

Comment Mover's Distanceを用いた 類似動画検索



Similar Video Search using Comment Mover's Distance

福田 寛公[†] 松本 和幸[‡] 吉田 稔[‡] 北 研二[‡]
[†]徳島大学大学院先端技術科学教育部
[‡]徳島大学大学院社会産業理工学研究部



Faculty of Engineering
The University of Tokushima

研究背景

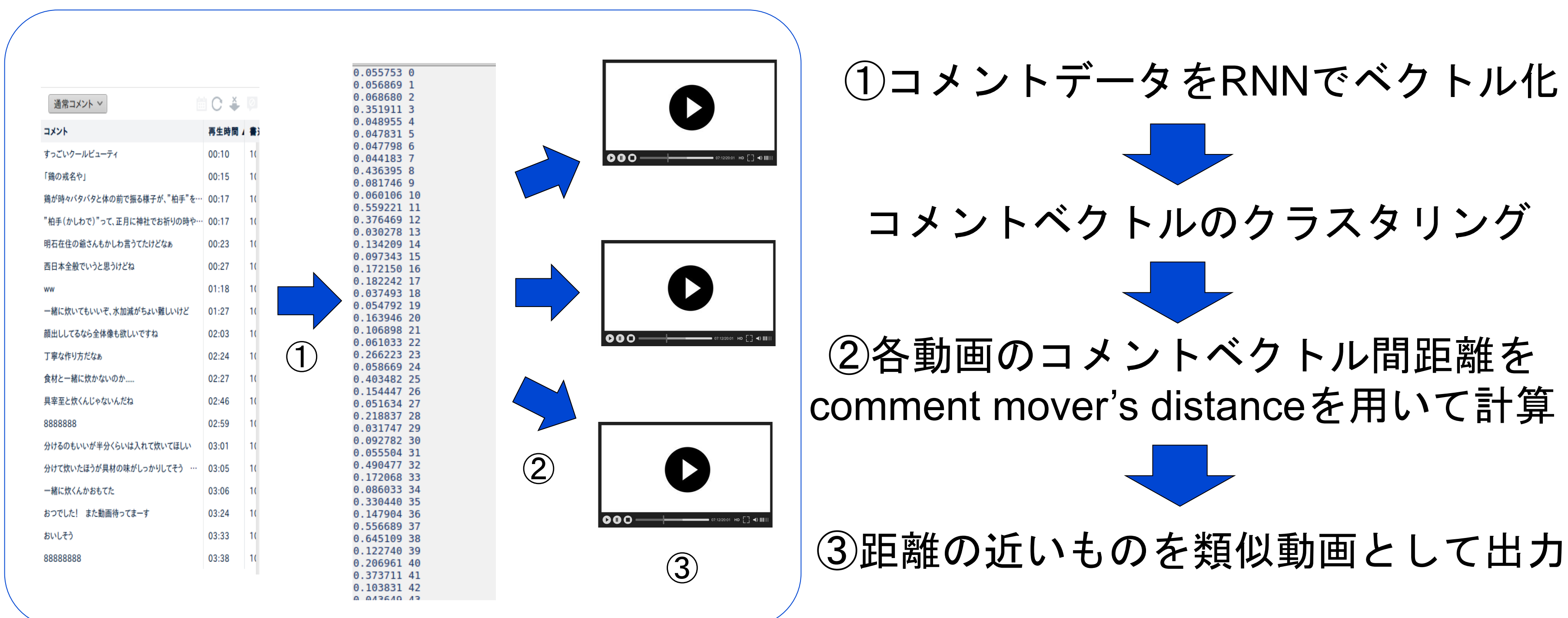
- 動画共有サイトの発達
 - YouTube, ニコニコ動画 etc...
- 投稿される動画が急増
- 興味ある動画と似ているものを探すことが困難に

