Application of Rough Set as Knowledge in Transfer Learning for Medical Corpus Analysis

Huaze Xie, Yuanyuan Wang, Mohd Anuaruddin Bin Ahmadon, Shingo Yamaguchi

Graduate School of Sciences and Technology for Innovation,

Yamaguchi University,

2-16-1 Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

E-mail: {g501vg, y.wang, anuar, shingo}@yamaguchi-u.ac.jp

Abstract Supervised transfer learning working on natural language processing, we propose a new rough set embedding model. For a query about a given context paragraph from medical documents semantic analysis, requires modeling complex interactions between the context and the query. In our previous work, we proposed a rough set preprocessing work in diabetes medical records. For the standardized diabetes corpus analysis, the RNN with rough set works better. And we expect this model to work on transfer learning. In this paper, we introduce the Rough Set Attention Flow Network, we use the rough set as the knowledge work in an unstandardized diabetes corpus semantic analysis.

Keyword Transfer learning, Rough set, Semantic analysis, Medical document.

1. INTRODUCTION

The recurrent neural network (RNN) and transfer learning have gained significant popularity over the past few years within the natural language processing and documents semantic analysis. In many different tasks and image domains, the end-to-end training system can get more promising results. For the application of NLP in medical, due to the popularity of electronic medical records and the high reliability of medical literature, mechanical learning, deep learning and transfer learning in medically are more useful. This is because most of the medical literature is based on the validated medical theory. Over time, a complete system will be formed. Therefore, we expect to propose a more suitable natural language analysis model for medical documents.

In our previous work, we proposed the semantic analysis model with rough set preprocessing [1]. And for this model, we expect to apply the rough set as the knowledge work in an attention transfer learning flow, so we proved the effectiveness of rough set as a supervised attention model retrieval ability [2]. In order to make our model convenient for practical application. We designed some tests in the medical corpus and improved our model to make it more suitable for machine comprehension (MC). This MC model is improved from the Automatic Scoring (AS) System (Y. Wang et al. 2015) [3] and the BI-Directional Attention Flow (BI-DAF) model (M. Seo et al. 2017) [4]. And this MC test makes our medical semantic analysis model more suitable for answering simple questions or extracting answers from relevant literature. The AS system is improved for medical corpus vector generation.

In this paper, we introduce the Rough Set Attention Flow Network (RS-AFN), a hierarchical multi-stage architecture for modeling the representations of the context paragraph at different levels of granularity (Figure 1). For the neural machine translation by jointly learning to align and translate [5], we improved the retrieval efficiency by rough set preprocessing. Initially, we only expected to separate the overly close word relationship through the rough set. Then we found that rough set can absorb the results after each training and transform it into the target vocabulary before the next training. So we make use of the advantage of the rough set to transform it into knowledge that can be transferred between feature-mapped corpora in transfer learning.

Our rough set attention mechanism offers following improvements to the previously popular attention paradigms. First, our attention layer is inherited from BI-DAF model, attention layer is not used to summarize the context paragraph into a fixed-size vector. Instead, the attention is computed for every time step, and the attended vector at each time step, along with the representations from previous layers, is allowed to flow through to the subsequent modeling layer. This reduces the information loss caused by early summarization [4]. Second, when the rough set as the knowledge transforms in the unlabeled corpora. We designed a selective absorption algorithm to

