

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 25 日現在

機関番号：35302

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25350365

研究課題名(和文) 学習者の言語的背景と講義に連動した日本語学習システム開発

研究課題名(英文) Japanese learning system development in conjunction with the linguistic background and lecture of learners

研究代表者

椎名 広光 (Shiina, Hiromitsu)

岡山理科大学・総合情報学部・准教授

研究者番号：40299178

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：これまでの遠隔講義用のVODシステムは、使いやすいことを目的にシステムが開発されている。そのため講師の日本語の発話を聞いているだけで、講義内容の理解が難しい。本研究では、日本語を母語としない学生に、日本語で実施されている講義資料中の語彙の難易度推定、専門用語の検索システム、及び専門用語の学習システムの開発を行った。専門用語の検索システムでは、日本語の学習者という観点から日本語の学習状況に応じた日本語での検索結果を表示するために、機械学習を利用して、やさしい表現と判断される日本語表現の推定手法を提案した。また、専門用語の学習システムでは、学習曲線を推定し、問題のレベルを変更する仕組みを開発した。

研究成果の概要(英文)：Distance learning VOD systems were developed with ease-of-use in mind. For that reason, it was difficult for students to comprehend lecture content just by listening to the lecturer speak in Japanese. In this research, we estimated the degree of difficulty of the vocabulary used in Japanese language lecture materials that have been carried out in Japanese, for non-native speakers. Additionally, we developed a search system and a learning system for technical terms. In order for the search system to return search results that apply to the study of Japanese from a student's point of view, we utilized machine learning, and proposed estimation methods that return Japanese expressions that have been deemed simple. Furthermore, for the learning system, we developed methods that estimates a learning curve and changes the difficulty level of problems.

研究分野：教育工学

キーワード：e-ラーニング 日本語学習 専門用語学習支援 多言語検索 機械学習 話題推定

1. 研究開始当初の背景

これまで遠隔講義で利用されている VOD 講義は、その利用システムを主眼にシステムが提供されている。これらを利用する場合、講義資料したがって講師の日本語の発話を聞いているだけでは理解が難しいと考えられる。そこで、講義理解のための補助教材提供やシステムが必要となってきた。しかし、単純に理解の補助といっても受講者の理解度との関係が大きい。そこで日本語での講義と受講者の学習状況を連動させたシステムが必要となってきた。また、言語処理技術の発展により、日本語のほか多くの言語を解析するソフトウェアが利用できるようになり、加えて Web など言語情報を電子情報の形で取得しやすくなってきており、言語情報を利用することが容易になってきている。よって、多言語の解析結果を学習補助に利用できる状況にある。

2. 研究の目的

本研究では、日本語で行われている講義を受講する日本語を母語としない学生向けに、講義の理解を補助するシステムの開発を目的としている。これを実現するために、語彙の難しさとその学習、検索への補助、留学生が検索する際に問題となる漢字の読み方の学習、日本語を利用したアルゴリズムの学習、VOD 講義での話題にしている場所の補助を行うシステムを開発することで、学習支援につなげることを目的とした。

3. 研究の方法

研究目的に対して、6 項目に分けて研究を行った。項目ごとに研究の手法を示す。

- (1) 日本語で実施されている講義資料中の語彙の難易度推定
日本語の単語難易度の推定では、これまで辞書の意味記述を利用していたが、Web 検索で記述量が平均的な記事を取得しその記事中に現れる周辺単語分布から単語難易度を推定する手法を開発した。特に、講義資料で利用される専門後に多い複合語の難易度の評価に利用した。
- (2) 講義で検索される専門用語の検索システム
講義内容で検索される専門用語に対する日本語説明文の読みやすさを、日本語を母語としない留学生向けの評価を行い、検索結果に対して簡潔と簡潔でない野ラベル付けを行い、それを決定木学習を行うことで評価すべき項目を洗い出した。
- (3) 日中、中日語彙学習システム
(1)での日本語の難易度の推定の他に、中国語の形態素解析ソフト NLPiR と(1)の手法組み合わせで中国語の単語の難易度を推定し、それぞれの言語での難易度が同程度であるもとと、そうでないものを考慮して、語彙学習を行うシステム

