

PAPER • OPEN ACCESS

English Semantic Recognition Based on Abstract Meaning Representation and Vector Space Model

To cite this article: Yu Han Gao 2018 *J. Phys.: Conf. Ser.* **1087** 052017

View the [article online](#) for updates and enhancements.

You may also like

- [Dynamic magnetic susceptibility and electrical detection of ferromagnetic resonance](#)
Yin Zhang, X S Wang, H Y Yuan et al.
- [Resonant precession of magnetization and precession—induced DC voltages in FeGaB thin films](#)
Prabesh Bajracharya, Vinay Sharma, Anthony Johnson et al.
- [ACCRETION SHOCKS IN CLUSTERS OF GALAXIES AND THEIR SZ SIGNATURE FROM COSMOLOGICAL SIMULATIONS](#)
Sandor M. Molnar, Nathan Hearn, Zoltán Haiman et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

English Semantic Recognition Based on Abstract Meaning Representation and Vector Space Model

Yu Han GAO

College of Computer Science and Technology, Jilin University, Changchun, Jilin
130012, China

gaoyuhan0923@foxmail.com

Abstract. With the development and application of computer science and technology innovation, artificial intelligence industry has achieved sharp development, and its technology has been widely used in many fields, such as natural language processing, robot manufacturing, semantic recognition, information system development, etc. In terms of natural language processing, the recognition and understanding of human language based on artificial intelligence has become the focus of research. And as semantic recognition and processing play an important part of natural language processing, they have achieved satisfactory results in the Abstract Meaning Representation (AMR) and Vector Space Model (VSM) development and application. Based on this, this paper takes the English recognition as the research object, adopts the scientific research methods such as literature analysis, practice investigation and experimental research, further analyzes English semantic recognition based on AMR and VSM from the aspects of semantic recognition difficulty, principles of AMR and VSM and its application in English recognition, as well as the practical applications. The author hopes to provide references for future research.

1. Introduction

As the implied meaning of the application process of natural language, semantics has a pivotal position in the dissemination of information. In the context of today's information age, with the continuous development and application of Internet technology, computer technology, information technology and other scientific technologies, achievement of the semantic identification and understanding has become the major issue based on artificial intelligence technology under the natural language processing, which is of great significance to strengthen the application of language diversification, and to optimize language mechanical system. Because of the existence of polysemy and ambiguity in semantics, its recognition and understanding is an important challenge in the field of natural language processing, and it is also the focus of language resource construction and application. It can be seen that the study of semantic recognition in English is of great significance and practical significance.

In recent years, with the improvement of information resources, the research on the theory and practice of natural meta-processing has made some progress, such as digital signal transmission, naming entity identification, semantic annotation, English mechanical translation optimization, semantic excavation and other related theories and practice. Since Homer Dudley [1] invented the Vocoder in 1939, digital language has been widely used in the field of communications. In recent years, the development of network technology has promoted the application of network information resources, and how to effectively tap the network information resources requires semantic analysis techniques and mining methods for text clustering and mining. The sentence-level semantic representation Abstract Meaning Representation (AMR) [2] and Vector Space Model (VSM) based on text feature have unified



the semantic annotation specification, improved text poly class method [2], and provided technical support for the computer language processing. On this basis, the relevant research on AMR, VSM application has caused stir among scholars at home and abroad [3, 4, 5, 6, 7]. However, there are little relevant research of AMR and VSM under the English recognition. Therefore, this paper explores the AMR and the English semantic identification of VSM in the literature analysis.

2. Difficulties in Semantic Recognition

Due to the polysemy and ambiguity in semantics, the accuracy and authenticity of semantic recognition in natural language processing are relatively low. There are often errors in word meaning, systematic misreading and grammatical problems, which is not conducive to text retrieval and translation.

2.1. Understanding of words

Wrong understanding of words is one of the common problems in English semantic identification in natural language processing. This paper analyzes the English word comprehension with the actual case of English mechanical translation equipment. It has found that there are always misunderstandings for some of the polysemous words, adjectives and abstract nouns.

