International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 13, Issue 9 (September 2017), PP.18 - 21

Semantic Textual Similarity for Machine Translation Evaluation

[1]A. Sai Hanuman, [2]LikhitaChintala
[1]Department of Computer Science and Engineering.
[2]GokarajuRangaraju Institute ofEngineering and Technology,Hyderabad, India
Corresponding Author: A. Sai Hanuman

ABSTRACT:- Machine translation translates a speech of text from the source language to target language. This paper introduces machine translation evaluation by calculating the semantic textual similarity between the machine translated sentences. The similarity score varies by using different values of alpha, beta and ranges semantic similarity [0,5]. The experiment is carried out on SemEval 2017 datasets. The experiment resulted in the highest accuracy for the Spanish-Spanish dataset with Pearson coefficient correlation 0.7969.

Keywords:- Sentence similarity, semantic nets, corpus, Machine Translation evaluation, natural language processing.

I. INTRODUCTION

Machine translation is the translation of the text by a computer without human intervention. It can also be referred as automated translation. As the internet opens up the wider multilingual text, research and development in Machine Translation continue to grow at a rapid rate. The Evaluation of Machine Translation (MT) hasbecome important in semantic information. Semantic textual similarity plays an important role in Natural language processing. Semantic Textual Similarity assess the degree of equivalence between two sentences. The semantic similarity score ranges from [0-5].0 indicates non- relevant and 5 indicates relevant. In this paper Semantic similarity measured using Pearson correlation coefficient. Semantic Textual Similarity evaluates in text-related research and applications, in areas such as machine translation, Web page retrieval, and text mining. The paper is organized as follows: Section 2 gives the description of about Literature survey. Section 3 describes the architecture of the system to compute our metric, then Section 4 describes corpus based and lexical based features, and Section 5 presents few results produced with Pearson coefficient correlation, and finally, Section 6 gives an outlook on the conclusion.

II. LITERATURE SURVEY

Yuhua Li et al[1], has explained The use of a lexical database that enables our model to maintain human common sense knowledge and the incorporation of corpus statistics allows our method to be versatile to different domains. Julio Castillo et al [21]presents a new approach to Machine Translation evaluation based on the recently defined task Semantic Textual Similarity. This problem is addressed using a textual entailment engine entirely based on WordNet semantic features. Simone Magnolini et al[22] present a work to evaluate the hypothesis that automatic evaluation metrics developed for Machine Translation (MT) systems have a significant impact on predicting semantic similarity scores in Semantic Textual Similarity (STS) task for English, in light of their usage for paraphrase identification. Liling Tanet al[23] explains the extended work on using machine translation (MT) metrics in the STS task by automatically annotating the STS datasets with a variety of MT scores for each pair of text snippets in the STS datasets.Sarah Kohailet al,[24]gives a clear idea about an unsupervised approach, whichestimates a word alignment-based similarity score, and supervised approach, which combines dependency graph similarity and coverage features with lexical similarity measures using regression methods.

III. SYSTEM DESCRIPTION

An Unsupervised system is used to measure semantic similarity between monolingual and cross lingual sentences. These sentences are translated into English using google translator. The data is preprocessed to get accurate results. The pre-processingsteps include misspelling corrections, lowercase conversion, contraction replacement. Two features corpus based and lexical statistics are generated. These features are combined using unsupervised model. The Unsupervised model calculates a similarity score based on the alignment of the input pair of sentences. Each pair has a score on the scale [0-5].

