13

単語の意味や関係を賢く学習

~ 開発者利便性が高い単語埋め込みベクトルの獲得法~



どんな研究

単語間の意味的な関連性の演算を可能とする「単語埋め込みベクトル」は、計算機による人間の直感に即した自然言語処理/理解の実現に必須な要素技術です。この技術を用いてシステムを構築する際に、システム開発者が不便に感じる欠点を機械学習技術を駆使して解消します。

どこが凄い

従来法と同等の性能を100分の1程度の必要メモリ量で実現する単語埋め込みベクトルです。これにより計算リソースが限定的な携帯端末のような環境でも利用可能となります。また、計算結果の再現性を担保し、利用者が安心して使える環境を提供することができます。

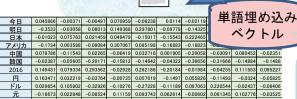
めざす未来

開発者利便性を高めることにより、コンピュータによる自動翻訳/質問応答/対話エージェントといった自然言語を介するあらゆる人工知能関連システムで使われる基盤技術の構築を目指しています。また、本技術を発展させることで、実システムの性能の底上げを実現します。

単語埋め込みベクトル

- 単語を(通常50から1000次元程度の)ベクトルで表現したもの
- 意味的な処理に基づく高度な自然言語処理/ 理解を実現する重要な要素技術

大規模な言語データ 例:新聞記事、 web全体



単語間の意味的な「関連性」

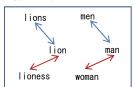
車 0.044333 -0.12678 -0.00515 自転車 -0.08889 -0.15497 2867

→ ベクトル間の「距離」「角度」で表現 また、関連性を「演算」することが可能

単語間の類似度 ⇒ <mark>距離</mark> 単語間の関連性 ⇒ 角度



例:意味的に類似した 単語は空間内で近い位 置に配置される



例:単数複数の関係(青) 性別(男女)の関係(赤)

ニューラルネット /深層学習 A L L 知能 言語理解 自然言語処理のあらゆる 応用システムで活用 質問応答/情報検索

従来法により生成された 単語埋め込みベクトルは、 利用者(研究者(開発者)によっ

利用者(研究者/開発者)にとって使いにくい 欠点があることが知られている

● 例1:計算結果が毎回違う● 例2:膨大なメモリ量が必要

[再現性の向上] 本成果:いつ、誰が、どこで実行しても同じ結果を獲得可能

毎回同じ結果が安定的に得られることで、 システム開発を安定して行うことが可能に

[省メモリ化] 単語埋め込みベクトル利用時の必要メモリ量を10~100分の1に削減

● 携帯端末のような計算リソースが少ない環境でも 単語埋め込みベクトルが利用可能に

関連文献

- [1] J. Suzuki, M. Nagata, "Right-truncatable neural word embeddings," in Proc. of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 1145-1151, 2016.
- [2] J. Suzuki, M. Nagata, "Learning compact neural word embeddings by parameter space sharing," in *Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2046-2052, 2016.

連絡先

鈴木 潤 (Jun Suzuki) 協創情報研究部 言語知能研究グループ E-mail: suzuki.jun(at)lab.ntt.co.jp