

令和 4 年 6 月 4 日現在

機関番号：12102

研究種目：基盤研究(B) (一般)

研究期間：2019～2021

課題番号：19H04417

研究課題名(和文) 解説型ウェブページの分かり易さ・見易さの自動評定とそれを用いたページ推薦システム

研究課題名(英文) Measuring Beginner Friendliness / Visual Intelligibility of Web Pages explaining Academic Concepts and Procedures / Know-hows and their Recommendation System

研究代表者

宇津呂 武仁 (Utsuro, Takehito)

筑波大学・システム情報系・教授

研究者番号：90263433

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：本研究では、特定の物事、概念、出来事、手順等の解説を掲載しその内容を分かり易く伝えるウェブページを「解説型ウェブページ」と呼ぶ。本研究では特に、図解等の視覚効果を多く含む「解説型ウェブページ」である理数系学術用語解説ページを代表例として、分かり易さと見易さの因子を特定した。そして、「解説型ウェブページ」を自動評定するとともに評定結果の理由を提示するシステムを開発した。さらに、習熟者・初心者、大まかな概念が知りたい・詳細な解説が欲しい、等を考慮して「解説型ウェブページ」を推薦するシステムを開発した。

研究成果の学術的意義や社会的意義

本研究では、用語解説ウェブページの分かり易さ・見易さを評定する際の因子となる特徴量についてその分析結果を示している。また、本研究では、「解説型ウェブページ」を自動評定するとともに評定結果の理由が提示できることを示している。これらの成果により、ウェブ学習者の視点では、ウェブ検索を用いた学術用語学習における学習効率の改善につながる。また、「解説型ウェブページ」開発者の視点では、評定結果の理由を知ることができ、より容易にページを改善することが可能となる。以上の点で、本研究の学術的・社会的意義は大きい。

研究成果の概要(英文)：In this project, we focused on Web pages explaining certain concepts, events, procedures etc. in a beginner friendly manner. In those Web pages, visual intelligibility in Web page layout and beginner friendly Web page texts are the requirements of Web contents for beginners. As a typical case study, this project focused on the factors of Web page visual intelligibility as well as beginner friendliness for explaining academic concepts. Then, this project proposed how to measure those factors as well as to identify those factors. Finally, in this project, we developed a system for recommending those Web pages explaining academic concepts based on those factors.

研究分野：情報工学

キーワード：学術用語解説ウェブページ 解説型ウェブページ 深層学習 ページ推薦 手順・ノウハウ解説

1. 研究開始当初の背景

研究開始当初、ウェブ検索で「分かり易く見易い用語解説」が上位に得られるか否かを検証した結果においては、十分に高いと言える性能は達成できなかった。そこで、この改善のため、本研究では、用語解説ウェブページの分かり易さ・見易さを評定する際の因子となる特徴量について分析すること、深層学習を用いてそれらの特徴量を自動評定することを研究課題とすることとした。また、当初は、AI 技術によって判定結果は得られるものの、判定理由を知ることができない点が問題視されており、「説明可能な AI」技術が注目され始めていた。

2. 研究の目的

本研究では、特定の物事、概念、出来事、手順等の解説を掲載しその内容を分かり易く伝えるウェブページを「解説型ウェブページ」と呼ぶ。本研究では特に、図解等の視覚効果を多く含む「解説型ウェブページ」である理数系学術用語解説ページを代表例として、分かり易さと見易さの因子を特定する。そして、それらの特定された因子をふまえて、「解説型ウェブページ」を自動評定するシステムを開発する。特に、このシステムでは、「説明可能な AI」技術に着目して、評定結果の理由を提示するシステムを開発する。さらに、習熟者・初心者、大まかな概念が知りたい・詳細な解説が欲しい、等を考慮して「解説型ウェブページ」を推薦するシステムを開発する。