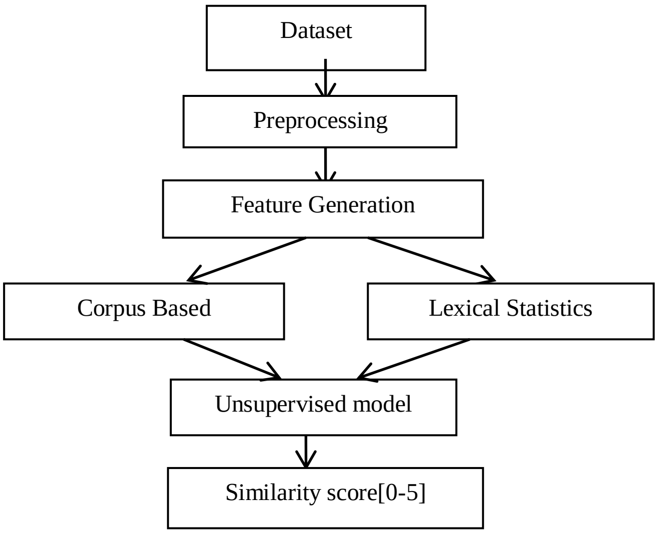

FIG. 1: System Architecture

IV. FEATURES

Semantic features deal with the meaning of the words in the sentences.

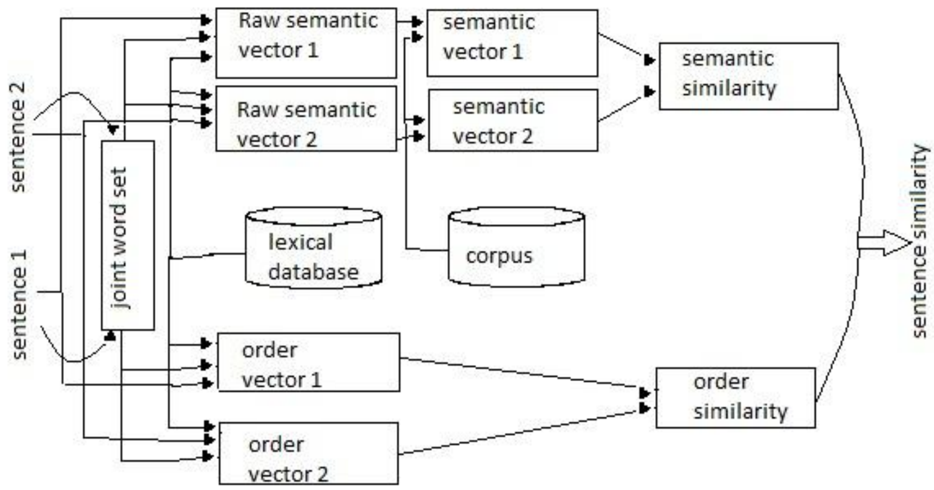

FIG. 2:Computing the sentence similarity between two candidate sentences.

The figure shows the procedure for computing the sentence similarity between two candidate sentences. The method automatically builds a joint word set using all the distinct words in the pair of sentences. For each and every sentence, a raw semantic vector is derived from a lexical database. A word order vector is formed for each sentence, using information from the lexical database [1]. Each word from a sentence conveys dissimilar to the meaning of the complete sentence, the reason of a word is weighted by using information content derived from a corpus. Combining the raw semantic vector from a corpus, and a semantic vector is measured for each of the two sentences. Semantic similarity is evaluated based on the two semantic vectors. An order similarity is measured using the pair of order vectors. Finally, the sentence similarity is calculated by combining semantic similarity and word order similarity. The following sections represent a clear description of each of the above procedure steps. Semantic similarity between words is used both in driving sentence semantic and word order similarity, we will first describe our method for measuring word semantic similarity[1].

A. Knowledge based feature

For finding the semantic similarity between a sentence pair, each sentence is mapped to the unique word vector(UWV) to form semantic vectors. If the unique word (UW_i) is present in the sentence then the i^{th} entry in the semantic vector(SV) is 1 otherwise i^{th} entry in the semantic vector(SV) is the highest semantic similarity value computed between the UW_i and every word in the sentence using $S(UW_i, W_j)$. For computing $S(UW_i, W_j)$ lexical database is used i.e., WordNet.

$$S\left(UW_i,\ W_j\right) = e^{\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}$$

Where l is the shortest path between the words $UW_i$and W_j , h is the depth measured in the WordNet and α, β are the constants. The α value and β values are 0.2 and 0.45 respectively which are found to be best by Li[17].