2.2. Misreading

In the web search, since the retrieval system semantic recognition function is low, there are often misreading and other issues. If the phrases and sentences need to be retrieved are not cut off properly, there will be inaccurate search results and other problems.

2.3. Grammatical problem

Mistaken semantic recognition and understanding are most likely to have grammatical errors. As for the retrieval system, under the absence of logical recognition, it will lead to misunderstanding of semantics, resulting in a retrieval error or a translation error.

3. The Application of AMR in English

Abstract Meaning Representation (AMR), as an emerging semantic representation of sentences, mainly refers to the conversion of semantic abstraction of an English sentence into "single directed acyclic graph". That is, the AMR specification is established and the Smatch metric method [8] is used to predict and output the structure of the AMR according to the sentence, so as to achieve sentence semantic analysis.

Take the sentence "He tries to affect a British accent" as an example (See Figure 1), AMR abstracts the words in the sentence to form a semantic concept node; The abstract relationship between the real word and the real word is transformed into an "arc" with a certain semantic relation mark and direction. On this basis, the morphological changes and the semantics of the function words are ignored, which to a certain extent, realizes the reservation of the sentence body structure, and effectively complements the implied and omitted part of the sentence, so as to achieve the maximum extent of the sentence and to comprehensively describe the sentence semantics, as well as to promote the automated formation of it.

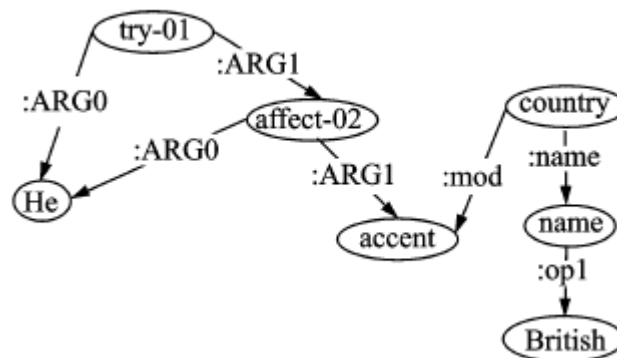


Figure 1. An AMR example “He tries to affect a British accent.”

4. Vector Space Model

The Vector Space Model (VSM) is a text retrieval system that was proposed and applied by Salton et al. [9] in the 1970s. Its working principle is to construct the VSM to simplify the text content in the vector space, and to use the spatial similarity to achieve the text semantic similarity performance according to the vector operation, which makes the text easier to be understood and identified. At the same time, by transforming the document into the text space vector, along with the vector calculation, the text clustering in a short period of time can be achieved, and similar text content can be retrieved.

In general, the VSM is an algebraic model applied to information retrieval, filtering, clustering and similarity evaluation. The main contents of it are as follows: Firstly, it has multiple disordered feature "ti"; Secondly, applying a feature term vector to represent a text keyword, either a phrase or a root; Thirdly, the weights of the vector are calculated and similarity is compared with training documents N; Finally, optimized results can be obtained through cosine calculation or inner product calculation.

5. Future Challenges

The rapid development of Internet technology has promoted the popularization of the Internet application. According to the survey, by the year of 2014, the number of Internet users from Asia, Africa and other regions has been increasing, whose population has exceeded 50% of the global Internet users, which is twice the value of the ones from Europe and the United States [10]. In addition, the process of economic globalization is accelerating, and the application language of network information shows the diversified development trend of language. In this context, the contradiction of the diversity of network information language and the simplification and limitation of user to grasp the language becomes more prominent, which puts forward higher natural language processing requirements. In order to realize the excavation and utilization of multi-language information resources, the innovation and perfection of language recognition technology become the inevitable potential of development.

Alzahrani, Salim [11] and others pointed out that in the application of language information resources, people should pay attention to multi-language text semantic annotation. By extracting the relationship between the content of the web page and the text concept and applying the traditional knowledge organization system, the multi-language search based on the website is realized. At the same time, Perez et al. [12] pointed out in the related research of RDF dataset that based on the analysis of the characteristics of the data source, the establishment of the relevant data model and the generation of the relevant data language recognition and coding are needed to realize the multi-linguistic development of the related data. It can be seen that in the future development, language recognition technology should meet the needs of multi-language.

