Answer selection using Word Alignment based on Part of Speech Tagging in community question answering

*Ade Sutedi 1, Moch. Arif Bijaksana 2, Ade Romadhony 3

1,2,3 Department of Informatics, Graduate School of Computing, Telkom University, Indonesia

Email: *adesutedi@sttgarut.ac.id

Abstract. This paper explain about answer selection using word alignment based on POS tagging in Community Question-Answering (CQA). This online community allowed the user to ask and reply related to the question problems which has no restrictions. This causes inappropriate comments with the question problems proposed before. To solve these problems, combining lexical and semantic features has been developed with result conclude that the approach more adequate for similarity task rather than question answering. According to the previous research, there is several problems that can be enhanced. First, vector representation counts exactly matched words, so it does not effective to cover other words that have relatedness between two pairing words. Second, noun overlap for similarity measure in pairing words can’t define that the two words are similar. So, it must be define that the pairing POS tag is the same meaning or relatedness. In this study, unsupervised lexical and semantic similarity method employed with different approach from previous method in verbatim and contextual similarities. The data was taken from SemEval 2017 competition which focus on Question-Answer Similarity task. The experiment result for precision (Mean Average Precision) score shows the improvement from 0.674 to 0.6845, 1.03 % higher than previous research in CQA. This improvement comes from lexical similarity, which is not just from noun pattern but also taken from verb pattern. Furthermore, semantic similarity has an important role in determining which words that have same pattern and meaning to define relevancy between them.

1. Introduction

In recent years, Community Question-Answering (CQA) system becomes popular and increased significantly. CQA was successfully used on many website application such as community forums, social media, and basic website that support forum activities like Qatar Living1, etc. A user in this community can post any question or give comments to the other question that proposed by another user with fewer restrictions. The services have a contribution to user collaboration since there is benefit each other when a user needs help with their problems.

In order to get the appropriate answer, there some problem that faced to determine which answer should be selected. Due to CQA content have a lot of irregular forms, irrelevant mention, typos, and informal language used, the answers can be solved the problem or it just not relevance. In other words, the existing data being not in accordance with the context of the question, which does not provide the

1 https://www.qatarliving.com/forum
right solution. The motivation to this research is how to develop lexical and semantic similarity with different approach according to the previous method [4] to get the pattern between question and answers.

2. Related work

There are so many methods adopted to gain similarity feature in question answering problem. In [9] similarity between question and answer obtained by using Lexical, Syntactical, and Semantic similarity based supervised machine learning approach and a manifold of features including word n-grams, text similarity, sentiment dictionaries, the presence of specific words, the context of a comment, some heuristics, etc. Another research [4] was developed using bidirectional distributed word alignment. It measures similarity by cosine similarity, word overlap, noun overlap, and n-gram feature between question and comment (answer). Combining with the distributed representation of text, word alignment, knowledge graph, and common frame which generated from BabelNET semantic network and FrameNET lexical database.

Semantic learning by parse tree kernel of text pair (intrapair and interpair) from sentence has been proposed by [8]. It can evaluate the syntactic similarity between two texts and capture emerging pairwise patterns. Single-relation queries approach [7] has been developed to answer any question in open-domain question system. This method developed and adapted from ReVerb system which automatically identifies and extracts binary relationships from English sentences which called PARALEX. Another method was developed by combining the syntactical pattern and word alignment which can minimize informal text and avoid parsing error without using parsing [5].

3. Methodology

In order to find similarity and relevancy between two pairing text, this research focus on two different levels in lexical and semantic similarity, which was adopted from previous research [4]. The state of the art in this research can be found for several stage which distinguish between this research and previous research. This stage describes on Figure 1:

![Flow diagram for lexical and semantic similarity](image)

**Figure 1.** Flow diagram for lexical and semantic similarity

3.1. Lexical similarity

In this section, lexical features such as Tokenization, Stopwords, and POS Tagging will apply to get a relation between two pairing texts. The POS tag pattern result from pairing text between question and answer will evaluate using the Jaccard similarity, which calculates sets of tags that related each other. This measure was used to overcome the lack of vector representation [1] in lexical form. If Jaccard similarity scores more than zero it will evaluate by synset similarity, otherwise it will process to next stage of alignment based on POS tag.

