(/) Community • Q Sign up (/signup?redirect\_to=%2FHironsan%2Fitems%2F8f7d35f0a36e0f99752c) Log in (/login?redirect\_to=%2FHironsan%2Fitems%2F8f7d35f0a36e0f99752c)

(/Hironsan)@Hironsan (/Hironsan) Edited at 2017-03-23

いますぐ使える単語埋め込みベクト ルのリスト

第1页 共3页 2017/12/10 17:36

Python(/tags/Python)

自然言語処理(/tags/%E8%87%AA%E7%84%B6%E8%A8%80%E8%AA%9E%E5%87%A6%E7%90%86)

機械学習(/tags/%E6%A9%9F%E6%A2%B0%E5%AD%A6%E7%BF%92)

DeepLearning(/tags/DeepLearning) gensim(/tags/gensim)

**118** 

## はじめに

単語埋め込みとは、単語を低次元(と言っても200次元程度はある)の実数ベクトルで表現する技術のことです。近い意味の単語を近いベクトルに対応させることができるとともに、ベクトルの足し引きで意味のある結果(例えば king - man + women = queen)が得られるのが特徴です。

単語埋め込みベクトルは品詞タグ付け、情報検索、質問応答などの様々な自然言語処理アプリケーションに使われる重要な技術です。とはいえ、実際に用意するのはなかなか骨の折れる作業です。大規模なデータをダウンロードし、前処理をかけ、長い時間をかけて学習させ、結果を見てパラメータチューニングを行う・・・。

そのため、**単に使いたいだけなら学習済みのベクトルを使うのが楽**です。というわけで、そんな単語埋め込みベクトルの中から今すぐ使えるものをピックアップしてみました。

埋め込みベクトルの情報は以下のリポジトリにまとめているので、こちらもよろしく。 awesome-embedding-models (https://github.com/Hironsan/awesome-embedding-models)

## まずは定番な3つ: Word2Vec, GloVe, fastText

## Word2Vec

一言コ メント	言わずと知れたWord2Vecの事前学習済みベクトル。何を使えばいいのかわからないならこれを使っておけば間違いはない。
発表年 数	2013年
URL	https://code.google.com/archive/p/word2vec/ (https://code.google.com/archive/p/word2vec/)

日本語を含めた多言語の学習済みベクトルは以下のリンク先からから入手することができる:

• https://github.com/Kyubyong/wordvectors (https://github.com/Kyubyong/wordvectors)

## **GloVe**

一言コメント	Stanfordが誇るGloVe。Word2Vecより良い性能だと謳っている。グローバルな行列分解のモデルとlocal context windowのモデルを組み合わせてよい単語ベクトルを学習した。
発表年 数	2014年
LIDI	har. 11.1. a. 1.1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.