6. Applications

The semantic recognition technology and method based on Abstract Meaning Representation and Vector

Space Model are widely used in many fields, such as website information retrieval system construction, English mechanical equipment translation and speech coding. In terms of machine translation, it is an important technique applied in the field of multilingual information retrieval, and it is a unified process of text based on certain rules.

The application of AMR provides a new idea for the improvement of English translation of the machine. By weakening the grammatical rules, the semantic information feature is identified, and the VSM algorithm can be used to realize the details of the language processing to a certain extent and improve the accuracy of the English machine translation. For example, as the dealer of software development and artificial intelligence in the world, Google devotes vast human and financial resources to develop an effective and efficient translation application, called Google Translate. Google Translate utilizes both semantic recognition and neural networks to achieve high accuracy. Considering the complex of grammar and wording scheme in different languages, there is still a long way in the machine translating so far.

Another application of AMR is about the emotion recognition. Lots of social media platforms, like Facebook and Twitter, are working on the emotion recognition to deplore the true interests of their users. Generally, they may implement algorithms in semantic recognition and machine learning, especially deep learning, to dig useful information.

7. Conclusion

Based on the analysis of literature, this paper studies the semantic recognition problem based on Abstract Meaning Representation and Vector Space Model. Based on the common problem of semantic recognition, the principle, structure and application of AMR, VSM are analyzed, and the development direction of semantic recognition technology is put forward. At the same time, its application is analyzed in detail with the assistant of actual cases. From the above analysis, it is found that AMR and VSM can enhance the semantic recognition and quality of understanding in English to a certain extent, indicating the importance of semantic structure in practical application. In the future development, the relevant theory and model innovation and the practical application of technology need to be further improved. On the basis of exploring the semantic structure and rules of natural language processing, we can realize the automatic recognition of semantic relations among text keywords.

References

- [1] Dudley, H. (1939). Remaking speech. *The Journal of the Acoustical Society of America*, 11(2), 169-177.
- [2] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., ... & Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 178-186).
- [3] Su Y., Zheng C., & Ma Z.J. (2011). Semantic-based VSM model improvement. (Doctoral dissertation).
- [4] WANG J., LIU B.G., FEI Z.S., ZHANG J.Y. (2016). Anti-packet loss algorithm for mobile network voice transmission based on AMR-WB codec. *Journal of Beijing Institute of Technology*, 36 (10), 1048-1052.
- [5] Sun J., Wu Z.Y., & Du T. (2004). Parametric layer coding conversion scheme from AMR to G.729. *Information and Electronic Engineering*, 2 (1), 17-20.
- [6] Guo Q.L., Li Y.M., Tang Q. (2008). Research on Text Similarity Computation Based on VSM. *Computer Application Research*, 25 (11), 3256-3258.
- [7] Zhao D.P., Xiao T.J., Shi J., & Jiang Z.G. (2014). A Method of Identifying Airport and Oil Depot Targets Based on Significant Semantic Model. *Journal of Computer Aided Design & Computer Science*, 26 (1), 47-55.
- [8] Cai, S., & Knight, K. (2013, August). Smatch: an Evaluation Metric for Semantic Feature Structures. In *ACL* (2) (pp. 748-752).
- [9] Salton, G. (1970). Automatic processing of foreign language documents. *Journal of the*

- Association for Information Science and Technology, 21(3), 187-194.
- [10] Duggan, M., & Smith, A. (2015). Social media update 2013. Pew research center, 19.
 - [11] Alzahrani, S., & Salim, N. (2010). Fuzzy semantic-based string similarity for extrinsic plagiarism detection. *Braschler and Harman*, 1-8.
 - [12] Pérez, J., Arenas, M., & Gutierrez, C. (2010). nSPARQL: A navigational language for RDF. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4), 255-270.