

特 集

◇自然言語処理・要約

Automatic Summarization on Various Domains with Combinatorial Optimization and Machine Learning

西川 仁

nishikawa.hitoshi@lab.ntt.co.jp

奈良先端科学技術大学院大学

指導教員：松本 裕治

博士 (工学), 2013 年 9 月 25 日 取得



キーワード：自然言語処理, 自動要約, 組合せ最適化, 整数計画問題, 構造学習.

概要：インターネットの普及に伴い電子化されたテキストの種類、量が増加するに従って、自動要約において新しい課題が生じている。課題は大きく分けて三つ存在する。第一に可読性の問題であり、第二に内容性の問題であり、最後に要約速度の問題である。本研究の目的は、自動要約において新しく生じたこれらの課題に対して、新しい自動要約の方法を提案し、これを解決することである。

本研究の主たる貢献は、これらの問題それぞれに対して解決の方法を示した点にある。第一の可読性の問題に対しては、文の抽出と順序付けを同時に行う要約モデルを新しく提案し、これを整数計画問題として表現し解くことによって、文の抽出と順序付けを別個に行う既存の手法に比べて可読性が改善された要約を生成できることを示す。第二の内容性の問題に対しては、転移学習を利用し異なる分野のテキストを教師事例として用いることで内容性が改善された要約を生成できることを示す。第三の要約速度の問題に対しては、新しい要約モデルである冗長性制約付きナップサックモデルと、ラグランジュヒューリスティックに基づく高速なデコーディングアルゴリズムを提案し、良好な要約を高速に生成できることを示す。最後に、これらの結果を踏まえ、さらなる研究の方向性を示す。

主な公表論文：西川 仁, 長谷川 隆明, 松尾 義博, 菊井 玄一郎: 文の選択と順序付けを同時に行う評価文書要約モデル, 人工知能学会論文誌, Vol. 28, No. 1, pp. 88-99 (2013)

現職：日本電信電話株式会社 NTT メディアインテリジェンス研究所

論文の入手先：<https://sites.google.com/site/nishikawahitoshi/>

抱負：我々の社会を取り巻く情報の多くはテキストで表現されています。情報の量が増加するにつれて、重要な情報を素早く理解できるように人間を助ける技術である自動要約の重要性は増えています。より良い社会の実現のため、今後も自動要約の研究に取り組んでいきます。

◇自然言語処理

A Framework for Applying Language Testing Methods to Support Systems of Second Language Use

江原 遥

ehara@nii.ac.jp

東京大学大学院情報理工学系研究科

指導教員：中川 裕志

博士 (情報理工学), 2013 年 9 月 27 日 取得



キーワード：ユーザ適応, 機械学習, 言語教育, 項目反応理論, 支援システム.

概要：本論文は、言語テスト手法を、読み書き支援など第二言語使用の支援システムに適用するためのフレームワークを提案する論文である。このような支援システムは、従来、自然言語処理、知的ユーザインタフェース、教育学などの分野で提案されてきた。一方、TOEIC や TOEFL などに使われているとされる言語テスト手法は、数理的に体系化されているが、主たる目的がテストの作成であるためか、支援システムとは独立に発展してきた経緯がある。言語テスト手法を用いると、数理モデルのパラメータをユーザの能力値や、語などの項目の難しさとして解釈することが可能となるので、支援システムでも言語テスト手法を積極的に用いたい動機がある。

そこで、本研究では、1) 従来の支援システムの研究を調査・分類し、言語テスト手法が利用できる支援システムの条件を特定することにより、言語テスト手法を適用するためのフレームワークを提案している。

さらに、以下も本論文の貢献である。

2) 言語テスト手法の側の数理モデル (項目反応理論) を改良し、支援システムに適用しやすい数理モデルを提案した。このモデルでは、語の難易度を個人化し、「学習者にとっての語の難易度」を導入することで、分析の幅を広げつつ効率的な計算を可能にしている。

3) 提案するフレームワークに基づく実際の支援システムの一つとして、文書中で学習者が知らない語を予測し、自動的に訳をつけるシステムを提案した。

4) 一人約 12 000 語に及ぶ語彙知識のデータセットを 16 人分作成・公開し、語彙知識の予測性能の評価を可能とした。

主な公表論文：Ehara, Y., Sato, I., Oiwa, H. and Nakagawa, H.: Mining words in the minds of second language learners: learner-specific word difficulty, *Proc. 24th Int. Conf. on Computational Linguistics (COLING-2012)*, pp. 799-814 (Long paper), Mumbai, India (Dec. 2012)

現職：日本学術振興会特別研究員 (PD) (受入：国立情報学研究所官尾研究室)

論文入手先：<http://yoehara.com/>

抱負：私は、人に内在する言語モデルの個人差や分布に主たる興味があり、その個人差が最も大きく出る言語教育分野のユーザ適応から研究を始めた。自然言語処理には、ユーザ適応や個人化が導入される余地が多く残されているので、これからは言語教育以外の自然言語処理にもこれらを広げていきたい。