

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 27 日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2011～2015

課題番号：23700113

研究課題名(和文)集合知による情報の信頼度算出手法

研究課題名(英文)A research on measuring quality of information using collective intelligence

研究代表者

鈴木 優 (Suzuki, Yu)

奈良先端科学技術大学院大学・情報科学研究科・特任准教授

研究者番号：40388111

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：Web上のデータは、多様性の増大と情報量の爆発が起きている。これは主に、UGC(User Generated Contents)が容易に生成可能となったことに起因する。一般にUGCの情報は質の面で玉石混淆であるが、書籍のように十分推敲された情報と比較すると質が相対的に低下している点が問題である。本研究課題では、Web上で公開されている情報の外部要因、内部要因を用いることによって、爆発的に増大し続けているWeb上のデータから質の高い情報を抽出するための基盤技術として、外部要因、内部要因の分析、体系化技術、統合利活用技術について実用的な時間で計算可能であるアルゴリズムの開発を行った。

研究成果の概要(英文)：In the Web, there are many types of, a lot amount of data. This is because, many persons publish information on the Web, which is called UGC (User Generated Contents). Generally, the quality of these information vary, but the quality should be lower than reviewed contents. In this research, we automatically measure the quality of Web contents using the outside element and the inside element of the information.

研究分野：情報検索，データマイニング

キーワード：質 Wikipedia 内部要因 外部要因 編集行動解析

1. 研究開始当初の背景

近年増加している情報の一つに UGC (User-Generated Contents; 閲覧者が編集可能なコンテンツ) が挙げられる。このようなコンテンツの例は Wikipedia やヤフー知恵袋などである。これらのコンテンツに含まれる情報は、必ずしも正確な情報、分かりやすい情報ばかりではなく、質の低い情報も数多く含まれている。そのため、情報の質を自動的に算出する方法を構築することが急務である。

2. 研究の目的

本研究では、大量の UGC からそのコンテンツの質を算出する方法の研究開発を行う。現在までに、情報検索、テキストマイニングに関する研究は一般に行われているが、コンテンツそのものの質を計測する方法は研究が不十分である。特に、インターネット上の情報が社会に与える影響が大きくなるに従って、重要な研究課題と予想される。また、情報量の増加に伴って、大量の情報に対しても実用的な時間で計測が可能であることは重要である。

3. 研究の方法

本提案では、情報の質を測定するための方法を確立した。情報を測定するための方法として、提案者は次の二つの方法によって情報の質を計測した。

- 外的要因に基づく測定方法: 情報に付随する情報として著者の情報と閲覧者の情報を利用した。つまり、より多くの閲覧者に閲覧される情報であり、さらにその情報が消去されない場合、その情報を記述した著者は信頼できると考えられる。
- 内的要因に基づく測定方法: 信頼度の高い情報に含まれている可能性が高い単語、低い単語が存在する。そのため、信頼度に関係すると考えられる単語を特定し、その単語の出現度合によって信頼度を算出した。

4. 研究成果

Wikipedia の外的要因に基づく質の計測を行うために、編集履歴を用いる方法、h-index を用いる方法など、様々な測定方法を用いた。基本的な考え方として、ある記述が Wikipedia に書かれ、その記述を多くの人が書き換えたとき、その記述の質は低く、多くの人がそのまま残存させたときその記述の質は高いと考えた。この考え方を基に記事の質を算出し、Web サービスとして公開した。ところが新たな問題として、記述が書き換えられたかどうかを判定することが困難であるという点が発生した。もしある記述が別の編集者によって大幅に書き換えられたとしても、その記述が意味することは同じである場合もあれば、異なる場合もある。書き換えられる前と後で共通する文字列の数だけでは、記述の意味が変化したかどうかを測定することはできない。そこで、クラウドソーシ

ングを用いて意味の変化を測定した。その結果、質の測定精度は向上した。

5. 主な発表論文等

[雑誌論文](計5件)