自動で類似動画を検索できるシステム

- 動画検索の方法
 - 動画の情報を用いる
 - 内容の類似を検索できない
 - タグを用いる
 - 恣意的な編集が可能

コメントに着目した動画検索システム

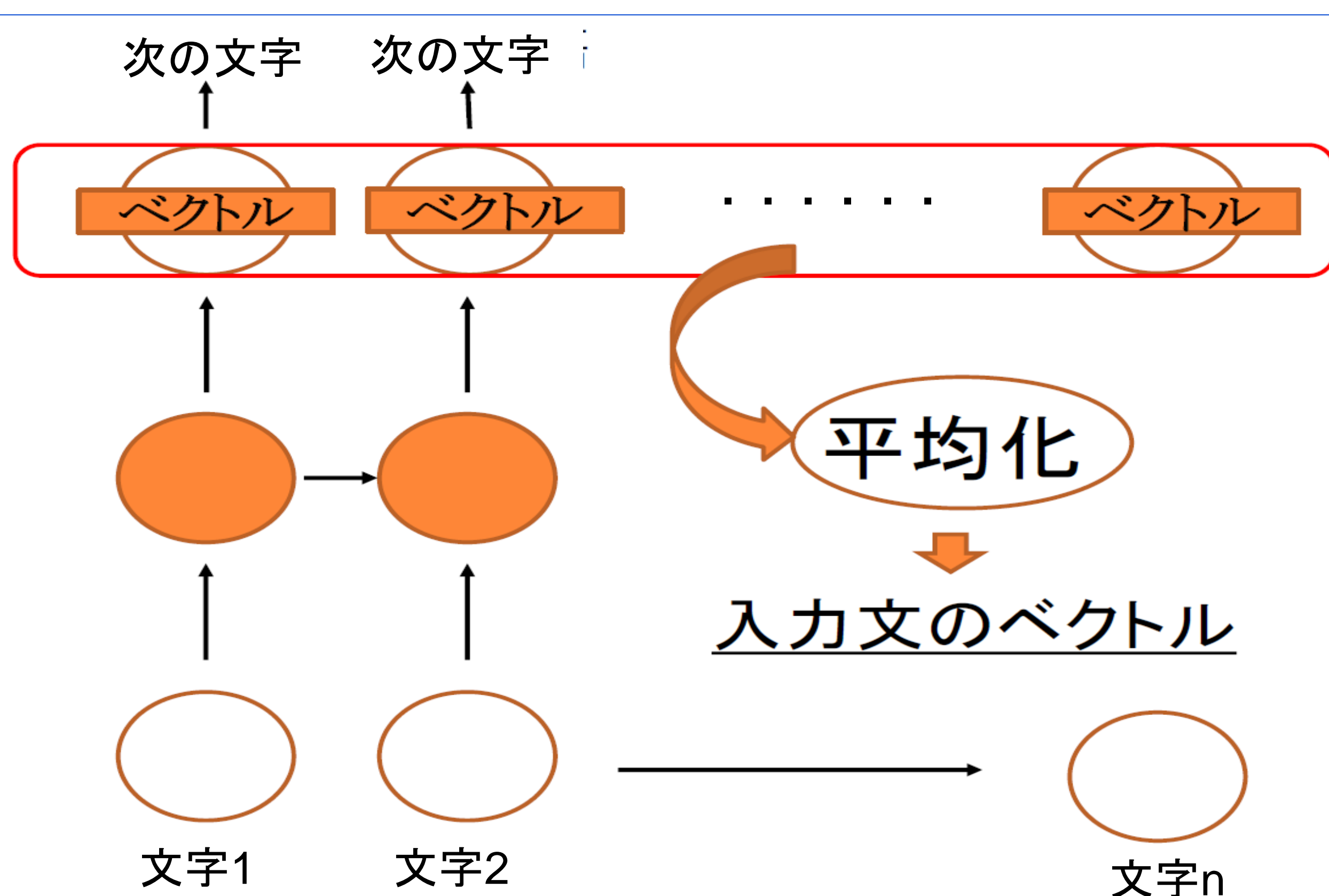
研究手法



RNN

- 時系列情報を用いることが可能
 - コメントは時系列情報が重要
 - コメントを文字単位で扱える
 - コメントは形態素解析できないものが大半
 - 違う言葉でも同じ意味、
同じ言葉でも違う意味を持つコメントの出現
- 例) ○○は神ゲー ←それな○○は神
 例)心がぴよんぴよんするんじゃ
 ↓
 心/が/ぴよんぴよん/する/ん/じゃ
 例)このダメージはヤバい ⇔
 魔法石100個配布はヤバい