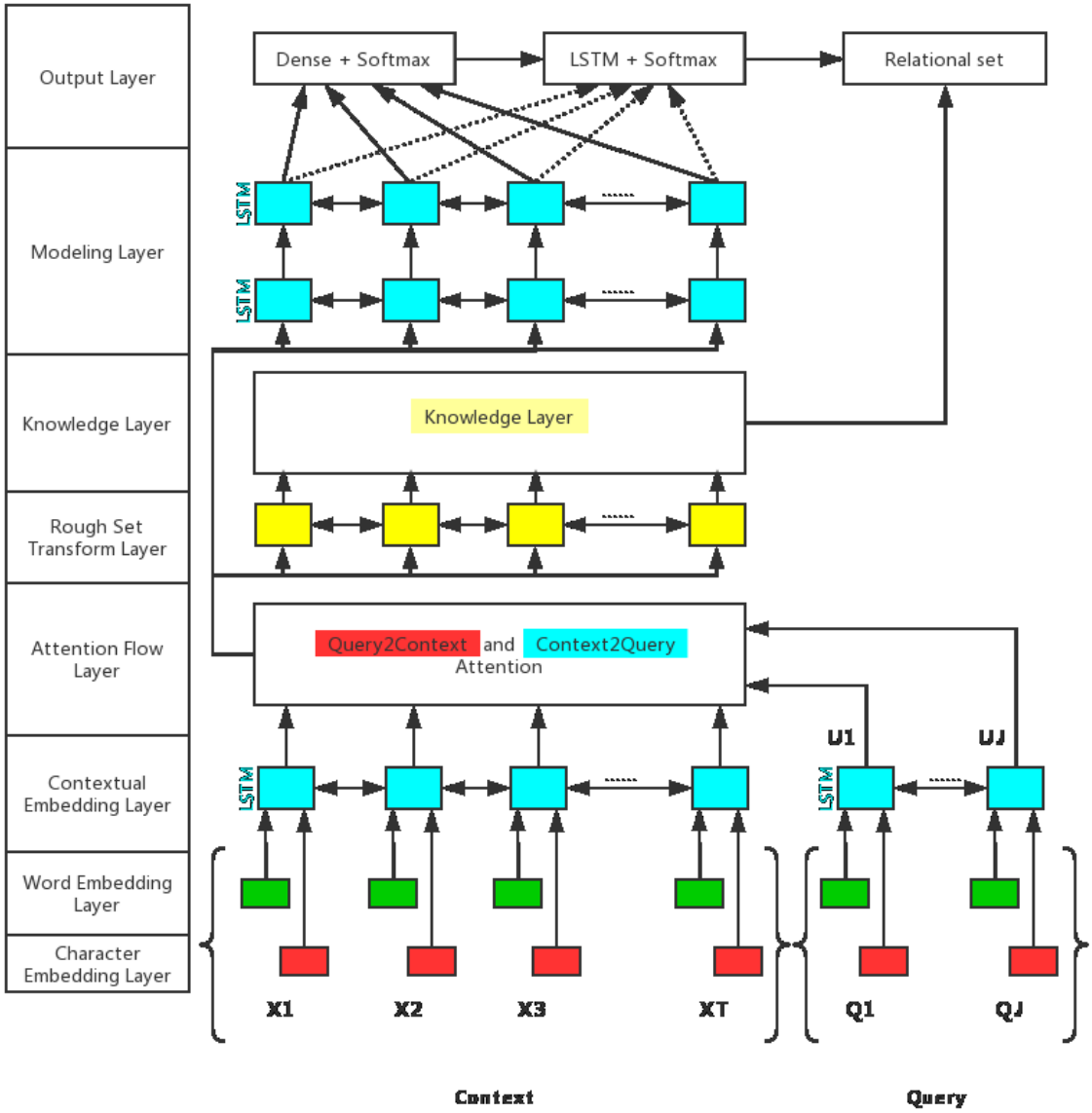


Figure 1: Rough Set Attention Flow Network Model

assist the attention layer, it does not mean we abandoned memory-less attention mechanism, the ordinary attention at each time step is a function of only the query and the context paragraph at the current time step and does not directly depend on the attention at the previous time step. We adopt the absorbability of rough sets to absorb knowledge from the current attention model and transmit it to an automatic scoring system, this process does not affect the normal memory-less attention mechanism working. Third, in the RNN-LSTM training, we continue to use attention mechanisms in both directions, query-to-context and context-to-query, which provide complementary information to each other.

2. PRELIMINARY

A. Rough set

Rough set theory is a mathematical method to describe the incompleteness and uncertainty of data. It can effectively analysis and dispose of inaccurate and incomplete data, finds hidden information from the rough set.

Rough set advantages:

a) It can dispose of all kinds of data, including incomplete data and many variables.

b) It can dispose of deterministic and non-deterministic data.

c) The minimum expression of knowledge and the granularity of knowledge can be obtained.

Rough sets attribute reduction is to eliminate redundant attributes or redundant features and play a role of dimensionality reduction.