Yu Suzuki: "Quality Assessment of Wikipedia Articles Using h-index", Journal of Information Processing (JIP), 査読有り, vol.23, No.1, pp.22-30, January 2015

Yuya Fujiwara, Yu Suzuki, Yukio Konishi, Akiyo Nadamoto: "Extracting Lack of Information on Wikipedia by Comparing Multilingual Articles", International Journal of Business Intelligence and Data Mining (IJBIDM), 査読有り, Vol.8, No.1, pp. 45-60, Inderscience Publishers, 2013

鈴木 優: 「Wikipedia における情報の質」(サーベイ論文), 情報処理学会論文誌: データベース, 査読有り, Vol. 6, No. 4 (TOD59), pp. 46-58, 2013年9月

Yu Suzuki: "Effects of Implicit Positive Ratings for Quality Assessment of Wikipedia Articles", Journal of Information Processing (JIP), 査読有り, Vol. 21, No. 2, Information Processing Society of Japan, pp. 342 - 348, April 2013

横山 拓也, 石川 佳治, 鈴木 優: 「Hadoop 環境における空間分割による並列全 k 近傍問合せ処理」, 日本データベース学会論文誌, 査読有り, Vol. 11, No. 1, pp. 25-30, 2012年6月

[学会発表](計19件)

Yu Suzuki, Satoshi Nakamura: "Assessing the Quality of Wikipedia Editors through Crowdsourcing", Proceedings of Wiki Workshop 2016 in conjunction with WWW 2016, Montreal, Canada, April 2016

鈴木 優: 「ビッグデータと可視化」, 第11回協定講座シンポジウム 計算科学とビジュアル・アナリティクス, 神戸大学百周年記念館(神戸), 2016年3月

水上 雅博, Graham Neubig, 吉野 幸一郎, Sakriani Sakti, 鈴木 優, 中村 哲: 「快適度推定に基づく用例ベース対話システム」, 言語処理学会第22回年次大会(NLP2016), 東北大学(宮城県仙台市), 2016年3月

芥子 育雄, 鈴木 優, 吉野 幸一郎, 大原 一人, 向井 理朗, 中村 哲: 「単語・パラグラフの分散表現を用いた Twitter からの日本語評判情報抽出」, 第8回データ工学と情報マネジメントに関するフォーラム論文集, ヒルトン福岡シーホーク(福岡), 2016年2月

Hiromitsu Ohara, Yu Suzuki, Akiyo Nadamoto: "Detection of Missing Tweets based on Browsing Interval and Topic Granularity", iiWAS2015, Brussels, Belgium, December 2015

前田若菜, 鈴木優, 吉野幸一郎, Graham Neubig, 中村哲: 「ソーシャルメディアにおける非構造化テキストデータの k-匿名化によるプライバシー保護」, WebDB Forum 2015, 芝浦工業大学(東京), 2015年11月

鈴木 優, 吉野幸一郎, 中村 哲: 「クラウドソーシングによる Wikipedia 編集者の質推定」, 信学技報・電子情報通信学会データ工学研究会, 東大寺総合文化センター(奈良市), 2015年8月

Yu Suzuki, Masatoshi Yoshikawa: "Assessing Quality of Wikipedia Articles using Mutual Evaluation of Editors and Texts", Proceedings of ACM Conference on Information and Knowledge Management (CIKM 2013), ACM Press, San Francisco, CA, USA, 2013

Akira Nakamura, Yu Suzuki, Yoshiharu Ishikawa: "Clustering Editors of Wikipedia by Editor's Biases", Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence (WI 2013), IEEE CS Press, Atlanta, USA, 2013

Yuya Fujiwara, Yu Suzuki, Yukio Konishi, Akiyo Nadamoto: "Complementary Information for Wikipedia by Comparing Multilingual Articles", Proceedings of Asia-Pacific Web Conference (APWeb 2013), Springer Lecture Notes in Computer Science, pp. 260 - 267, Sydney, Australia, April 2013.