を作成した。

- (4) 専門用語の漢字学習システム
講義に現れる漢字を含む単語の読みを学習するシステムを開発し、単語の読み方の学習状況をロジステック関数をもとにした学習曲線で把握し、より学習者のレベルに合った学習を提供した。また、非漢字語圏からの留学生で評価実験を行った。
- (5) アルゴリズムの学習システム
日本語の文章の整列によるプログラミング手順テストシステムで、日本語の理解力とプログラムのアルゴリズムの理解の補助を行うシステムを開発した。また、非漢字語圏からの留学生で評価実験を行った。
- (6) 発話字幕と PPT との対応付けシステム
スライド(マイクロソフトパワーポイント)内のパラグラフと字幕の文ごとに共起語グラフを作成し、共起語グラフの類似によって、その対応関係を求める手法を提案し、システムを開発した。

4. 研究成果

6 項目ごとで得られた成果について述べる。

- (1) 日本語で実施されている講義資料中の語彙の難易度推定
辞書データのみを利用した場合の精度 53.6%であったが、Web データを利用した場合では 60.0%に精度が向上した。また、講義資料のページ中の単語難易度やページごとの発話の単語難易度を出力できるようにした。講義名「データベース」では、講義資料の平均単語難易度が、日本語能力試験の難易度換算で 2.25 級、発話が 2.65 級であった。その他の講義でも講義資料の方が難易度級判定が低く難しめのあることが裏付けられた。
- (2) 講義で検索される専門用語の検索システム
講義内容で検索される専門用語に対する日本語説明文の読みやすさを、日本語を母語としない留学生向けの評価を行い、検索結果として評価すべき項目を洗い出した。その結果を利用して、異なり語彙数が語彙レベルよりも重要な項目であることが判明し、検索結果として提示すべき辞書情報のシステムを作成した。
- (3) 日中、中日語彙学習システム
日本語と中国語の単語難易度の推定を利用した日中単語学習システムでは単語の選択問題をお互いの難易度の同程度のものから生成するシステムで、タブレットコンピュータ上に実装を行った。また、単語学習を日本語の単語の問題に中国語の単語を選択して学習する手法、中国語の単語の問題に日本語の単語を選択して学習する手法、それらを続けて学習するだけでなくお互いの単語の問

題を交互に学習する仕組みを導入し、交互に学習する手法では、より高い学習効果が得られることを確認した。

- (4) 専門用語の漢字学習システム
単語テストシステムによる単語の読みを選択する問題で学習状況を把握し講義資料に提供する漢字に付ける振り仮名率はロジスティック関数を基に作成した振り仮名率曲線によって講義回ごとに決め、単語テストシステムの結果を基に随時更新していくようにすることで、より学習者のレベルに合った学習を提供する。本システムを、非漢字語圏からの留学生で実践利用を行ったところ、個人間の学習状況の違いを把握することができ、講義の間隔や専門用語によって理解が相違することが分った。
- (5) アルゴリズムの学習システムの開発
アルゴリズムの学習システムでは、非漢字語圏からの留学生で評価実験を行ったところ、C言語のプログラムの学習項目のうち、条件分岐や繰り返しの制御及び関数の理解までに有効にであった。ポイントの項目については、効果が少なかった。ポイントの仕組みの理解が難しく効果が表れる状況までに至らなかったと考えられる。
- (6) 発話字幕とPPTとの対応付けシステム
発話字幕とPPTとの対応付けシステムでは、対応関係の一致率については、学生アンケートによると、丁度良いかすこし少ないという評価が多かった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 10件)

