

佐々木 稔 (Minoru Sasaki)

所属 (Domain) 情報科学領域 (Domain of Computer and Information Sciences)

・ 博士後期課程社会インフラシステム科学専攻 (Major in Society's Infrastructure Systems Science)

● 研究テーマ (Research theme)

① 文の中に存在する単語の語義を国語辞典の中から自動分類する研究

(Word Sense Disambiguation Using Machine Learning Techniques)

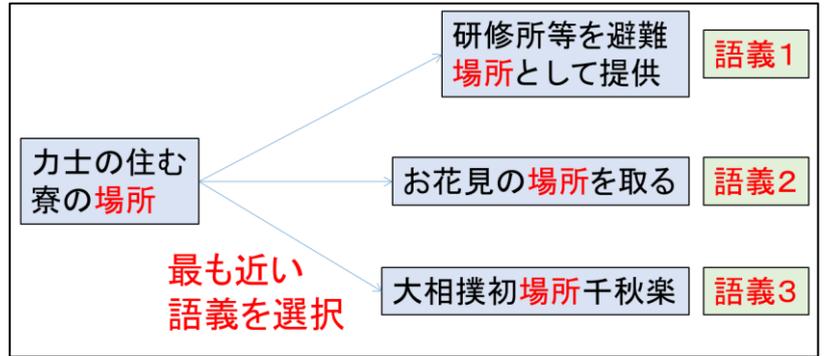
② 単語の分散表現を用いた書籍・新聞・雑誌などの領域における出現単語の特徴分析

(Semantic Relation Analysis among Domains using Multi-Domain Word Embeddings Learning)

① 対象単語を含む文を与えたとき、その単語が辞書中のどの意味に属するか識別します。識別をするためには、語義と付加情報が付いた用例文をあらかじめ用意し、語義の自動分類を行うための識別モデルを作ります。識別モデルができれば、新しい文を与える事で、どの意味に属するか自動的に判定できます。

Word Sense Disambiguation (WSD) is the problem of assigning the appropriate meaning to a given

word in a text or discourse. In this problem, supervised learning systems are typically employed, as they can be trained using a number of hand-labeled examples and then applied to classify a set of unlabeled examples.



Target words and domain	Similar Words	Target words and domain	Similar Words
市場_書籍	市場_新聞, 企業, システム, 競争, 取り引き, 化, 市場_雑誌, 需要, 構造, 銀行	見せる_書籍	見せる_新聞, 見る, 見せる_雑誌, 教える, 感ずる, 与える, 伝える, せる, 広げる, 引き立てる
市場_新聞	市場_書籍, 市場_雑誌, 利下げ, 企業, 大手, 貿易, 下落, じり高, 軟調, 買戻	見せる_新聞	見せる_雑誌, 見せる_新聞, 見る, 面変わり, 語る, 決める, 見せ付ける, 知る, 熟す, 広げる
市場_雑誌	市場_新聞, 市場_書籍, 経済, デフレ, 企業, 大手, 株安, 資本, グローバル, 集権	見せる_雑誌	見せる_新聞, 見せる_書籍, 見る, 感ずる, 教える, 付ける, 与える, 出す, 覚える, 変える
口紅_書籍	リボン, ピンク, チーク, パウダー, ウール, 同系, 裏地, ラメ, ブラウス, コート	わくわく_書籍	うんざり, ときどき_雑誌, 感激, ときどき_新聞, 狼狽, ぐずぐず, 浮き浮き, ときめく, 嬉々, 奇立つ
口紅_新聞	ドリトス, 衣料, 子馬, 特売, 押麦, 茶粥, プラスチック, 後払い, ステンレス, コーンビーフ	わくわく_新聞	イレン, 操る, 町歩き, しゃざり, R U, フェスタ, 触れ合い, むずかる, ときどき_新聞, 宝恵
口紅_雑誌	前著, パーブル, 茶系, パール, マフラー, イミテーション, メロー, ラメ, 風姿, 重修	わくわく_雑誌	ときどき_雑誌, ほろり, 台無し, 閉口, 居た堪れる, 息継ぎ, おたおた, ときまぎ, はらはら, まごまご

② 領域ごとの単語ベクトル生成手法を考案し、各領域における単語の特徴分析を行います。書籍・雑誌・新聞の3つの分野(領域)の日本語のコーパスに対し、指定した対象単語に対して単語ベクトルを作成する。対象単語のベクトルを用いて、対象単語が類似する単語を各領域において抽出し、単語の使用傾向などの分析を行います。

In this research, I propose a word embeddings learning method for not less than three different domains and analyze characteristics of word embeddings for three different domains. I use three

domains, Published Books, Published Magazine and Published Newspaper to construct word embeddings using word2vec for each domain. For the obtained word embeddings in each domain, I extract top-k similar words to the target word and analyze similarities and differences among the three domains.

キーワード (Keyword)

専門分野 (Specialized Field)

共同研究可能技術 (Possible Technology of Cooperative research)

関連論文・特許情報 website

(Related articles・patent information)

研究設備 (Research Facility)

研究室URL (Lab. URL)

E-mail

語義曖昧性解消 (Word Sense Disambiguation) 機械学習 (Machine Learning) データマイニング (Data Mining)

自然言語処理 (Natural Language Processing)

機械学習や人工知能によるテキスト解析技術

(Text Analysis Using Machine Learning and Artificial Intelligence)

<https://info.ibaraki.ac.jp/Profiles/5/0000495/profile.html>

実験用コンピュータ

<http://sas.cis.ibaraki.ac.jp>

minoru.sasaki.01@vc.ibaraki.ac.jp