Bag-of-wordsなどより適している

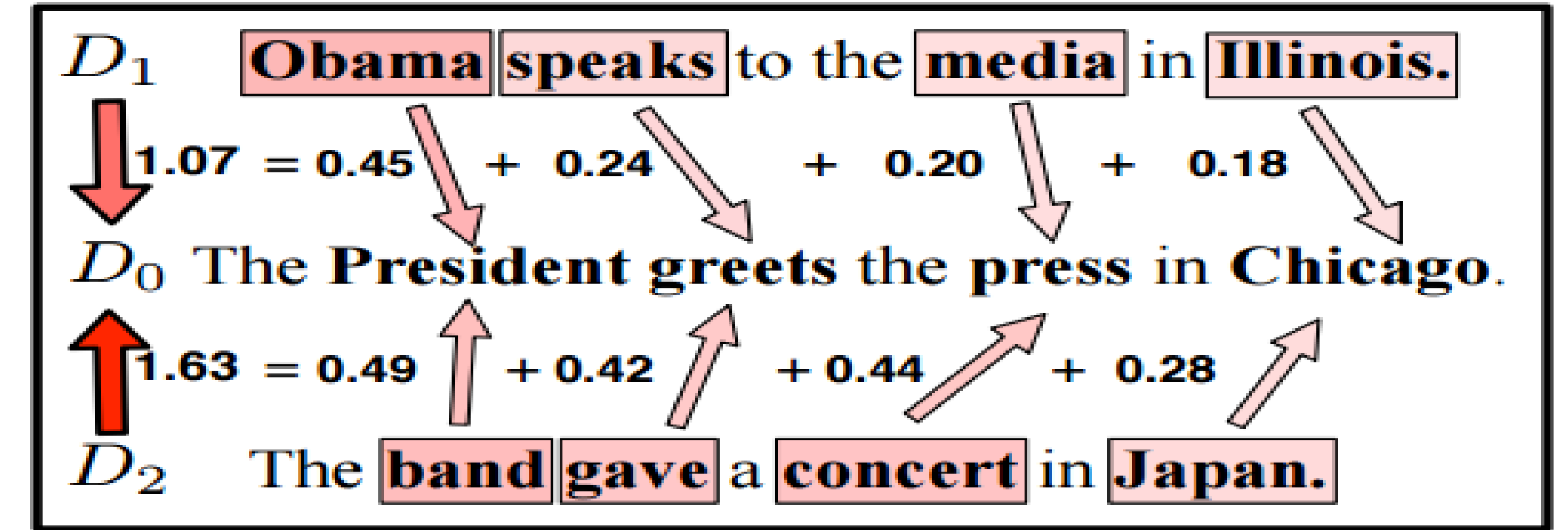


Comment Mover's Distance

- word mover's distanceをコメントに拡張

word mover's distance : 文書間距離を計算するアルゴリズム

コストの総和を文書間距離と考える



WMDのアルゴリズム

- fastEMDを用いて計算を高速化
 - ↳WMDは計算が遅い
- コメントベクトル間距離を計算
 - ↓
 - 距離に近い順に出力することで検索

実験結果

ゲーム,アニメ,音楽,政治,科学,スポーツの6ジャンルにおいて、
ジャンルが一致するかの実験を行った
(各ジャンル100件、計600動画に対し実験)

カテゴリー	再現率	適合率	F値
ゲーム	0.540	0.675	0.600
アニメ	0.427	0.494	0.458
音楽	0.546	0.338	0.417
政治	0.520	0.600	0.557
科学	0.500	0.450	0.473
スポーツ	0.370	0.536	0.438

実験結果

ベクトルのクラスタ数 : 135
 1つの検索単位のコメント上限 : 3
 各コメント集合ごとの類似候補数の上限 : 20

今後の課題

- ジャンル一致は40~60%の精度を出せた
- 投稿された年代の近い動画が検索結果として多く出現していた
 - その当時流行していた言葉が多くコメントに投稿されていた
- 実験結果を基にしたアンケートの実施
 - アンケートにより内容の類似を確認
- クラスタリングのクラスタ数を変更し、精度の変化を確認
 - 現在は135
- 各パラメータを変更し精度向上

参考文献

- 青木秀憲. ニコニコ動画における映像要約とサビ検出の試み. 情報処理学会研究報告
- 多胡厚津史. ソーシャル Annotations とキーワード群に基づく動画要約. 情報処理学会シンポジウム論文集
- 高間康史. 動画要約のための並列映像提示手法の提案.
- 高野祐太郎. 投稿型動画視聴におけるユーザ間リアルタイムコミュニケーション支援システム. システム開発論文特集
- 住元宗一郎. コンテンツ投稿型SNSにおける未知性と意外性を考慮した推薦エージェントの提案.
- Nan Yang. Data Analysis System for Online Short Video Comments
- 北研二. Earth Mover's Distanceの高速検索ライブラリfastEMDの開発
- 田中克己. 視聴者のコメントに基づく動画検索及び推薦システムの提案
- ニコニコ動画 : <http://www.nicovideo.jp/>
- ニコニコ大百科 : <http://dic.nicovideo.jp/?header>