2016/04/07

A New Approach to Embedding Semantic Link Network with Word2Vec Binary Code

1. この研究はどのようなものか？

word2vecを用いて SLNを構築している。

2. 先行研究と比較してすごいところは何か？

従来の方法に比べて、3%良いクラスタリングを行える。

3. 技術や手法のポイントはどこか？(目的関数に何を使ったか、Scoring metrics, major 等々)

ドキュメント→グラフ→semantic グラフ→kernel 行列→クラスタリング（SVMを用いる）

semantic graph を作るときにword2vec を用いている。

4. どのように提案手法が有効だと検証したか？（どのようなデータを用いたかも記述すること）

データ：20個のジャンルに分けられた20Newsgroupと、SARRの4796の記事を使った。

従来の方法であるTF-IDF、LDA、Shortest-Path Kernel、LDA+Kernel と比較した。比較した内容は、正確さとrecall率。

5. 気になったこと、気づいたことはあるか？

比較対象である、Recall がわからなかった。SLNについての知識がうすすぎる

6.次に読むべき論文は何か？

概要

SLNをドキュメントと、類似点を比較した新しいグラフを取り入れて再表現する。グラフは同時に起こった関係同士を表現したもの。ニューラルネットワーク言語モデルによって得られた豊富な意味情報を組み込んでいる。

結論

コーパスのドキュメントをそれぞれ再表現した。コーパスは、オブジェクトの意味情報をより含むことができる意味グラフである。グラフのNodeのラベルはWord2Vecで学習した

そして、グラフカーネルは、意味グラフのために設計された。意味グラフは、以前は、カーネル行列（2つのグラフの類似性を示すことができる）の計算するものだった。得られたカーネル行列はSVMのドキュメントを正確に分類するクラスフィルタで学習するために使われている。たくさんの経験は、私達の方法の効率と有効性を証明して設計された。実験の結果は示された、その方法は満足な分類を達成することができた。

ところで、この論文は、いくつか不足している。私たちは、明確の目的のないよりシンプルなドキュメントのグラフしか考慮していない。近い将来、より意味情報を含んだDAG(有向非巡回グラフ)を考慮したい。ドキュメントの意味情報をより表現できるエッジの意味記述をより豊かにしたい。画像処理のような他のタスクに私達の方法を適応することを挑戦したい。