B. Corpus based feature

The information content values are calculated by incorporating corpus statistic[25]. The corpus statistics is incorporated to make this feature work for various domains. The value at the i^{th} entry in semantic vectors generated using knowledge based feature i.e., $S(UW_i, W_j)$ is normalized by multiplying $S(UW_i, W_j)$ with $IC(UW_i)$ and $IC(W_j)$ to generate normalized semantic vectors (NSV). The IC(w) is the information content of word and it is defined as:

$$IC(w) = 1 - \frac{\log(n+1)}{\log(N+1)}$$

The total number of words in the corpus is N. The term frequency of the word 'w' in the corpus is indicated by 'n'. The semantic with corpus similarity is the cosine value between the normalized semantic vectors of the sentence pair.

V. EXPERIMENTS AND RESULTS

The datasets Arabic-Arabic, Arabic-English, Spanish-Spanish are taken from SemEval 2017. Each dataset contains 250 pairs of sentences. All the datasets are translated into English using Google translator.
The dataset is pre processed before building the model to generate the features correctly. The Preprocessing steps involve in Contraction replacement, Lower case conversion, Spelling correction. Features Corpus based and Lexical statistics are combined using Unsupervised model. Semantic similarity measured using Pearson correlation coefficient.Table 1 depicts similarity score for a pair of sentences the sentences are taken from Arabic – English dataset.

TABLE 1: Similarities between Selected Sentence Pairs

Sentence pair	Sim	sentence pair	Sim
Two men laughing. Two men are crying.	0.5176	A young girl carrying a camera. A girl has binoculars.	0.2966
The girl riding her bike. A bike riding blond child	0.7817	Two men are leading the way to another. Two people ride red bikes.	0.3823
People near the road. People are in the road.	0.7402	A group of traders working in the market. The man works in the local market.	0.5219
The windows open. The windows are tall.	0.5519	Woman wearing glasses. A woman in red wears glasses and long earrings.	0.4722
Police arrest man. A man is being chased by the police.	0.4097	Man outdoor shake park. Skateboarder in midair at a park.	0.2919

The results obtained higher accuracy for Spanish-Spanish dataset. The accuracy varies by changingα, β values.

TABLE 2: Similarity Correlations

Datasets	A	B	Accuracy
Track 1-ar-ar	0.9	0.7	0.7339
Track 2-ar-en	0.9	0.9	0.6372
Track 3-es-es	0.9	0.9	0.7969

VI. CONCLUSION

This paper described Machine Translation evaluation based on Semantic Textual Similarity. Semantic similarity is derived by generating word order similarityand corpus based features. The lexical knowledge base models common human sense knowledge about words in a natural language [1]. A corpus reflects the actual usage of words and language. Thus, our Semantic Textual Similarity (STS) not only focus common human sense knowledge with information, but it is also able to adapt to an application area using a corpus specific to that application.

REFERENCES

[1]. Yuhua Li, David McLean, ZuhairA.Bandar, JamesD.O'Shea, and Keeley Crockett,"3Li.pdf", "Sentence Similarity Based on Semantic Nets and Corpus Statistics", IEEE TRANSACTIONS ON KNOWLEDGE ANDDATA ENGINEERING, VOL. 18, NO. 8, AUGUST 2006

[2]. J. Atkinson-Abutridy, C. Mellish, and S. Aitken, "Combining Information Extraction with Genetic Algorithms for Text Mining," IEEE Intelligent Systems, vol. 19, no. 3, 2004.

[3]. G. Erkan and D.R. Radev, "LexRank: Graph-Based Lexical Centrality As Salience in Text Summarization," J. Artificial Intelligence Research, vol. 22, pp. 457-479, 2004.

[4]. Y. Ko, J. Park, and J. Seo, "Improving Text Categorization Using the Importance of Sentences," Information Processing and Management, vol. 40, pp. 65-79, 2004.

[5]. Y. Liu and C.Q. Zong, "Example-Based Chinese-English MT," Proc. 2004 IEEE Int'l Conf. Systems, Man, and Cybernetics, vols. 1-7, pp. 6093-6096, 2004.

[6]. D. Michie, "Return of the Imitation Game," Electronic Trans. Artificial Intelligence, vol. 6, no. 2, pp. 203-221, 2001

[7]. C.T. Meadow, B.R. Boyce, and D.H. Kraft, "Text Information Retrieval Systems", second ed. Academic Press, 2000.

