 37 (2007)