- (1) Nobuyuki Kobayashi, Shingo Nakamura, Hiromitsu Shiina and Fumio Kitagawa: Estimating Movie Segment Association in VOD Lectures Using Mixed Distribution and Co-occurrence Words, 査読有, International Journal of Computer & Information Science, Vol.16, No.3, pp.1-11, 2015.
- (2) Hiromitsu Shiina, Takayoshi Mihara, Nobuyuki Kobayashi: Classification by Approximation of Normal Distribution for k-neighborhood Data using Steepest Descent Method, 査読有, The Standard International Journal Transactions on Computer Science Engineering & its Applications, Vol.3 No.8, pp.105-110, 2015.
- (3) Nobuyuki Kobayashi, Kohei Sakane, Fumio Kitagawa, and Hiromitsu Shiina: Programming Lecture Support System Linked to a Lecture Towards Non-kanji Countries, 査読有, IIAI International

Conference on Advanced Applied Informatics 2015, pp.18-23, 2015. DOI 10.1109/IIAI-AAI.2015.258

- (4) Masahiro Takeda, Nobuyuki Kobayashi, Fumio Kitagawa, and Hiromitsu Shiina: Regional Information Video Searches Using Word Searches Generated by Twitter Posts, IIAI International Conference on Advanced Applied Informatics 2015, 査読有, pp.127-131, DOI 10.1109/IIAI-AAI.2015.256
- (5) Kohei Sakane, Nobuyuki Kobayashi, Hiromitsu Shiina and Fumio Kitagawa: Audio Subtitle Mapping System between Slide and Subtitles by Co-occurrence Graphs on VOD Lecture, 査読有, pp.337-342, The 22th International Conference on Computer Education, 2014.
- (6) Nobuyuki Kobayashi, Shingo Nakamura, Hiromitsu Shiina, Fumio Kitagawa: Estimating Relevant Movie Segment for Co-occurrence Word from Japanese subtitles of VOD Lecture, 査読有, IIAI International Conference on Advanced Applied Informatics 2014, pp.13-18, 2014. DOI 10.1109/IIAI-AAI.2014.15
- (7) Nobuyuki Kobayashi, Kiyonori Nakanishi, Hiromitsu Shiina, Fumio Kitagawa, Estimating Word Difficulty through Machine Learning using Web Data and Dictionary Definitions: 査読有, Conference of the Pacific Association for Computational Linguistics(Pacling2013), Sep2-8, pp.1-6, 2013.
- (8) Pei Qing Zang, Hiromitsu Shiina, Nobuyuki Kobayashi: Paraphrasing for Simple Expression of Japanese Subtitles on VOD Lecture, 査読有, Pacling2013, Sep2-10, pp.1-6, 2013.
- (9) Pei Qing Zang, Hiromitsu Shiina, Nobuyuki Kobayashi, F.Kitagawa: Chinese and Japanese Word Learning System by Estimation of Word Difficulty, 査読有, pp.320-323, The second asian conference on information systems, 2013.
- (10) Hiromitsu Shiina, Shingo Nakamura, Nobuyuki Kobayashi, Fumio Kitagawa.: Detecting movie segments using mixed normal models and mixed beta model in VOD lectures with Japanese subtitles, 査読有, International Conference on Applied and Theoretical Information Systems Research 2013, pp. 1-12, 2013.

[学会発表](計 31件)