[8]. P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," Proc. 14th Int'l Joint Conf. AI, 1995.

[9]. Julio Castillo,Paula Estrella,WMT-2012-Castillo.pdf, "Semantic Textual Similarity for MT Evaluation"http://www.mt-archive.info/WMT-2012-Castillo.pdf

[10]. S12-1092.pdf, FBK: Machine Translation Evaluation andWord Similarity metrics for Semantic Textual Similarity Jos´eGuilherme C. de Souza, MatteoNegri, YasharMehdad , http://www.aclweb.org/anthology/S12-1092

[11]. Chris Dyer Jonathan Clark AlonLavie Noah A. Smith ,Alignment.pdf, "UnsupervisedWord Alignment with Arbitrary Features", https://www.cs.cmu.edu /~jhclark/pubs /alignment.pdf

[12]. Menno van Zaanen and Simon Zwarts ,"b9a2b40015ce762ddcecbb5e795cd43541ef.pdf", "Unsupervised Measurement of Translation Quality using Multi-Engine, Bi-Directional Translation", https://pdfs.semanticscholar.org/95f2/b9a2b40015ce762ddcecbb5e795cd43541ef.pdf

[13]. Tatiana Gornostay, "Machine Translation Evaluation", "Machine_Translation_Evaluation.pdf".

[14]. SemEval-2017 Task 2: Multilingual and Cross-lingual SemanticWord Similarity Jose Camacho-Collados*1, Mohammad TaherPilehvar*2, Nigel Collier2 and Roberto Navigli1,https://www.researchgate.net/ publication/228368890 Using_machine_translation_ evaluation_techniques_todetermine_sentencelevel_semantic_equivalence?enrichId=rgreq,

[15]. ErcanCanhasi ,"Paper_5.pdf", "Measuring the sentence level similarity".

[16]. Chinmay More2, AbhishekKulkarni1 , Mayur Kulkarni3& Vishal Bhandekar4,"268.pdf", "Text Analytic Tools for Semantic Similarity", http://www.onlinejournal.in

[17]. P.W. Foltz, W. Kintsch, and T.K. Landauer, "The Measurement of Textual Coherence with Latent Semantic Analysis," Discourse Processes, vol. 25, nos. 2-3, pp. 285-307, 1998.

[18]. C. Burgess, K. Livesay, and K. Lund, "Explorations in Context Space: Words, Sentences, Discourse," Discourse Processes, vol. 25, nos. 2-3, pp. 211-257, 1998.po

[19]. ShubhamGautam,Text Entailment and Machine Translation Evaluation, http://www.cfilt.iitb.ac.in /resources/surveys/te-mt_shubham-june14.pdf

[20]. AlonLavie, AMTA 2010 Tutorial, Evaluating the Output of Machine Translation Systems, https://amta2010.amtaweb.org/AMTA/papers/6-04-LavieMTEvaluation.pdf

[21]. Julio Castillo,PaulaEstrella," Semantic Textual Similarity for MT evaluation",p52-castillo.pdf

[22]. Simone Magnolini, Ngoc Phuoc An Vo, Octavian Popescu," Learning the Impact of Machine Translation Evaluation Metrics forSemantic Textual Similarity", R15-1052.pdf

[23]. Liling Tan, Carolina Scarton, Lucia Specia,and Josef van Genabith," Semantic Textual Similarity with Machine Translation Evaluation Metrics and (eXtreme) Boosted Tree Ensembles", S16-1095.pdf

[24]. Sarah Kohail," Scoring Semantic Textual SimilarityUsing Supervised and Unsupervised Ensemble", unsupervised MT.pdf

[25]. V. Sowmya, Vishnu Vardhan B, BhadriRaju M S V S," Importance of Syntactic Features for Semantic Textual Similarity "

A. Sai Hanuman. "Semantic Textual Similarity for Machine Translation Evaluation." International Journal of Engineering Research and Development, vol. 13, no. 09, 2017, pp. 18–21.