B. Transfer learning

Compared with the traditional training model of data mining, transfer learning is a learning model based on a training set. In actual data testing, the data distribution of the training set and test set has big differences. This is also the main reason that even though the training model effect is working better, but the testing effect maybe have too many problems. This is not caused by the over-fitting of the model, the inconsistency between the training set and the test set is the main cause.

In our testing of open source medical data on diabetes, our task requires training about 50,000 medical records. But in fact, our training set collection only comes from about 800 medical documents on diabetes. So we use transfer learning to make our method adapt to the test set.

Different from the general transfer learning model, we decided to start transfer learning after modeling. It means if the test result is not very bad, we expect to improve the accuracy of the model, so we will use the test set with its predicted results as training samples. It is like there is only one correct answer, but the wrong answer is different from each other.

The correct answers given in the test set will accumulate because they are the results of the correct or the same model, but the wrong answers have different error patterns. If the parameters of the model are limited and not over-fitting, the model will eliminate these wrong patterns and tend to the right answers.

It is similar to Figure 1, the basic word embedding is training from the test corpus and the semantic is obtained by training Word2Vec with training corpus and test corpus.

3. METHOD

In our machine comprehension model (Figure 1), we make the processing work in a hierarchical multi-stage and it consists of eight layers:

1) Character Embedding Layer: maps each word to vector space by word2vec

2) Word Embedding Layer: maps rough set target words to another vector space.

3) Contextual Embedding Layer: utilizes contextual cues from surrounding words to refine the embedding of the words. These first three layers are applied to both the query and context.

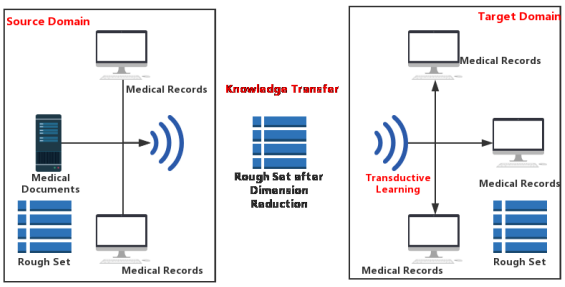


Figure 2: Rough set in transfer

4) Attention Flow Layer: produces a set of query aware feature vectors for each word in the context.

5) Rough Set Transform Layer: the strong correlation is extracted from the mapping set of the target word relationship from the pre-rough set.

6) Knowledge Layer: rearrangement of existing rough set target word relations.

7) Modeling Layer: an RNN-LSTM training to scan the context [1].

8) Output Layer: the answers in the query are sorted out and the knowledge learned in this training is added to the rough set.

4. ROUGH SET IN TRANSFER

In our transfer learning training model shows in Figure 2, we proposed the rough set training method to collect and disseminate target words in medical documents [1]. And the new target words also can be added into rough set [1].