Yuya Fujiwara, Yu Suzuki, Yukio Konishi, Akiyo Nadamoto: "Extracting Lack of Information on Wikipedia by Comparing Multilingual Articles", Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (iiWAS2012), ACM Press, Bali, Indonesia, Dec. 2012

Yu Suzuki, Yuya Fujiwara, Yukio Konishi, Akiyo Nadamoto: "Good Quality Complementary Information for Multilingual Wikipedia", Web Information System Engineering (WISE 2012), Springer Lecture Notes in Computer Science, pp. 185-198, Paphos, Cyprus, Nov. 2012

Yu Suzuki, Masatoshi Yoshikawa: "Mutual Evaluation of Editors and Texts for Assessing Quality of Wikipedia Articles", Proceedings of

the 8th International Symposium on Wikis and Open Collaboration (WikiSym 2012), Online, ACM Press, Linz, Austria, Aug. 2012

Yu Suzuki: "Assessing Quality Values of Wikipedia Articles Using Implicit Positive and Negative Ratings", Web-Age Information Management (WAIM 2012), pp.127-138, Springer Lecture Notes in Computer Science, Harbin, China, Aug. 2012

Takuya Yokoyama, Yoshiharu Ishikawa, Yu Suzuki, "Processing All k-Nearest Neighbor Queries in Hadoop", Web-Age Information Management (WAIM 2012), LNCS 7418, pp. 346-351, Harbin, China, August 2012

Yu Suzuki, Masatoshi Yoshikawa: "QualityRank: Assessing Quality of Wikipedia Articles by Mutually Evaluating Editors and Text", Proceedings of the 23rd ACM Conference on Hypertext and Social Media (HT 2012), pp. 307-308, ACM Press, Milwaukee, WI, USA, June 2012

Yuya Fujiwara, Yu Suzuki, Yukio Konishi, Akiyo Nadamoto: "Extracting Difference Information from Multilingual Wikipedia", Asia-Pacific Web Conference (APWeb 2012), Springer Lecture Notes in Computer Science, pp. 496-503, Kunming, China, April 2012

Yu Suzuki, Akiyo Nadamoto: "Credibility Assessment using Wikipedia for Messages on Social Network Services", Proceedings of IEEE International Conference on Social Computing and its Applications (SCA 2011), IEEE CS Press, pp. 887-894, Sydney, Australia, December 2011

Akiyo Nadamoto, Yu Suzuki, Takeshi Abekawa: "Gist of a Thread in Social Network Services based on Credibility of Wikipedia", Proceedings of the 44th Annual Hawaii International Conference on System Services (HICSS 2011), IEEE CS Press, pp. 1-10, Hawaii, USA, January 2011

〔図書〕(計2件)

NAIST-IS書籍出版委員会 中村 哲, 戸田智基, 鈴木 優ほか: 「シンギュラリティ: 限界突破を目指した最先端研究」, 近代科学社, 2016年, 総ページ数324ページ
監訳: 田中克己, 角谷和俊, 翻訳: 大島 裕明, 岡本 真, 加藤 誠, 北山 大輔, 山本 岳洋, 小林 竜己, 是津 耕司, 山本 祐輔, 中島 伸介, 宮森 恒, 鈴木 優, 中村 聡史: 「情報検索のためのユーザー

インタフェース」, 共立出版, 2011 年,
総ページ数 415 ページ

〔その他〕〔総説〕(計 1 件)

平井(森田)晶, 西原 典孝, 大橋 美
名子, 富田 圭子, 小島 誠也, 大西
啓介, 稲村 真弥, 生島 早紀子, 佐藤
啓大, 小野 直亮, 黄 銘, 鈴木 優,
中村 哲, 加藤 博一, 金谷 重彦:
「新技術と研究紹介-ビッグデータ・フ
ード・サイエンス: 料理レシピと世界の食
品アクセスにおけるデータ・サイエンス」,
明日の食品産業, pp.33-44, 2016 年 3 月

6. 研究組織

(1) 研究代表者

鈴木 優 (SUZUKI, Yu)
奈良先端科学技術大学院大学情報科学研
究科・特任准教授
研究者番号: 40388111