

2017/05/08

sunouchi

Word2vecとは何か

- Googleが開発 <https://code.google.com/archive/p/word2vec/>
- Mikolovらの論文 <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- 公式リファレンスが分からない
- githubで見つけたdanielfrg氏のをを使った <https://github.com/danielfrg/word2vec>

Word2vecで何ができるか

- distance: 類似語を抽出する
- doc2vec: ドキュメントの類似度を求める？
- word-analogy: A->B, C->X のような感じで、A,B,Cの3語を与えると、Xを求めてくれる
- compute-accuracy: ?

こんなことができるよ

- Doc2vecによる文書分類(47あたり) <https://www.slideshare.net/recruitcojp/ss-56150629>
- 参考記事
 - 自然言語処理をなにも知らない私がword2vecを走らせるまで <http://saiyu.cocolog-nifty.com/zug/2014/02/word2vec-1867.html>

疑問

- Word2vecのSkip-gramという仕組みをいまいよく理解できていない
 - 絵で理解するWord2vecの仕組み <http://qiita.com/Hironasan/items/11b388575a058dc8a46a>
- 学習してみると、CPU300%を超えた。計算効率を上げるのは、スペック高いマシンを使うだけで良いのか？GPUを使えば良いのか？ほかには？
- Word2vecのreferenceの本家ってどこなの？