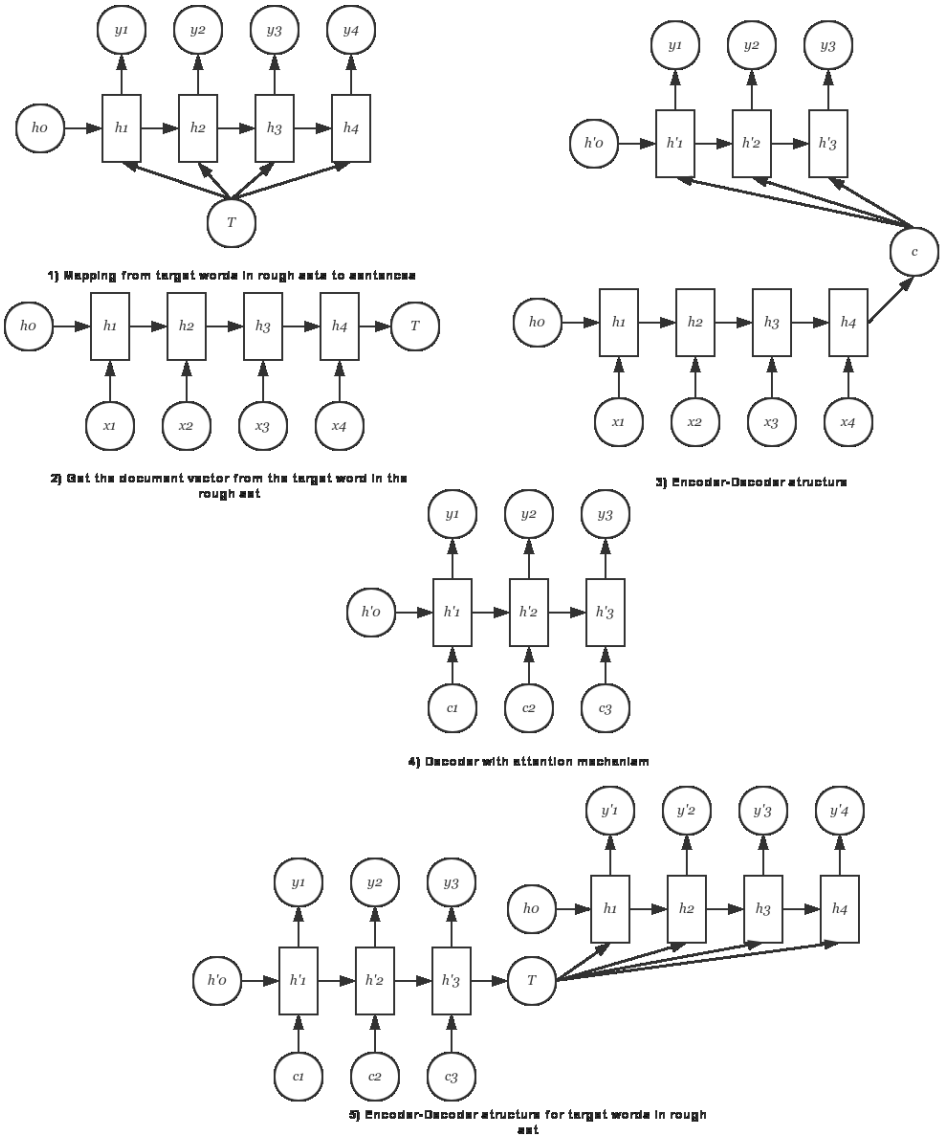
The rough set attention training method task is given by rough set target words at first (see Figure 3. 1)). The input is target words and outputs a sequence, this is like a very simple question answering system by RNN, we can get the output by (1).

$$h_1 = f(U_T + Wh_0 + b), y_1 = \text{Softmax}(Vh_1 + c) \quad (1)$$

Of cause we also need to find out the most relevant words in the output keywords and add them to the rough set (see Figure 3. 2)). We get documents vectors from target words by (2).

$$h_1 = f(U_{x1} + Wh_0 + b), T = \omega(h_1, h_2, h_3, h_4) \quad (2)$$

For training tasks with excessive length, we usually use Encoder-Decoder structure to solve this problem (see Figure 3. 3)). At first Encoder-Decoder structure encodes the input data into a context vector c. After getting c, decode it with another RNN network. This part of RNN



network is called Decoder. This is done by using c as input for each step in Decoder. Because this Encoder-Decoder structure does not limit the sequence length of input and output, we will detect the context containing keywords again to find the hidden attributes of the target words. And we also use Encoder-Decoder to encode and decode the questions in the input medical literature and medical records to get the answers.

In the Encoder-Decoder structure, Encoder encodes all input sequences into a unified semantic feature c , and then decodes them. Therefore, c must contain all the information in the original sequence, and its length becomes the bottleneck restricting the performance of the model. Attention mechanism solves this problem by entering different c at each time. We introduce Decoder with attention mechanism (see Figure 3. 4)). Each c automatically selects the most appropriate context information to y being output. Specifically, we use a_{ij} to measure the correlation between the h_j of phase j in Encoder and the phase i in decoding. Finally, the input context information c_i of phase i in Decoder comes from the weighting of a_{ij} by all h_i.

In the attention mechanism, we can easily apply the it to the semantic learning of medical literature and the medical records question answering system. However, due to the keywords strong relationship in the medical literature, attention mechanism is difficult to separate the keywords, so, it will lead to the difficulty in obtaining the relevance from the target words in the medical paper. So we propose an attention mechanism with rough sets in order to improve the performance of the automatic scoring system (see Figure 3. 5)).