3.2. Semantic similarity

The result of POS tag pattern with the lexical structure such as a noun, verb, adjective, and adverb was evaluated by WordNet. Based on [3] similarity can be applied to a noun and verb structure. This process will determine synonym sets (synsets) of words that have the same meaning and relation. So, it defines
that the pairing text only within classes of words with the same part-of-speech [13]. For example the pairing text:

\[ q = \text{"Do you know how long it will take to get Schengen Visa from Embassy of Greece? 3 days? 1 week? any idea? Thanks."} \]
\[ a = \text{"Standard is 2 weeks maximum, but when I apply my schengen visa from Hungarian Embassy took 5 weeks."} \]

We applied synset similarity based on [3],[13] for any \( q \) and \( a \) sentences with result:

\[ q = [(\text{\textquoteleft know\textquoteleft}, \text{\textquoteleft VB\textquoteleft}), (\text{\textquoteleft take\textquoteleft}, \text{\textquoteleft VB\textquoteleft}), (\text{\textquoteleft visa\textquoteleft}, \text{\textquoteleft NN\textquoteleft}), (\text{\textquoteleft embassy\textquoteleft}, \text{\textquoteleft NN\textquoteleft}), (\text{\textquoteleft greece\textquoteleft}, \text{\textquoteleft NN\textquoteleft}), (\text{\textquoteleft week\textquoteleft}, \text{\textquoteleft NN\textquoteleft})] \]
\[ a = [(\text{\textquoteleft weeks\textquoteleft}, \text{\textquoteleft NNS\textquoteleft}), (\text{\textquoteleft apply\textquoteleft}, \text{\textquoteleft VBP\textquoteleft}), (\text{\textquoteleft visa\textquoteleft}, \text{\textquoteleft NN\textquoteleft}), (\text{\textquoteleft Hungarian\textquoteleft}, \text{\textquoteleft NNP\textquoteleft}), (\text{\textquoteleft Embassy\textquoteleft}, \text{\textquoteleft NNP\textquoteleft}), (\text{\textquoteleft took\textquoteleft}, \text{\textquoteleft VBD\textquoteleft}), (\text{\textquoteleft weeks\textquoteleft}, \text{\textquoteleft NNS\textquoteleft})] \]

The score result is:

- Get (NN) \( \rightarrow \) Visa (NN) \( \rightarrow 0.307692307692 \)
- Schengen (NN) \( \rightarrow \) nothing in database
- Visa (NN) \( \rightarrow \) Embassy (NN) \( \rightarrow 1.0 \)
- Embassy (NN) \( \rightarrow \) Hungarian (NN) \( \rightarrow 0.3 \)
- Greece (NN) \( \rightarrow \) Weeks \( \rightarrow 0.769230769231 \)
- Days (NN) \( \rightarrow \) Weeks (NN) \( \rightarrow 1.0 \)
- Weeks (NN) \( \rightarrow \) Visa (VB) \( \rightarrow 0.307692307692 \)
- Idea (NN) \( \rightarrow \) Visa (VB) \( \rightarrow 0.615384615385 \)
- Thanks (NN) \( \rightarrow \) Apply (NN) \( \rightarrow 0.5 \)
- Know (VB) \( \rightarrow \) Took (NN) \( \rightarrow 1.0 \)

### 3.3. POS tag alignment

The pairing texts segmented by tokenization and filtered out by stopwords. After that process, the pairing texts will generate to the POS tag pattern then for each set of pattern computes using Jaccard similarity and synsets similarity. This similarity will determine the POS tag alignment process in order to find the semantic similarity between question and answers. In the alignment result, the highest score indicates that the pairing words are matching each other. It allows to align position such as 1-to-1, 1-to-0, 0-to-1, 1-to-many, many-to-1, and many-to-many alignment. Until this stage, the pairing text should have relevance between question and answers.

The words with highest similarity score indicate that similar each other. For any words that have not similarity score will be ignore and do not use in the alignment process.
The alignment process in figure 2, produce some word position related to their similarity in pairing sentences based on POS tag. At the end, the pairing sentence formed by lexical similarity and semantic similarity to get the important word which focus on primary of sentence in query of question and answer relevancy. In the alignment result, the highest score indicates that the pairing words are matching each other.

3.4. Bigrams and TF-IDF score
According to McKeown and Radev, (2000) in [2], “Collocation is generally defined as a group of words that occur together more often than by chance”. This process will produce simple combination of bigram paraphrase associated with lexical structure. In order to measure occurring words [4], TF-IDF will used to compute probability of bigrams of POS tag pattern that occurs in text.

4. Experiment and result

4.1. Experiment
In this research, the data was taken from SemEval 2017 competition which describes on the table 1.

| Corpus     | Original Question | Related Question | Answer  |
|------------|-------------------|------------------|---------|
| Train Part1| 200               | 1999             | 19990   |
| Train Part2| 67                | 670              | 6,700   |
| Dev        | 50                | 500              | 5000    |
| Test       | 293               | 293              | 2,930   |

The experiment was carried out by comparing two groups of data namely data relevance (IR) and data prediction (SYS) then process by MAP scorer. This process aims to produce a ranking of each answer that has relevance to a particular question. The evaluation measure from SemEval official provides a number of scores to assess the quality of the output of a system or tools. It is called Mean Average Precision (MAP), the baseline that will evaluate the 10 ranked answers which relevance to the question.