- (1) 塩飽朝美, 椎名広光: 単語評価推定に機

- 械学習を利用したアンケート自由回答欄評価, 2015年度第6回研究会, 教育システム情報学会研究報告, Vol.30, No. 6, pp.49-56, 2016年3月19日, 東京工芸大学.
- (2) 竹村直規, 椎名広光: Wikipedia と WordNet の多言語リンクを用いた留学生向け講義データ検索, 2015年度第6回研究会, 教育システム情報学会研究報告, Vol.30, No. 6, pp.21-24, 2016年3月19日, 東京工芸大学.
- (3) 三原隆義, 塩飽朝美, 小林伸行, 椎名広光: 分類木の類似性を利用した商品コメント評価, 言語処理学会第22回年次大会, pp.1061-1064, 2016年3月10日, 東北大学.
- (4) 武田昌大, 竹村直規, 小林伸行, 椎名広光: Wikipedia を用いた木構造カーネルによるコメント分類, 言語処理学会第22回年次大会, pp.1077-1080, 2016年3月10日, 東北大学.
- (5) 三原隆義, 椎名広光, 柳貴久男: TF・IDFを用いた類似性評価による商品推薦, 計算機統計学会第29回シンポジウム, pp.73-76, 2015年11月27日, 釧路市生涯学習センター.
- (6) Nobuyuki Kobayashi, Takayoshi Hiromitsu Shiina, Kikuo Yanagi: Classification by Approximation using the Neighborhood Range Estimation of Learning Data, 2015 International Workshop for JSCS 30th Anniversary in Okinawa. pp. 13-14, October, 30, 2015, Okinawa Institute of Science and Technology Graduate University, Okinawa.
- (7) 三原隆義, 椎名広光: 学習データの近傍幅の推定法を用いた近似による分類法, 平成27年度電子・情報関連学会中国支部連合大会, 21-6, 2015年10月17日, 山口大学.
- (8) 塩飽朝美, 椎名広光: 単語間の類似性による極性分析を利用した自由回答からの教員・講義評価, 平成27年度電子・情報関連学会中国支部連合大会, 27-10, 2015年10月17日, 山口大学.
- (9) 武田昌大, 椎名広光: 木構造カーネルを用いたSVMによる観光動画検索, 平成27年度電子・情報関連学会中国支部連合大会, 25-7, 2015年10月17日, 山口大学.
- (10) 竹村直規, 椎名広光: WordNet 及び word2vec を使用した自由アンケートの評価, 平成27年度電子・情報関連学会中国支部連合大会, 25-6, 2015年10月17日, 山口大学.
- (11) 塩飽朝美, 柳貴久男, 椎名広光, 小林伸行, 語彙の極性分析を利用した自由回答からの重要項目の抽出と特徴分析, 教育システム情報学会第40回全国大会, pp.215-216, 2015年9月2日, 徳島大学.
- (12) 中村慎吾, 小林伸行, 椎名広光, 北川文夫: 字幕付きVOD講義における発話内容のトピック分析, 言語処理学会第21回年次大会, pp.237-240, 2015年3月19日, 京都大学.
- (13) 三原隆義, 小林伸行, 椎名広光, 北川文夫: データ周辺の生起確率推定による機械学習によるコメントの評価クラス分類, 言語処理学会第21回年次大会, pp.908-911, 2015年3月19日, 京都大学.
- (14) 坂根耕平, 小林伸行, 椎名広光, 北川文夫: 非漢字語圏向け漢字及びプログラミングの学習システム, 電子情報通信学会教育工学研究会, ET2014-86, pp.7-12, 2015年3月14日, 四国大学交流プラザ.
- (15) 椎名広光, 武田昌大: コメント類似性を評価した観光施設動画の検索, OUS フォーラム 2014, pp.70, 2014年11月20日, 岡山プラザホテル.
- (16) 三原隆義, 椎名広光, 柳貴久男: 正規分布による生起確率を近似する機械学習法のパラメータ最適化, 日本計算機統計学会第28回シンポジウム, pp.85-88, 2014年11月14日, 沖縄科学技術大学院大学.
- (17) 中村慎吾, 椎名広光: VOD 講義の検索語に対する単語共起度による共起語の散布図表示システム, 平成26年度電子・情報関連学会中国支部連合大会, pp.22-23, 2014年10月25日, 福山大学.
- (18) 三原隆義, 椎名広光: 学習データの近傍の生起確率に正規分布の近似を利用した時系列データの分類学習, 平成26年度電子・情報関連学会中国支部連合大会, pp.179-180, 2014年10月25日, 福山大学.
- (19) 坂根耕平, 椎名広光: 講義に連動した非漢字語圏向け漢字学習システムとフリガナ表示システム, 平成26年度電子・情報関連学会中国支部連合大会, pp.16-17, 2014年10月25日, 福山大学.
- (20) 坂根耕平, 椎名広光, 北川文夫, 小林伸行: 留学生を対象としたロジスティック関数による漢字のフリガナ表示システム, 教育システム情報学会第39回全国大会, pp.51-52, 2014年9月10日, 和歌山大学.
- (21) 中村慎吾, 椎名広光, 北川文夫, 小林伸行: VOD 講義の話題区間共起語による類似性検出, 教育システム情報学会第39回全国大会, pp.53-pp.54, 2014年9月10日, 和歌山大学.
- (22) 三原隆義, 椎名広光, 柳貴久男: 正規分布で学習データの生起確率を近似する機械学習法の提案, 日本行動計量学会

