

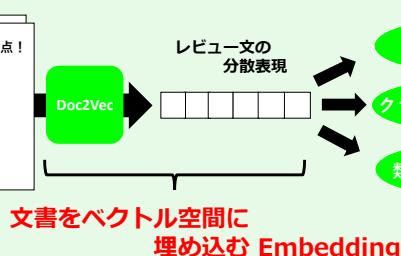
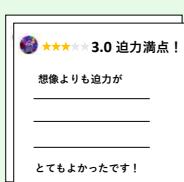
7E-01

レビューデータからの各次元が意味を持つ Disentangledな映画ベクトル表現の獲得

金田 悠路 (静岡大学), 藤田 澄男 (LINEヤフー株式会社), 莊司 廉行 (静岡大学)

機械学習でEmbeddingは当たり前

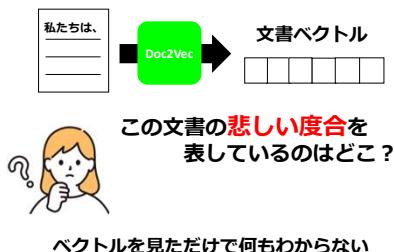
単語、テキスト、グラフ、etc.



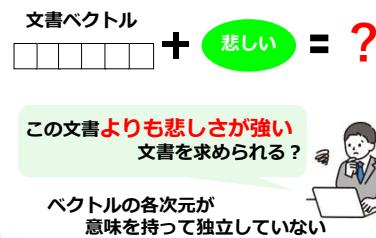
自然言語、画像、音声などの広い分野で
既存のEmbedding手法でのベクトル化が一般的

分散表現の各次元は意味を持っていない！！

人間が解釈できない



より意味的に踏み込んだベクトル演算ができる



ベクトルの各次元が重複のない意味を持つようにするには？

分散表現の…

ベクトルの各次元に意味を持たせることで

ちょい足し検索

映画を気になる観点で比較可能に！

類似度比較

「Disentangled」なエンコーダを作成！

- ・現実空間の距離関係が反映
- ・ベクトルの各次元が独立
- ・ベクトルの各次元が意味を持つ

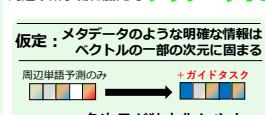
映画id ID:388069

レビュー

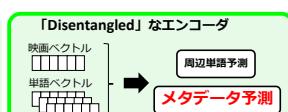


工夫① ガイドタスク学習

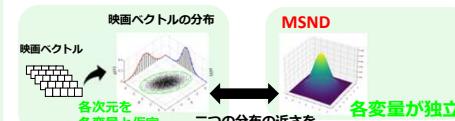
Doc2Vecの周辺単語予測に加えて メタデータ予測 も同時に予測



文書の内容 + 文書間の意味的関係性を学習！



工夫② β -factorをロスに適用

 β -VAEを参考に… ロス計算で MSNDとの距離を考慮
多変量標準正規分布

ベクトルの各次元の独立化を促す！

これを実現するために 2つの工夫 を Doc2Vec の学習過程に追加

トイ・データを使用した評価 工夫②の効果検証

比較手法：ランダム、Doc2Vec、工夫②のみ

知見① ベクトルの解釈可能性が向上

ベクトルの各次元の値の分散に注目
各次元が独立したベクトルは各次元の解釈可能性が高い

ベクトルの各次元の値の情報エントロピーに注目
各次元が独立したベクトルは各次元の情報量が制限

ベクトルの各次元の値 上位文書のトピックに注目
各次元がそれぞれ特定のトピックについての情報を持つように

実データを使用した評価

比較手法：提案手法、工夫①のみ、工夫②のみ、Doc2Vec

知見② ベクトルが現実空間の距離関係を反映

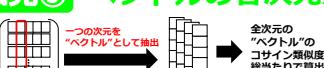


「Disentangled」なベクトルか検証

クエリ映画に似た映画ランキングを主観評価

どの手法でも類似映画が上位に！

知見③ ベクトルの各次元が独立



映画ベクトルの各次元が直交する組み合わせの数を評価

提案手法のほとんどの次元同士が独立

知見④ ベクトルの各次元が人間の解釈できない意味を持つ



映画ベクトルの各次元上位映画に共通点があるかを被験者・自動評価
各次元の意味的なまとまりが被験者の回答と一致せず

{ ガイドタスク MNSDとの近さ考慮 } の 2つの工夫 が 各次元の独立した分散表現の獲得 を可能に