3. 研究の方法

本研究では、まず、学術用語解説ウェブページの事例データベースを構築する。さらに、この事例データベースを対象として、学術用語解説ウェブページにおける利用者像別の分かり易さ・見易さの因子を同定したデータセットの作成を行う。次に、構築された事例データベースを対象として、分かり易さ・見易さの詳細な因子について分析を行う。そして、この分析結果をふまえることにより、「解説型ウェブページ」を自動評定するとともに評定結果の理由を提示するシステムを開発する。さらに、この自動評定結果をふまえて、学術用語解説サイト・ページ群の推薦を行うシステムについて研究を行う。

4. 研究成果

- (1) 学術用語解説ウェブページ事例データベースを構築した。この事例データベースにおいては、分かり易さ・見易さの因子となり得る情報を収集するとともに、全体評定・分かり易さ・見易さの良否、および、各因子の良否を格納した。
- (2) (1)の学術用語解説ウェブページ事例データベースを対象として、「分かり易さ」・「見易さ」の詳細な因子についての分析を行った。この研究では、特に、「分かり易さ」と「見易さ」のうち、「分かり易さ」のみが充足され、「見易さ」が充足されない場合、および、逆に、「見易さ」のみが充足され、「分かり易さ」が充足されない場合に焦点を当て、それぞれ「見易さ」あるいは「分かり易さ」を損なう因子群を網羅的に分析した。
- (3) (2)の成果をふまえて、「解説型ウェブページ」を自動評定するシステムについて研究を行った。事前学習済みのモデルとして、1,000 種の物体識別用大規模汎用深層モデル ResNet を用いた。事前学習済みの ResNet モデルに対して、見易いページ・見易くないページ各数百ページを教師用事例として与えて転移学習 (fine-tuning) し、「見易さ」自動評定システムを開発した。そして、ResNet による見易さ自動評定結果の理由を提示する方式として、Grad-CAM に基づく方式を実現した。この方式では、見易さ自動評定時に重要な手がかりを与えた画像上の領域を示すことができ、この方式により、一定以上の精度によって、人間の直観に合う領域を示すことができた。
- (4) (3)の発展として、深層学習手法に基づき、ページ構成要素(図 1)の同定モデル、ページ構成要素をふまえた見易さの自動評定モデル、および、ページ構成要素をふまえた見易さの自動評定理由提示モデルのモデル化を行った(図 2)。そして、多様な学術分野を対象として、深層学習モデル訓練用のページ構成要素データセットを作成した。さらに、作成したページ構成要素データセットを用いて、各モデルの訓練・評価を行い、提案モデルの有効性を示した。具体的には、大規模データセットを用いて事前訓練された物体検出モデルである YOLOv5 モデルを利用することにより、ページ構成要素の同定を自動で行う手法を提案した。ページ構成要素を考慮した見易さ自動評定タスクでは、ResNet-50 の部分アーキテクチャを利用して

モデルを設計し、ページ画像に加えてページ構成要素の領域情報を入力することで、自動評価精度を向上させることができた。また、ページ構成要素を考慮した理由提示タスクにおいては、見易さ自動評価モデルの勾配を利用することで、自動評価結果に最も寄与したページ構成要素を特定した。これにより、見易さ自動評価に加えてページ構成要素を特定した理由提示を実現することができた。

