

**Leveraging Conceptualization for Short-Text Embedding**

**Abstract:**

Most short-text embedding models typically represent each short-text only using the literal meanings of the words, which makes these models indiscriminative for the ubiquitous polysemy. In order to enhance the semantic representation capability of the shorttexts, we (i) propose a novel short-text conceptualization algorithm to assign the associated concepts for each short-text, and then (ii) introduce the conceptualization results into learning the conceptual short-text embeddings. Hence, this semantic representation is more expressive than some widely-used text representation models such as the latent topic model. Wherein, the short-text conceptualization algorithm used here is based on a novel co-ranking framework, enabling the signals (i.e., the words and the concepts) to fully interplay to derive the solid conceptualization for the short-texts. Afterwards, we further extend the conceptual short-text embedding models by utilizing an attention-based model that selects the relevant words within the context to make more efficient prediction. The experiments on the real-world datasets demonstrate that the proposed conceptual short-text embedding model and short-text conceptualization algorithm are more effective than the state-of-the-art methods.

**Existing System:**

They suffer severely from the data sparsity and the high dimensionality, and have very little sense about the semantics of the words. Recently, in the short-text representation and classification, the deep neural network (DNN) approaches have achieved the state-of-the-art results. Despite their usefulness, recent short-text embeddings face several challenges: (i) most short-text embedding models represent each short-text only using the literal meanings of the words, which makes these models indiscriminative for the ubiquitous polysemy; (ii) for short-text, however, neither parsing nor topic modeling works well because there are simply not enough signals in the input.

**Proposed System:**

We introduce a novel co-ranking framework to address the problem of short-text conceptualization. After co-ranking the words and the corresponding concepts simultaneously during an iterative procedure, not only the most expressive concepts but also the contextual keywords could be obtained.

We integrate the concepts and the attention-based strategy into short-text embedding, and allow the resulting conceptual short-text embeddings to model different meanings of a word under different concepts and different contexts.