

単語トピック特定性を用いた文脈単語の重み付け

中山雄貴 ホー・ツー・バオ

北陸先端科学技術大学院大学知識科学研究科

{s1450021, bao}@jaist.ac.jp

1章 4行目

(誤) 前後 N 単語内

(正) 前後 N 単語内(文脈ウィンドウ内)

1章 14行目、2.1節 16、20、22、23、26行目、2.2節 1、7、8行目、4、5、9、13、19、20行目、5章 30、31行目

(誤) T-test

(正) t 検定

3章 10行目

(誤) 我々はこの性質に着目し、単語のトピックの特定性から計算した文脈単語の有効性の指標

(正) 我々はこの性質に着目し、前述したジャンルを LDA におけるトピックとし、単語のトピックの特定性から計算した文脈単語の有効性の指標

3.2節 5行目

(誤) $P(w_{abstract}|k_i) = P(k_i) (\forall i \in \{1, 2, \dots, K\})$ (4)

(正) $P(k_i|w_{abstract}) = P(k_i) (\forall i \in \{1, 2, \dots, K\})$ (4)

5章 9行目

(誤) 主に珍しい単語の

(正) 主に Wikipedia において珍しい単語の

5章 15行目

(誤) Wikipedia のデータから抽出した 10 万文書を用いた。

(正) Wikipedia のデータから抽出した^[12]10 万文書を用いた。

5章 20行目

(誤) Spearman の順位相関係数を表わす。

(正) Spearman の順位相関係数を表わす。なお、共起情報を得るための文脈ウィンドウの大きさを 4 とした場合の結果だけを表わした。また、Freq、PMI、T-test、PMI+WTS、T-test+WTS の行はそれぞれ頻度のみ、PPMI による重み、t 検定による重み、PPMI と WTS を融合した重み、t 検定と WTS を融合した重みによって生成された単語の意味ベクトルのコサイン類似度と評価データセット WordSim-353、MEN、MTURK、RW の Spearman の順位相関係数を表わす。

5章 23行目

(誤) しかし、WordSim-353 においては、他の評価セットと比較して、あまり PPMI の重み付けがあまり改善されていないという結果となった。これは、WordSim-353 は絶対スコアを注釈付けしていったのに対し、MEN や MTURK は 2 つのペアを比較する形で注釈付けが行われたため、後者のほうがより正確な単語間類似度を示しているためだと考えられる。

(正) しかし、RW においては、他の評価セットと比較して、(※『あまり』を削除) PPMI の重み付けがあまり改善されていないという結果となった。これは、評価セット RW における単語ペアの単語は Wikipedia における特殊語であるため、出現頻度が非常に少なく、一般的な単語と比較して、PPMI によって単語のベクトルの成分に適切な重み付けを行うことができないのに加え、その PPMI によって与えられた適切でない重みを WTS によってより強調してしまう場合があるからだと考えられる(※考察変更)。

6章 16行目

(誤) 共起の必然性によってより多くのトピックへ広がるトピックモデル

(正) 共起の必然性によって単語のトピックに対する分布がより多くのトピックへ広がるトピックモデル

参考文献

[12]<http://www.psych.ualberta.ca/~westburylab/downloads/westburylab.wikicorp.download.html>