

解釈性のある意味表現学習によるデータ分析・可視化



Keyword

分散表現、word2vec、可視化、Twitter、単語意味ベクトル辞書、カテゴリ分析

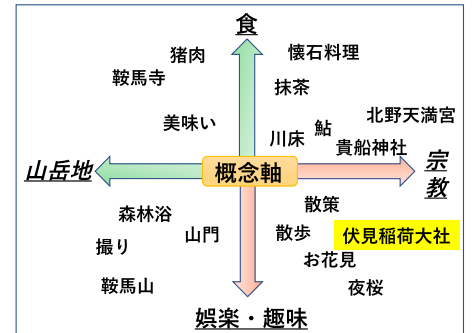
連絡先

電気電子工学科 教授 芥子育雄

電話 0776-29-2618

E-mail keshi@fukui-ut.ac.jp

ニューラルネットにより獲得される隠れ層と単語や文書間の重みに専門家が構築した辞書を反映する学習手法を提案しTwitter分析に適用しました。



Twitter 分析・可視化イメージ

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展していますが、自動抽出される特徴量の解釈性に課題があります。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す。そこで私達は、隠れノードを有限個で意味を代表する特徴単語で表現することとし、特徴単語との連想的、意味的関連性を記述した単語意味ベクトル辞書をシードベクトル（単語や文書と隠れノード間の重みの初期値）に導入することにより、特徴量の解釈性を改善しました（図 1）。また、特徴量の解釈性があるからこそ実現可能な特徴単語を概念軸とした Twitter のカテゴリ分析、可視化手法を提案しました（図 2）。

今後は、意味表現学習の性能改善および分析・可視化のプラットフォーム化を推進すると共に「Twitter に見る福井県の観光・産業」など、産学連携による福井県の地域振興や産業振興に役立つ分析を考えています。

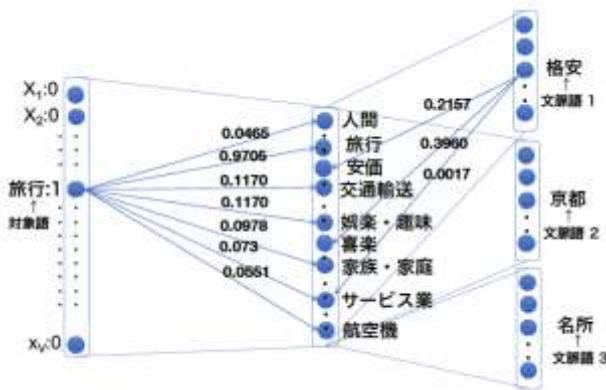


図 1 単語意味ベクトル辞書によるシードベクトル

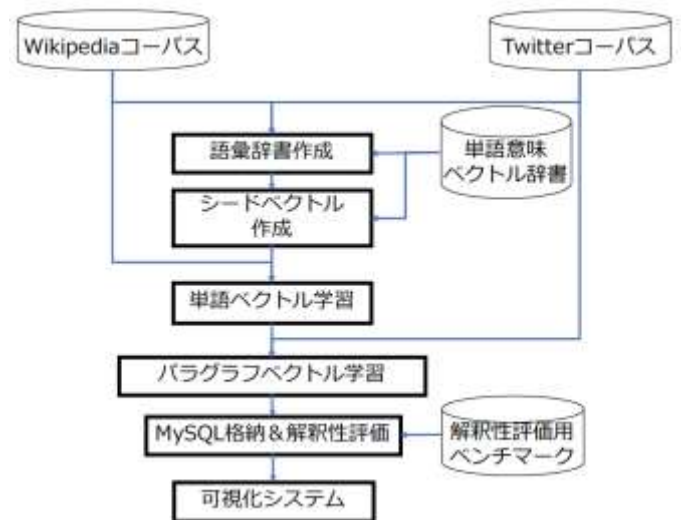


図 2 提案手法の流れ

《 共同研究の相手となる業界等 》

製造業、卸売業、小売業、サービス業、地方自治体

《 参考文献 》

Ikuo Keshi, Yu Suzuki, Koichiro Yoshino, Satoshi Nakamura, “Semantically Readable Distributed Representation Learning and Its Expandability Using a Word Semantic Vector Dictionary,” IEICE TRANSACTIONS on Information and Systems, Vol.E101-D, No.4, pp.1066–1078, Apr. 2018.

芥子 育雄, 松田 義貴, 鈴木 優, 吉野 幸一郎, 中村 哲, “意味表現学習におけるツイート分散表現の解釈性評価と可視化の提案,” 第 10 回データ工学と情報マネジメントに関するフォーラム, Mar. 2018.