5. RESULT

We trained the rough set attention flow network models in Keras. And we use about 300 labeled medical papers to train our method, we also have 500. All the medical papers are about diabetes, in order to make the training set and the test set are in the same feature space.

Question: How to increase satiety of diabetic patients? **Answer:** incretin hormone glucagon-like peptide 1, GLP1, insulin release. **Context:** The **incretin hormone glucagon-like peptide 1** (**GLP1**) potentiates **insulin release** and suppresses glucagon secretion in response to the ingestion of nutrients. **GLP1** also delays gastric emptying and **increases satiety**.

Form 1: Examples from RS-AFN training model to obtain the correct answer span.

Question: What is the effect of GLP1? **Answer:** normalize **Context:** In patients with **type 2 diabetes mellitus** (**T2DM**), supraphysiological doses of **GLP1 normalize** the endogenous **insulin response** during a hyperglycaemic clamp.
Question: How the GLP1 treatment the T2DM? **Answer:** receptor agonists **Context:** Owing to the short plasma half-life of native **GLP1**, several **GLP1 receptor agonists** (**GLP1RAs**) with longer half-lives have been developed for the **treatment of T2DM**.
Question: What are the genetic diseases caused by diabetes mellitus? **Answer:** diabetes insipidus **Context:** The **genetic** and molecular basis of **familial forms** of **diabetes insipidus** has been elucidated. Seven different **familial forms** of **diabetes insipidus** are known to exist. The clinical presentation, genetic basis and cellular mechanisms responsible for them vary considerably.

Form 2: Examples from RS-AFN training model to observe the causality.

Question: What is diabetes insipidus? **Answer:** diabetes insipidus, diluted urine, polyuria **Context:** **Diabetes insipidus** is a clinical syndrome characterized by the excretion of abnormally large volumes of **diluted urine** (**polyuria**) and increased fluid intake (**polydipsia**).
Question: What is neurohypophyseal diabetes insipidus? **Answer:** neurohypophyseal diabetes insipidus **Context:** The most common type of diabetes insipidus is caused by lack of the antidiuretic hormone arginine vasopressin (vasopressin), which is produced in the hypothalamus and secreted by the neurohypophysis. This type of diabetes insipidus is referred to here as **neurohypophyseal diabetes insipidus**.
Question: What is nephrogenic diabetes insipidus? **Answer:** nephrogenic diabetes insipidus **Context:** The syndrome can also result from resistance to the antidiuretic effects of vasopressin on the kidney, either at the level of the vasopressin 2 receptor or the aquaporin 2 water channel (which mediates the re-absorption of water from urine), and is referred to as **renal** or **nephrogenic diabetes insipidus**.

Form 3: Examples from the result comparison of similar problems.

In the ordinary training methods, there are many definitions of professional nouns in medical theory and training results of diversified information, the ordinary training method usually cannot give reasonable answers or objective information. So there are three kinds of examples given to observe our RS-AFN model training results.

The example test documents are the medical paper from Nature Reviews, here we use the paper: “Glucagon-like peptide 1 in health and disease”, and “Familial forms of diabetes insipidus: clinical and molecular characteristics” to talk our RS-AFN model test results.

From the form 1, we can see the RS-AFN model can give the correct answer span. It means our rough set obtain the correct target words from paper title and abstract.

From the from 2, for the causality from words relationship, our RS-AFN model can successfully grasp the relationship between target words in the rough set.

From the from 3, the similar problems are given in order to observe the same target word or the target word in the other target words like “diabetes insipidus” and “neurohypophyseal diabetes insipidus”, we can see the given answers are from the different paragraph, instead of the same paragraph shows many times.

The results of our RS-AFN training model and the target words with keywords competing for approaches on the medical documents, the visualization by dimensionality reduction of test set are showed in Figure 4.

In Figure 4, the star shows the target word in rough set, and each point is a keyword trained by target word.

We find that by increasing the number of target words in rough set, we can get more keyword output.

And the incompleteness of rough sets can allow strong relational target words to be added.

