

論文内容の要旨

博士論文題目 An Exemplar-Based Approach to Word Representation for Information Extraction

(情報抽出における単語表現のための規範的アプローチ)

氏名 濱口 拓男

(論文内容の要旨)

単語表現は自然言語処理における基盤的な情報である。特に実数値ベクトルによる単語の表現は意味を扱うことに優れており、ニューラルネットワークでも用いられるなど幅広い領域での研究が近年なされている。

先行研究における連続表現では、訓練された埋め込みベクトルはタスクの実行時には固定される。この固定化のため、それらの連続表現は未知語や多義性を扱うことが難しいという問題を持つ。

本論文では、それらの問題を解決するために、規範的アプローチに基づいた連続表現の枠組みを提案する。この枠組みでは単語表現を固定することはせず、単語と関係したオブジェクト(連想対象)を用いて単語表現を動的に計算する。連想対象は単語を特徴付けるものであり、具体例として表現したい単語を含む用例の集合を挙げることができる。本論文で提案する枠組みは、注意機構に基づくニューラルネットワークを用いることで、連想対象から獲得された単語表現からタスクに応じた1つのベクトルを抽出する。

本論文はエンティティ分類と知識ベース補完の2つの情報抽出のタスクに取り組んでいる。前者のタスクでは大規模コーパス内の用例を、後者のタスクでは知識グラフにおける近傍の情報を、それぞれ連想対象として用いる。提案する枠組みはこれら両方の設定において有用な実験結果を示した。

氏 名	濱口 拓男
-----	-------

(論文審査結果の要旨)

平成30年1月25日に開催した公聴会の結果を参考に平成30年2月13日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動が続けていくための十分な素養を備えていることを示すものと認める。

濱口 拓男は、本博士論文において、自然言語処理分野において近年盛んに研究が行われている単語の埋め込み表現が、未知語や多義性の問題にうまく対応できていない点を指摘し、それを克服するための新たな方法を提案した。さらに、提案した手法の応用として、知識ベース中に欠落している情報を補完するタスクに取り組み、その有効性を示した。

本論文の貢献は、以下のようにまとめることができる。

1. 従来の単語のベクトル空間への埋め込み表現では、大規模な言語データから事前に学習され訓練され、具体的なタスクへ利用される。タスクの学習時にそれらの表現を更新することも多いが、訓練時に含まれていなかった未知語や単語の多義性を扱うのが難しい。本論文では、これらの問題に取り組み、その一解決策を提案した。
2. 提案した枠組みでは、単語表現を事前学習によって得られた埋め込み表現に定せず、テキスト中あるいは知識ベース中で単語と関係するオブジェクトを用いて単語表現を動的に計算する。本論文では、注意機構に基づくニューラルネットワークを用いることで、連想対象から獲得された単語表現からタスクに応じた、新たなベクトル表現を得る手法を提案した。
3. 提案手法の応用として、エンティティの分類と知識ベース補完の2つの情報抽出のタスクに取り組んだ。エンティティ分類のタスクでは大規模コーパス内の用例を文脈情報として用い、提案手法が、エンティティの分類だけでなく、多義性の問題にも対応できることを示した。知識ベース補完の問題では、知識グラフにおける近傍の情報をそれぞれ連想対象として用い、エンティティ間の未知の関係の予測に有効に働くことを実験により示した。

単語の埋め込み表現における未知語や単語の多義性の問題に対処するための単語埋め込み表現の計算法を提案した本研究は、独創性が高く、しかもエンティティ分類や知識ベース補完などの実用的な問題への適用可能性も実験により示しており、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士（理学）の学位論文として価値あるものと認める。