
NLP 2019 参加報告

長岡技術科学大学
自然言語処理研究室
多田太郎

内容

- ・面白かった論文
- ・自身の発表
- ・所感

内容

- ・面白かった論文
- ・自身の発表
- ・所感

面白かった論文について

聴講できた範囲で特に面白かった論文

- ・文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か?
- ・複数粒度の分割結果に基づく日本語単語分散表現
- ・大域的な類似度と部分文字列を用いた未知語分散表現の生成手法
- ・疾患間類似度計算における分散表現の活用手法

面白かった論文について

聴講できた範囲で特に面白かった論文

- ・ **文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か? (C5-2)**

三田 雅人^{1,2} 水本 智也¹ 金子 正弘^{3,1} 永田 亮^{4,1} 乾 健太郎^{2,1}

¹理化学研究所 AIP センター ²東北大学 ³首都大学東京 ⁴甲南大学

- ・ 複数粒度の分割結果に基づく日本語単語分散表現
- ・ 大域的な類似度と部分文字列を用いた未知語分散表現の生成手法
- ・ 疾患間類似度計算における分散表現の活用手法

文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か?

既存の評価手法への問題提起

- ・このタスクでは多様な誤りについて対応する必要がある
- ・CoNLL-2014 Shared Task データ・セットが多様される
→シンガポール国立大の学生が書いた文書、言語に対する習熟度高
- ・各評価コーパスの概要、各評価コーパスでの各モデルの性能などは発表論文参照

まとめ

各コーパスでのモデルの性能順位が大幅に変動する可能性がある
コーパスを横断しての評価が必要

面白かった論文について

聴講できた範囲で特に面白かった論文

- ・文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か?

- ・**複数粒度の分割結果に基づく日本語単語分散表現 (P8-5)**

真鍋 陽俊¹ 岡 照晃² 海川 祥毅¹ 高岡 一馬¹ 内田 佳孝¹ 浅原 正幸²

¹株式会社ワークスアプリケーションズ ²人間文化研究機構 国立国語研究所

- ・大域的な類似度と部分文字列を用いた未知語分散表現の生成手法

- ・疾患間類似度計算における分散表現の活用手法

複数粒度の分割結果に基づく日本語単語分散表現

既存の日本語単語分散表現では固有名詞のような長い語の登録数が少ない

→内部の短い語との関係性を無視した学習となっていた

- ・Sudachiの長・中・短単位の各分割モードで分かち書きを行う(NWJCを使用)

→この複数の分割をしたコーパスを併用し学習を実施

- ・多くの長い語をカバーしながら内部の短い語との関連性も考慮

まとめ

複数分割したコーパスでの学習により単語間類似度と文書分類のタスクで精度向上
短単位語でも長単位を考慮した方が良い類似性を確認(詳細は発表論文参照)

面白かった論文について

聴講できた範囲で特に面白かった論文

- ・文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か?
- ・複数粒度の分割結果に基づく日本語単語分散表現
- ・**大域的な類似度と部分文字列を用いた未知語分散表現の生成手法**
五十川真生 † 梶原智之 ‡ 荒瀬由紀 † (P5-4)
† 大阪大学大学院情報科学研究科 ‡ 大阪大学データリテリィフロンティア機構
- ・疾患間類似度計算における分散表現の活用手法

大域的な類似度と部分文字列を用いた未知語分散表現の生成手法

未知語の分散表現の生成が困難

- ・大域的な類似度と部分文字列を利用
 - 大域的な類似度:他の単語との関係から学習
 - 部分文字列:単語の意味を部分文字列により構成
- ・単語間類似度推定タスクで評価

まとめ

大域的な学習と部分文字列を用いることで性能が向上

面白かった論文について

聴講できた範囲で特に面白かった論文

- ・文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か?
- ・複数粒度の分割結果に基づく日本語単語分散表現
- ・大域的な類似度と部分文字列を用いた未知語分散表現の生成手法
- ・**疾患間類似度計算における分散表現の活用手法 (P4-5)**

大村 舞 † 松本 裕治 †,§ 奥村 貴史 ‡, §

† 奈良先端科学技術大学院大学 情報科学領域 ‡ 北見工業大学 工学部

§ 理化学研究所 革新知能統合研究センター

疾患間類似度計算における分散表現の活用手法

疾患間類似度計算においても、分散表現は有効か？

- ・日本語の疾患名を疾患英和辞書データで英語に翻訳
- ・英語医学文献で学習したword2vecから英訳した疾患間のコサイン類似度を計算

まとめ

state-of-the-artには及ばないが高スコアを達成(0.7ポイント差)

シンプルな手法で高性能

内容

- ・面白かった論文
- ・自身の発表
- ・所感

自身の発表について

「分散表現による読影レポートの類似文書検索」という題目で発表
医師支援のための類似医療文書検索とこのタスクでの分散表現の有用性を確認したい

- ・読影レポートをdoc2vecを用い学習
- ・複雑な文書から、まず問題のシンプルなデータ・セットを作成し文書間類似度の算出
- ・データ・セットからどれだけ同じカテゴリの文書を得られるかで評価

発表にて

- ・たくさんのアドバイスをいただきました
- ・最初のステップとして問題をシンプルにしたが(データ・セット作成の困難さもあった) 題目と少し離れ誤解となる部分があったので、次回はもう少し考えたい

内容

- ・面白かった論文
- ・自身の発表
- ・**所感**

所感

初めて言語処理学会の年次大会に参加しました

- ・聴講した発表では固有名詞などの長い単語をどう扱うかというものが多い印象だった
- ・みんな優しく、建設的な質問やアドバイスばかりだった
- ・参加者が多く、次回のポスター発表聴講ではより準備が必要
- ・お祭りみたいでした

また来年も発表者として参加します