図1: 用語解説ウェブページの構成要素

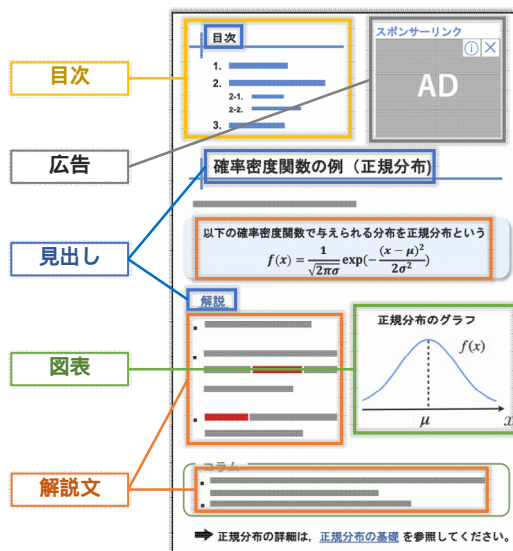
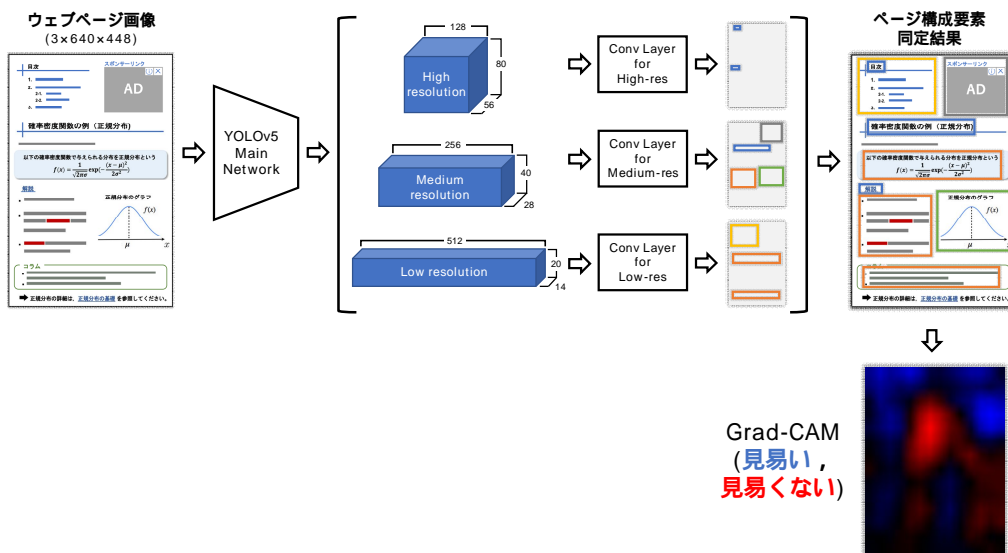


図2: 用語解説ウェブページの構成要素同定および見易さの評価



- (5) 学術用語解説ウェブページの分かり易さ・見易さに基づく閲覧・推薦システムの研究項目の一環として、複数の学術用語解説サイトのページに対して見易さの自動評価モデルを適用し、その評価結果をふまえて、評価の高い順に学術用語解説サイトを並べ替えて提示する閲覧インタフェースを作成した。そして、理数系学術用語解説ページを代表例としてこの閲覧インタフェースを適用した。
- (6) 学術用語解説ウェブページにおける利用者像別の分かり易さ・見易さの因子を同定したデータセットの作成を行った。具体的には、学術用語解説ウェブページの利用者像の一例として、<初学>：これから学習する初学者レベルで、この学術分野の全体像を体系的に学びたい、<習熟>：この学術分野は既に学習途中で中級レベルであり詳細な解説を求めている、の二通りを想定し、利用者像別に、分かり易さ・見易さの因子を手で同定したデータセットを作成した。

- (7) 「解説型ウェブページ」の閲覧・推薦システムの研究項目の一環として、(6)のデータセットにおける、利用者像別の分かり易さ・見易さの因子をふまえ、複数の学術用語解説サイトの間の特性の違いを見比べながら閲覧支援を行うシステム、および、学術用語解説サイト・ページ群の推薦を行うシステムについて研究を行った。
- (8) 手順・ノウハウ解説ウェブページ・サイトの推薦および手順・ノウハウを回答する手法の開発を行った。特に、あるまとまった学術分野における用語解説集、あるまとまったドメインにおける手順・ノウハウ、等を対象とした解説型ウェブサイトにおいて、機械読解モデルの枠組みにもとづき手順・ノウハウに関する質問に回答する手法を開発した。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Yuhei Oga, Kosei Soda, Kazuki Tanaka, Takehito Utsuro, Yasuhide Kawada
2. 発表標題 Annotation of Knowledge Amount and Learning Level to Web Pages explaining Academic Concepts
3. 学会等名 第35回人工知能学会全国大会(国際セッション) (国際学会)
4. 発表年 2021年