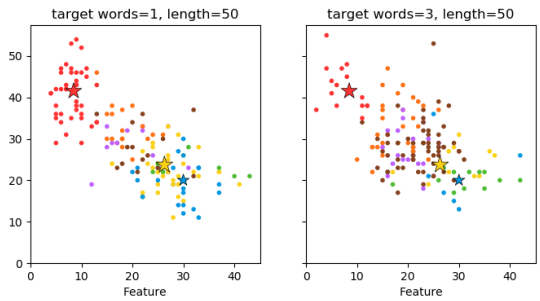


Figure 4: Different initial rough sets in the transfer

6. CONCLUSION

In this paper, we introduce RS-AFN a multi-stage hierarchical process that represents the context at different levels of granularity and uses a rough set attention flow mechanism to achieve a target words context representation. The experimental evaluations show that our model achieves the output of keywords in attention model is controlled by rough sets with target words.

The visualizations and results show that our model is learning the test set containing medical vocabulary representation for MC and is capable of answering complex questions by attending to correct locations in the given target words. And the target words lead the keywords to constitute the relevance of medical vocabulary. This relevance will be used to the electronic medical record. In our future work, we expect to combine our learning model with the actual medical records, accordingly, we can give some advice for the records.

References

[1] H. Xie, M.A.B. Ahmadon, and S. Yamaguchi. "Evaluation of Rough Sets Data Preprocessing on Context-Driven Semantic Analysis with RNN," Global Conference on Consumer Electronics (GCCE), pp.379-382, 2018.

[2] H. Xie, M.A.B. Ahmadon, S. Yamaguchi, and I. Toyoshima. "Random Sampling and Inductive Ability Evaluation of Word Embedding in Medical Literature," International Conference on Consumer Electronics (ICCE), 2019.

[3] Y. Wang, Y. Kawai, S. Miyamoto and K. Sumiya. "An Automatic Scoring System for E-Reports based on Student Peer Evaluation using Groupware," The Database Society of Japan Journal, pp.71-76, 2015.

[4] M. Seo, A. Kembhavi, A. Farhadi and H. Hajishirzi. "Bidirectional Attention Flow for Machine Comprehension," International Conference on Learning Representations (ICLR), 2017.

[5] C. Clark and M. Gardner. "Simple and Effective Multi-Paragraph Reading Comprehension," Annual Meeting of the Association for Computational Linguistics, pp.845-855, 2018.

[6] R. T. Sousa, L. A. Pereira and A. S. Soares. "Predicting Diabetes Disease Evolution Using Financial Records and Recurrent Neural Networks," Machine Learning for Health, 2018.

[7] J. Howard and S. Ruder. "Universal Language Model Fine-tuning for Text Classification," Association for Computational Linguistics, 2018.

[8] R. Volpi, H. Namkoong, O. Sener, J. Duchi and V. Murino, Silvio Savarese "Generalizing to Unseen Domains via Adversarial Data Augmentation," Neural Information Processing Systems, 2018.

[9] D. Xiao, G. Wang, F. Hu. "Fast Parallel Attribute Reduction Algorithm Based on Rough Set Theory " Computer Science, 2009.

[10] M. J. Kusner , Y. Sun, N. I. Kolkin, K. Q. Weinberger. "From word embeddings to document distances," International Conference on International Conference on Machine Learning, 2015.

[11] A.B. Dieng, C. Wang , J. Gao , J. Paisley. "TopicRNN: A Recurrent Neural Network with Long-Range Semantic Dependency" International Conference on Learning Representations, 2017.

[12] T. Mikolov, W. Yih, and G. Zweig. "Linguistic Regularities in Continuous Space Word Representations".North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.746-751, 2013.

[13] T. Mikolov, K. Chen, G. Corrado, and J. Dean. "Efficient Estimation of Word Representations in Vector Space" International Conference on Learning Representations, 2013.

[14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. "Distributed Representations of Words and Phrases and their Compositionality" Neural Information Processing Systems, pp.3111-3119, 2013.

[15] Z. Pawlak. "Ai and Intelligent Industrial Applications: the Rough Set Perspective" Cybernetics and Systems , pp.227-252, 2000.