

論 文 内 容 要 旨

報告番号	甲 先 第 332 号	氏 名	韓 朝
学位論文題目	Research on Knowledge Discovery and Affective Computing for Short Text Processing (短文テキスト処理のための知識発見と感情計算に関する研究)		
<p>内容要旨</p> <p>Short text is often used in the QA system, and the processing method of short text has an important influence on the performance of the QA system. This thesis deals with short text information from three aspects: semantics, knowledge and emotion. Details are:</p> <p>[1] Because of the uncertainty of both the language representation and the knowledge representation, the current methods for short text processing are not very effective. To solve the uncertainty of knowledge representation, a rough set knowledge discovery method for Chinese short text QA system is proposed. It uses the method of rough set equivalence partitioning to represent the rough set knowledge of the QA pairs, then uses the idea of attribute reduction to mine out the upper approximation representations of all the knowledge items. Based on the rough set QA knowledgebase, the knowledge match value of a QA pair can be calculated as a kind of knowledge item similarity. After all the knowledge similarities of one question and its answer candidates are given, the final matching values which combines rough set knowledge similarity with traditional sentence similarity can be used to rank the answer candidates. The experiment shows that the proposed method can improve the MAP and MRR compared with the baseline information retrieval methods.</p> <p>[2] A novel method for Short Text Information Retrieval based Chinese Question Answering is proposed. It is developed from the Discernibility Matrix based Rules Acquisition method. Based on the acquired rules, the matching patterns of the training QA pairs can be represented by the reduced attribute words, and the words can also be represented by the QA patterns. Then the attribute words in the test QA pairs can be used to calculate the matching scores. The experimental results show that the proposed representation method of QA patterns has good flexibility to deal with the uncertainty caused by the Chinese word segmentation, and the proposed method has good performance at both MAP and MRR on the test data.</p> <p>[3] The accurate extraction of knowledge subject is not only one</p>			

of the important processes affecting the matching accuracy of the QA system based on knowledge base, but also one of the important processes of knowledge subject positioning based on the knowledge map. A sequence labeling method for knowledge subject analysis for short text KBQA is proposed. From the perspective of rough set model and rough set attribute importance, combined with the existing named entity recognition and syntactic subject analysis, the sequence labeling method is used to optimize the sequence labeling results for knowledge subject extraction, and thus improve knowledge subject analysis ability of the overall system. The experiment results verify the validity of the method.

[4] In short text KBQA, the performance of knowledge predicate analysis can affect the overall matching result of knowledge triple. The knowledge predicate analysis of Chinese short text question is difficult because of the uncertainty of Chinese knowledge predicate representation. Based on the rough set theory, a new definition of knowledge predicate analysis of knowledge based question answering was given, and a new method was proposed to analyze the knowledge predicate of question. It can reduce the words which are weakly related with the knowledge predicate, and then the words which are more related with knowledge predicate representation will be used to match the knowledge triples to improve the overall performance of system. The experiment results verify the validity of the method.

[5] A sentence-level sentiment analysis method is proposed to deal with sentiment measurement and classification problems. It is developed from a model called Synthetic and Computational Language Model (SCLM), which represents modifying and modified information respectively using matrices and vectors. In the proposed method, a global modifying matrix of a sentence is constructed and determinant value of this matrix is calculated and adjusted, and then the final value is used as the sentiment value of the sentence. The regression experiment shows that the deviation between the output sentiment and target sentiment does not exceed a class distance of 5-classes. The classification experiment shows that the proposed method has improved most of the performance comparing to the simplified SCLM.