第 41 回大会 抄録集, pp.132-135, 2014 年 9 月 3 日, 東北大学.

- (23) 友廣翔太, 小林伸行, 椎名広光: 漢字難易度を素性としたベイズ推定による新聞記事の難易度推定, 言語処理学会第 20 回年次大会, pp.928-931, 2014 年 3 月 20 日, 北海道大学.
- (24) 藏培慶, 小林伸行, 椎名広光: 単語難易度推定による中日単語学習システム, 言語処理学会 第 20 回年次大会, pp.113-116, 2014 年 3 月 18 日, 北海道大学.
- (25) 中村禎吾, 坂根耕平, 椎名広光: 混合ベータ分布モデルを用いた共起語頻度による VOD 講義の関連映像区間の検出, 言語処理学会第 20 回年次大会, pp.250-253, 2014 年 3 月 18 日, 北海道大学.
- (26) 坂根耕平, 椎名広光, 北川文夫: 共起語グラフによる VOD 講義のスライド内容と音声字幕の対応付けシステム開発, 教育システム情報学会, 2013 年度学生研究発表会(四国会場), pp.13-14, 2014 年 3 月 10 日, 香川大学.
- (27) 中村禎吾, 坂根耕平, 椎名広光, 小林伸行: 混合ベータ分布モデルによる VOD 講義の映像区間の検出, 情報処理学会研究報告, Vol. 2013-DBS-158, No. 13, pp. 1-7, 2013 年 11 月 19 日, 京都大学.
- (28) 中村禎吾, 椎名広光, 北川文夫: ベータ分布による VOD 講義の話題区間の検出, pp.321-322, 平成 25 年度電子・情報関連学会中国支部連合大会, 2013 年 10 月 19 日, 岡山大学.
- (29) 友廣翔太, 椎名広光: 読み手の評価と漢字級別に基づく日本語文書の難易度の推定, pp.345-346, 平成 25 年度電子・情報関連学会中国支部連合大会, 2013 年 10 月 19 日, 岡山大学.
- (30) 坂根耕平, 椎名広光: Wikipedia における言語間のリンク情報の類似性評価, pp.347-348, 平成 25 年度電子・情報関連学会中国支部連合大会, 2013 年 10 月 19 日, 岡山大学.
- (31) 友廣翔太, 椎名広光, 柳貴久男: 漢字難易度を素性としたベイズ推定による Web ページのランキング提示手法, 日本行動計量学会第 41 回大会抄録集, pp. 224-227, 2013 年 9 月 6 日, 東洋大学.

6. 研究組織

(1) 研究代表者

椎名 広光 (SHIINA HIROMITSU)
岡山理科大学・総合情報学部・准教授
研究者番号: 40299178

(2) 連携研究者

小林 伸行 (KOBAYSHI NOBUYUKI)
山陽学園大学・総合人間学部・講師
研究者番号: 90412270