1. 発表者名 白 書霆, 李 廷軒, 鈴木 勢至, 宇津呂 武仁, 河田 容英
2. 発表標題 ノウハウ読解用コンテキストデータセットの作成および評価
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 李 廷軒, 白 書霆, 鈴木 勢至, 宇津呂 武仁, 河田 容英
2. 発表標題 コミュニティQAサイト上の質問回答事例に対するノウハウ読解
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 Tingxuan Li, Shuting Bai, Fuzhu Zhu, Takehito Utsuro
2. 発表標題 Evaluating a How-to Tip Machine Comprehension Model with QA Examples collected from a Community QA Site
3. 学会等名 the 35th Pacific Asia Conference on Language, Information and Computation (国際学会)
4. 発表年 2021年

1. 発表者名 白 書霆, 李 廷軒, 朱 福主, 宇津呂 武仁
2. 発表標題 コンテキストデータセットを用いたコンテキスト検索・ノウハウ読解
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2022年

1. 発表者名 大川遥平, 曾田耕生, 岡田 心太郎, 宇津呂 武仁, 河田 容英, 神門 典子
2. 発表標題 学術用語解説ウェブサイト群の見易さ評価結果閲覧インタフェース
3. 学会等名 第34回人工知能学会全国大会
4. 発表年 2020年

1. 発表者名 曾田 耕生, 大賀 悠平, 岡田 心太郎, 宇津呂 武仁, 河田 容英
2. 発表標題 大規模収集した学術用語解説ウェブページ群の見易さ評価結果閲覧インタフェース
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2021年

1. 発表者名 岡田 心太郎, 曾田 耕生, 大賀 悠平, 宇津呂 武仁, 河田 容英
2. 発表標題 学術用語解説ウェブページにおけるページ構成要素を考慮した見易さの評価および理由提示
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2021年

1. 発表者名 大賀 悠平, 岡田 心太郎, 曾田 耕生, 宇津呂 武仁, 河田 容英
2. 発表標題 学術用語解説ウェブページの充実度・対象読者の習熟度評定データセットの作成
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2021年

1. 発表者名 廣花智遥, 岡田心太郎, 宇津呂武仁, 河田容英, 神門典子
2. 発表標題 学術用語解説ウェブページの良否評定のための分かり易さ・見易さ因子の分析
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 Shintaro Okada, Chiharu Hirohana, Kota Kawaguchi, Kosei Soda, Takehito Utsuro, Yasuhide Kawada, Noriko Kando
2. 発表標題 Identifying factors of visual intelligibility of Web pages explaining academic concepts
3. 学会等名 2019 KDD Workshop on Deep Learning for Education (国際学会)
4. 発表年 2019年

1. 発表者名 曾田耕生, 大川遥平, 岡田心太郎, 廣花智遥, 宇津呂武仁, 河田容英, 神門典子
2. 発表標題 学術用語解説ウェブページ見易さ評定モデルのサイト単位適用事例の分析
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

ニューラルネットワークを用いたウェブページの見易さ判定
<http://nlp.iit.tsukuba.ac.jp/research/listxx-NN-CNN-Vis.html>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	吉岡 真治 (Yoshioka Masaharu) (40290879)	北海道大学・情報科学研究院・教授 (10101)	
研究分担者	乾 孝司 (Inui Takashi) (60397031)	筑波大学・システム情報系・准教授 (12102)	
研究分担者	亀田 能成 (Kameda Yoshinari) (70283637)	筑波大学・計算科学研究センター・教授 (12102)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------