4.2. Result
In this section, the experiment result will measure in several evaluation for “Good” and “Bad” answer ranking in thread (c1, ..., c10) according to their relevancy in question-answer similarity. The input file taken from question sentence with queue of answer sentence related to each thread. Both file compared by lexical and semantic similarity according to their relevance. If they have similarity score, it’s indicate that the pairing sentence have similarity and relevancy to be candidate of the “Good” answer and the other answer is “Bad”.

Due to data relevance was carried from official, then in this study only focus on producing predictive data to gain MAP score with the result:
Table 2. Comparison of the classification result

|        | Official Score | Previous Research | The Experiment |
|--------|---------------|-------------------|---------------|
| Accuracy | 0.4332        | 0.682             | 0.528         |
| Precision | 0.3444        | 0.526             | 0.6167        |
| Recall  | 0.7641        | 0.545             | 0.2429        |
| F1 Score| 0.4747        | 0.527             | 0.3486        |

Table 3. Comparison of the ranking result (MAP)

|        | Official Score | Previous Research | The Experiment |
|--------|---------------|-------------------|---------------|
|        | IR            | SYS               | IR            | SYS        |
| MAP    | 0.538         | 0.456             | 0.632         | 0.6762     | 0.726 | 0.6845 |
| AvgRec | 0.728         | 0.654             | 0.812         | 0.795      | 0.793 | 0.770  |
| MRR    | 63.13         | 53.50             | 72.5          | 77.1       | 82.37 | 76.77  |

Conclusion

After the experiments process, it concludes that the research has several improvements. The experiment result for precision (Mean Average Precision) score shows the significant improvement from 0.6742 to 0.6845, 1.03 % higher than previous research in CQA. Based on the previous section, the conclusion can be stated as follows:

- The weakness of vector representation in lexical similarity can be improved by changing to a set of POS tag, then compute the similarity using Jaccard similarity.
- Similar POS tag in pairing text can be defined by WordNet to prove that the pairing text in similar meaning.
- Relevancy between two pairing text can be defined by POS tag alignment to prove that the pairing text is similar to each other.

Due to semantic similarity depend on POS tag feature, it’s difficult to define how similar between two pairing word that caused by typo, non english language, and empty data. For example word “take”, which is verb type became “taake” that haven’t in POS tag. It becomes the opportunity to be developed in the future with another features methods.

References

[1] A. I. Zhiguo Wang, 2015 FAQ-based Question Answering viaWord Alignment
[2] Z Liu, H Wang, H Wu, S Li 2009 Collocation Extraction Using Monolingual Word Alignment Method no August pp 487–495
[3] T Pedersen, J Michelizzi 1998 WordNet: Similarity - Measuring the Relatedness of Concepts HLT-NAACL--Demonstrations '04 Demonstr Pap HLT-NAACL 2004 no July pp 38–41
[4] M Franco-Salvador, S Kar, T Solorio, P Rosso 2016 UH-PRHLT at SemEval-2016 Task 3: Combining Lexical and Semantic-based Features for Community Question Answering in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) pp 814–821
[5] K Liu, L Xu, J Zhao 2013 Syntactic Patterns versus Word Alignment: Extracting Opinion Targets from Online Reviews Proc 51st Annu Meet Assoc Comput Linguist (Volume 1 Long Pap pp 1754–1763

[6] J Chang, H Chen 1994 Using Part-of-speech Information in Word Alignment in AMTA pp 16–23

[7] A Fader, L Zettlemoyer, O Etzioni 2013 Paraphrase-Driven Learning for Open Question Answering Acl pp 1608–1618

[8] S Filice et.al 2016 Learning Semantic Relations between Questions and Answers in Proceedings of SemEval-2016 San Diego California

[9] M Nicosia et.al 2015 Answer Selection for Community Question Answering – in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) Denver Colorado

[10] P Nakov, L Marquez, W Magdy, A Moschitti J Glass and B Randeree 2015 SemEval-2015 Task 3: Answer Selection in Community Question Answering in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) Denver Colorado

[11] P Nakov, D Hoogeveen 2017 SemEval-2017 Task 3: Community Question Answering

[12] Miller George A 1995 WordNet: A Lexical Database for English Communications of the ACM Vol 38 No 11: 39-41

[13] R Mihalcea, C Corley, C Strapparava 2006 Corpus-based and knowledge-based measures of text semantic similarity Proc 21st Natl Conf Artif Intell vol 1 pp 775–780