

# **Modelling the Semantic Significance in Non-Factoid Question Answer Pairs in Online Discussion Forums Based on Deep Belief Networks**

M.V.P.T. Lakshika

*Department of Science and Technology, Uva Wellassa University, Badulla, Sri Lanka.*

Modelling the semantic significance between questions and answering (QA) is essential for the detection of precise answers in Online Discussion Forums (ODF). QA can be divided as factoid and non-factoid. Traditional methods of modelling semantic relevancy lead to the sparsity of the word features due to the short texts in non-factoid QA pairs. Textual features and word co-occurrence features that commonly used in factoid answer quality predictions are irrelevant to ODF. Hence we are proposing a model to extract textual features in non-factoid QA pairs based on Deep Belief Network (DBN). DBN is modelling the semantic relationship between QA pairs by reconstructing QA pairs into a low dimensional semantic feature space. The DBN is capable of demonstrating the semantic relevance between QA pairs by modelling the semantic information hidden in the answers. Dimensions of the DBN feature space are minimized using word frequency and occurrence of function words as word features. The model is learning the semantic information from the solved question threads and then model is training to reconstruct the question using its answers. Cross entropy error function and gradient descent optimization algorithm are used to fine tune the weights of DBN. The candidate answers with the smallest distance computed by level by level calculation is considered as the best answer for the given question. Precision (P) and Mean Reciprocal Rank (MRR) methods are used to evaluate the performance of the DBN model over the Cosine Similarity, HowNet similarity and KL-divergence Model. Result shows HowNet is unable to calculate the semantic similarity between QA pairs with high precision. Compared to the cosine similarity, KL- divergence achieved more perfection. The DBN model showing a significant difference of 5.66% in P and 3.4% in MRR when applying fine tuning. The reason of growth in the DBN model is fine tuning and training the model to learn the semantic relevancy in QA pairs from the training set.

*Keywords:* Semantic significance, Non-factoid questions, Data belief networks, Online discussion forums