

電気電子工学科

キーワード 人工知能、機械学習、ディープラーニング、ニューラルネットワーク、ソーシャルメディア、テキストマイニング、自然言語処理、データサイエンス



教授 / 工学 (博士)

芥子 育雄

Ikuo Keshi

学歴

大阪大学 工学部 電子工学科、大阪大学大学院 工学研究科 電子工学専攻 修士課程、奈良先端科学技術大学院大学 情報科学研究科 博士後期課程

経歴

シャープ (株) e-ライフ開発室長/研究開発本部 技師長
マサチューセッツ工科大学 人工知能研究所 客員研究員
奈良先端科学技術大学院大学 非常勤講師

相談・講演・共同研究に応じられるテーマ

SNSにおける評判分析など "AI x マーケティング" (地域振興や産業振興) に関する共同研究、「AI・IoTがもたらす未来」に関する講演と技術相談

メールアドレス

keshi@fukui-ut.ac.jp

主な研究と特徴

「単語意味ベクトル辞書を用いた可読性のある意味表現学習と拡張性」

ニューラルネットワークを利用した単語や文書の分散表現学習は研究が進んでいるが、自動抽出される特徴量の解釈性に課題がある。また、分散表現を利用したタスクの精度を高めるためには、大規模データでの学習が必須であり、大部分のWebデータやIoT家電などにより収集されている小規模なデータに適用することは困難である。ニューラルネットワークにより獲得される分散表現に対して、本研究では、辞書編纂の専門家が構築した単語意味ベクトル辞書を導入する。本辞書では、単語は266種類の概念分類から成る特徴単語との関係 (関係あり, なし) により意味を表現する。専門家が特徴単語との関係を付与した単語を基本単語と呼ぶ。我々はニューラルネットワークの隠れノードに特定の意味 (特徴単語) を割り当て、単語意味ベクトル辞書を元に約2万語の基本単語に初期の重みを設定することにより、学習後も重みの大きな隠れノードの意味が維持されることを示す (図1)。ユーザテストの結果、教師なし学習によるツイートの分散表現に対して、重みの大きな上位5ノードの52.4%がツイートと関連があることが示された。また、分散表現学習を前提に単語意味ベクトル辞書を改良し、特徴単語数を264種類とした。さらに多様性のある大規模な極性分析ベンチマークを構築し、提案手法の拡張性評価を行った。最後にWikipediaコーパスを用いて、提案手法のドメイン独立性の評価を行った。これら客観的、主観的評価により、隠れノードは特定の意味を維持することを確認し、提案手法は分散表現学習の可読性を改善することを示した。

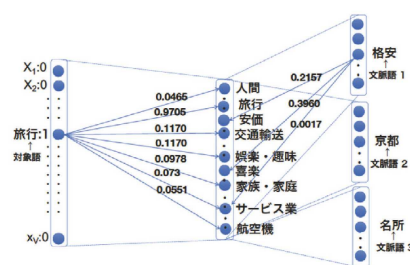


図1. 分散表現学習 (Skip-gram) モデルへの単語の初期重み (シードベクトル) の設定

「単語意味ベクトル辞書を用いた Twitter からの評判情報抽出」

LeとMikolov は、文書の分散表現を単語と同様にニューラルネットワークで学習できるパラグラフベクトルのモデルを提案し、極性分析ベンチマークを用いて最高水準の分類精度を示した。パラグラフベクトルを用いたツイートの極性分析における実用上の課題は、単語のスパース性を解消するパラグラフベクトルの構築のために大規模文書が必要なことである。本研究では、Twitterの文に対して評判情報抽出を適用する際、その出現単語のスパース性に由来する性能低下を解決するため、人手により構築された単語意味ベクトル辞書を導入する。意味ベクトルとして、各次元が266種類の特徴単語に対応し、約2万語に付与されている単語意味ベクトル辞書を使用する。この辞書を用いて単語拡張したツイートをパラグラフベクトルのモデルで学習するという、単語意味ベクトルとパラグラフベクトルの統合化手法を提案する (図2)。これにより、単語がスパースでも特定分野の文脈情報を学習できることが期待される。この評価のため、クラウドソーシングを利用してスマートフォン製品ブランドに関する極性分析ベンチマークを作成した。評価実験の結果、約1万2千ツイートから構成される特定のスマートフォン製品ブランドのベンチマークにおいて、提案手法は、ポジティブ、ニュートラル、ネガティブの3クラス分類におけるポジティブ予測とネガティブ予測のマクロ平均F値71.9を示した。提案手法は従来手法であるパラグラフベクトルによるマクロ平均F値を3.2ポイント上回った。

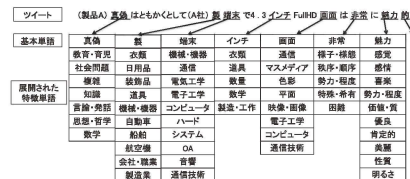


図2. 単語意味ベクトル辞書を用いて、ツイートから抽出された基本単語、及び基本単語から展開された特徴単語の例

今後の展望

以上に述べたように、ニューラルネットワークにより獲得される分散表現に専門家が構築した辞書を反映する手法を提案してきた。本辞書の基本単語は百科事典から抽出された2万語であり、特徴単語は264種類の概念分類に対応する。今後の展望は、以下に示す通り、学習結果の解釈性が重要な医療分野への展開を目指す。

第1の目標は医療分野の専門用語に意味表現学習を最適化することである。数10万件の電子カルテにおける退院サマリを元に、患者の経過要約を対象とした病名推定を本研究のフィールドとし、医学辞書をニューラルネットワークに取り込む方法論を確立する。

第2の目標は文書の特徴抽出にディープラーニングを用いることにより、病名推定の性能向上を図ることである。課題は解釈性を維持しながら、メジャーな病気の病名推定精度を改善することである。

第3の目標は病名推定の有用性検証のためのプラットフォーム化である。事例が少ないレアな病気に対しては、主訴の演算 (足し算・引き算) により対話的に病名推定が可能な手法の確立を目指す。

上述した医療分野に意味表現学習を最適化することにより、学習結果の解釈性を維持しながら病名推定の精度が高く、主訴から対話的に病名推定が可能なプラットフォーム化を今後の展望とする。

所属学会

- 一般社団法人 電子情報通信学会員 (1981年~現在まで)
- 一般社団法人 情報処理学会員 (1982年~現在まで)
- 一般社団法人 人工知能学会員 (1992年~現在まで)

主要論文・著書

Ikuo Keshi, Yu Suzuki, Koichiro Yoshino, Satoshi Nakamura "Semantically Readable Distributed Representation Learning and Its Expandability Using a Word Semantic Vector Dictionary" IEICE Transactions E101-D(4), pp.1066 - 1078, April 2018.

芥子育雄, 鈴木優, 吉野幸一郎, ニュービッググラム, 大原一人, 向井理朗, 中村哲

"単語意味ベクトル辞書を用いたTwitterからの評判情報抽出" 電子情報通信学会論文誌 J100-D(4), pp.530 - 543, 2017年4月

芥子育雄, 池内洋, 黒武者 健一

"百科事典の知識に基づく画像の連想検索" 電子情報通信学会論文誌 79 D-2(4) pp.484 - 491 1996年4月