単語の埋め込み表現を用いた Evernoteの 自動タグ付けの検討

TFUG Utsunomiya #13 LT

自己紹介

- 横山雅季 (@myblackcat7112)
- M2
- DNN音声合成
- Python
- 面白そうなことは、どんどんやっていくスタイル



ぼくのEvernoteの使い方

- Webクリッパーとして
 - 気になった記事をあとで…
 - とりあえず、役に立ちそうなページを…
- Twitterの発言ログの保存先として
 - ・あとで見返すと面白い
 - そのとき何を考えていたのか
 - アイデアの宝庫?
- リマインダーとかメモとして…



ぼくのEvernote(現状)



Twitterのログ

仮想通貨 Twitterの漫画

お金持ちの特徴

効率の良い勉強方法 ねこのかわいい画像 一度は行ってみたいスポット

片付けたい! (朝ドラのタイトルみたい...)



Evernoteのノートブックがグチャグチャになってきたので抜本的に大整...

tipstour.net/ever-note/9948 ▼

2015/02/24 - しまった…たたみきれない! …はい。 はいこんにちわ。 来年ワーホリでニュージーランドに行くために8月末で仕事を辞める予定のやまぶきです。 **Evernote**をかれこれ5年近く使い続けているわけなのですが、最近どうにもあまり有効活用できて …

スタックとノートブックの整理 ...・分析 2、Evernoteの特性を ...・方針 2、ToDo管理も ...

evernoteタグ整理術。逆転の発想でタグの機能を150%引き出す使い方...

gtdtips2.blogspot.com/2017/10/evernote.html ▼

2018/03/30 - タグを使いこなすと、様々なevernoteのノートを、きちんと整理し、必要な情報を取り出せるようになり、「すべてのことを記憶 ... 本記事は、仕事術gtdの整理術や、仕事管理の考え方をevernoteに応用するものだ。gtdのガイドブック記事で、gtdの ...

【2017年版】Evernoteの整理方法:タグを使うだけで全て解決! - 鈴森...

https://suzumoritakumi.com/how-to-organize-evernote-tags/ ▼

@beck1240 さんのマネをして作成している、いわゆる「自分の...

Evernoteの整理方法や、使い方に悪戦苦闘していませんか?私も最近までその一人でした。ですが、タグの本当の使い方を知った今、Evernoteは毎日使う最強のツールになりました。 私が使っているタグの使い方を全て紹介しています!

タグを使えば良いのは、わかってるんだよぉ!

記事のタイトルだけ見てタグ付ける?

手で数千記事のタグ付けは勘弁して…

問題は、そのタグをどうやって付ければ良いか?

Evernoteさん、自動でタグ付けてくれよ…



記事のタイトルを見て、ふさわしそうな カテゴリ名をタグとして付ければいいんだ!

たとえば…

実行してはいけないLinuxコマンド(1) Ubuntu Serverで『rm -rf /』を実行|マイナビニュース

PCとかLinuxとか、サーバとかのタグがつけば良い

クロールのバタ足、速くなる効果なしむしろ水の抵抗増

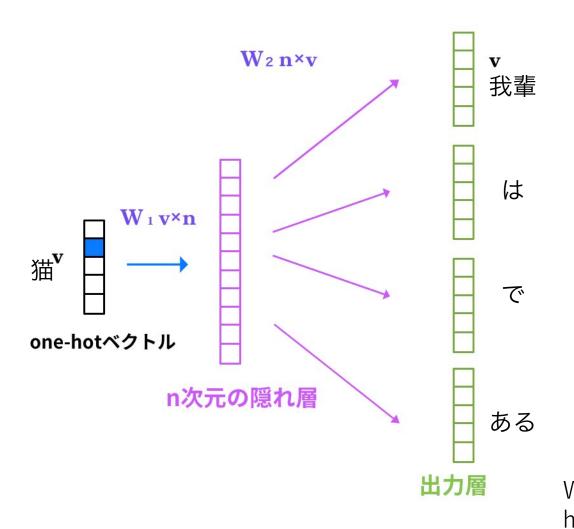
スポーツとか水泳とか、クロールとかのタグがつけば良い



自動タグ付けの流れ

- 1. 各ノートのタイトルを単語の埋め込み表現(ベクトル)を用いて表す
- 2. 付けたいカテゴリ名の単語の埋め込み表現(ベクトル)を得る
- 3. タイトルとカテゴリ名の持つ意味が近ければベクトルは似て くるはず。2つのベクトルの類似度を計算
- 4. 全てのカテゴリ名とタイトルのベクトルの類似度を求め、最も類似度が高いものをタグとする

単語の埋め込み表現



- 単語をベクトル化する方法
 - ・意味の計算が可能
 - 対象とする単語の近くにある単語は 意味的に近いはず
- n次元の隠れ層からベクトル表現 を得られる。
- ・ベクトル間の類似度を求めれば、 どれだけ似た単語なのかが分かる

Word2Vec:発明した本人も驚く単語ベクトルの驚異的な力 https://deepage.net/bigdata/machine_learning/ 2016/09/02/word2vec_power_of_word_vector.htmlより引用

実験

- カテゴリ: 26
- タグを付けるEvernoteの記事数: 1000
- 単語の埋め込み(Word2Vec)の学習に用いる記事数: 10000
- Word2Vecの次元数: 100
- 学習回数: 29(early stopping)
- 単語の種類数:9906
- タグを付けたいEvernote記事のタイトルを単語に分解して、 出てきた単語のベクトルの総和を求めタイトルを表すベクトル とする

結果(その1)

TEDを英語学習に活用する~TED勉強法の王道6Step~ | リスニング | TOEIC Lab: TOEICを短期間で高得点を目指す人のためのサイト

→'ネタ': 0.2055904, 'グルメ': 0.19872004, 'レシピ': 0.17538078

- ディープラーニングを用いて「写真の見た目の特徴」を別の写真に転送してしまう「Deep Photo Style Transfer」 GIGAZINE
 - 'ゲーム': 0.22755796, 'ネタ': 0.18459052, '科学': 0.14852937

一応あってる?

学習に用いた全ての単語をタグ候補に

• ArduinoをRaspberry Piで使いこなす——Raspberry Piに搭載できるArduino互換ボード「π duino(パイデュイーノ)」 「fabcross

'Raspberry': 0.8765589 'Pi': 0.80370343 'Arduino': 0.65621388 'USB': 0.61371064

・もう科学論文のねつ造は許さない。画像データの不正を検出するAI

'科学': 0.62866557 '論文': 0.62556201 '読み方': 0.55326563,

'神経': 0.5172227

まあまあできてそう!

まとめ

- Evernoteのタグ付けに、単語の埋め込み(Word2Vec)を使って やってみた
- 全ての単語をTag候補と見なせば結構、それらしいものが得られた