

論文内容の要旨

博士論文題目

Semantic Operations on an Embedding Space

埋め込み空間における意味演算

氏 名 Yoichi Ishibashi

(論文内容の要旨)

This thesis addresses the formulation of semantic representation and operations in the embedding space. Word embeddings can mathematically represent ambiguous and diverse information in language. In the pre-trained embedding space, it is known that the semantic operation such as king - man + woman = queen, also called additive compositionality, can be computed. Semantic operations have various practical advantages such as generality and are worth studying for our understanding of the nature of the embedding space. However, no semantic operations other than additive compositionality have been discovered. In this thesis, we formulated two types of semantic operations to tackle this limitation. First, we propose a binary attribute transfer that inverts the binary attributes of words. An analogy of word vectors can transfer word attributes, but it requires explicit knowledge of whether the input word has the attribute or not (e.g., the gender attribute of “king” is male). However, this knowledge cannot be developed for various words and attributes in practice. We define an ideal transfer function without using the knowledge and propose reflection-based word attribute transfer. We demonstrate that this method achieves high accuracy in transferring words with the binary attribute to be transferred and high stability in not transferring other words. Next, we formulate word sets and set operations in the pre-trained embedding space. Set operations are critical because they can be general-purpose tools in NLP. We propose subspace-based formulations inspired by quantum logic. First, we

quantitatively and qualitatively demonstrate that the proposed method is valid as a semantic operation using a word set dataset. Moreover, we demonstrate that our proposed set operation improves performance on downstream tasks such as sentence similarity.

氏 名	Yoichi Ishibashi
-----	------------------

(論文審査結果の要旨)

Word embedding is a fundamental basis of the representation of natural language in recent neural network-based technologies. This thesis work addresses semantic operations in a word embedding space for general and intuitive use of the natural language nature of semantics in a computational manner. The first contribution is formulating binary word attribute transfer based on model-based reflection using parameterized mirrors. The second contribution is formulating the representation of word sets and their operations using linear subspaces inspired by quantum logic. Their effectiveness was verified empirically through sufficient experiments, analyses, and discussions.

The thesis work is fundamental but promising for future natural language technologies; it bridged the gap between symbolic semantic operations and the vector space representations of natural language and shed light on the further potential of the current framework of word embeddings from the viewpoint of semantic operations. This thesis opened new research directions to symbolic perspectives and operations of real-valued vector representations of natural language. Through a series of research, one article in a high-quality domestic journal and one peer-reviewed international workshop paper were published. As a result of the discussions in the thesis committee, this thesis sufficiently qualified as a Doctoral thesis of Engineering.