

平成 30 年 5 月 21 日現在

機関番号：32663

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730162

研究課題名（和文）統計的データ解析と被験者実験を用いたテキスト・コミュニケーション研究

研究課題名（英文）A study on text communication using statistical data analyses and experiments

研究代表者

鈴木 崇史（SUZUKI, Takafumi）

東洋大学・社会学部・准教授

研究者番号：70507037

交付決定額（研究期間全体）：（直接経費） 2,700,000 円

研究成果の概要（和文）：本研究では、多様な媒体、ジャンルからテキストデータを収集し、統計的データ解析、被験者実験等により、テキスト特徴量とコミュニケーション・スタイルの関係を実証的に分析した。社会学、言語学、情報学的、社会情報学的見地から分析を行い、テキスト解析の方法論的拡張をはかるとともに、テキスト・コミュニケーションの実態を実証的に明らかにした。

研究成果の概要（英文）：This study collected a various types of text data from a wide variety of media, and conducted empirical researches on the relation between textual characteristics and communication styles using statistical data analyses and experiments. This study extended the methods on text analyses and obtained empirical findings on text communication.

研究分野：図書館情報学、人文情報学

キーワード：計算文体論 計量情報学 テキスト分析 統計的データ解析

1. 研究開始当初の背景

自然言語処理技術の発展と入手可能なデータの爆発的増大により、計算機ベースのテキスト解析は、その射程をますます広げている。多種・大量のテキストデータを新たな応用へと役立てることが学術的、社会的に急務の課題となっている。

同時に、Web の発達 は既存のメディア環境に変更を迫っている。掲示板、知識交換コミュニティ、口コミサイトなど CGM(消費者生成メディア) の興隆は、人々のコミュニケーション・スタイルに大きな影響を与えている。CGM は、多くの分野で期待され、利用されているが、同時に、なりすまし、炎上等、ネガティブな影響も看過できない。このような状況のもとで、多様化する人々のコミュニケーション・スタイルをテキスト解析技術によって実証的に明らかにすることは重要な研究課題である。

2. 研究の目的

本研究では、多種テキストを収集、整理し、テキスト特徴量とテキスト発信者のコミュニケーション・スタイルの関係を明らかにする。さらに、被験者実験を行うことで、それらとテキスト受信者の態度の関係を明らかにする。計算文体論の手法を用い、さらに、被験者実験を行うことで、社会科学分野のテキスト解析にも実質的、方法論的貢献をもたらす。

本研究により、多様なメディアの普及により変化しつつある現代のコミュニケーション・スタイルの一側面が実証的に明らかになる。これは、将来のさらに多様なメディア環境を構想し、そこでの、円滑なコミュニケーションのあり方を考えるために有用な知見を提供する。

3. 研究の方法

すでに収集、整理したテキストに加え、

新たに多種テキストを収集、整理する。Blog、Twitter 等 Web データをクロールし、新聞、歌詞等、OCR を利用し、多種テキストを作成する。作成したテキストファイルに対して、形態素解析、係り受け解析を適用し、基本的なテキスト統計量(文章長、パラグラフ長、文長、文字 (n-gram) 頻度、形態素 (n-gram) 頻度、共起頻度、係り受け頻度、要約特徴量)を計算する。

多変量解析を適用することで、多種テキストの位置関係を明らかにし、テキスト特徴量への影響要因(著者、ジャンル、時代ほか)を明らかにする。

コミュニケーション・スタイルの構成要素を類型化し、統制要因を排除した上で、テキスト特徴量から、各類型をあらわす指標を作成する。コミュニケーション類型ごとに、機械学習を適用する。分類に際しての変数の重要度を計算することで、両者を区別する特徴量を明らかにする。

4. 研究成果

まず、テキストデータ、実験データの収集を行うとともに、基礎的なテキストデータ解析を行うことで、諸テキストの特徴を明らかにした。Web データのクロール、各種テキストデータの収集、整理、整形、必要なメタデータの挿入を行い、作成したテキストファイルに対して、形態素解析、係り受け解析を適用し、基本的なテキスト統計量(文章長、パラグラフ長、文長、文字 (n-gram) 頻度、形態素 (n-gram) 頻度、共起頻度、係り受け頻度、要約特徴量)を計算した。多変量解析を適用することで、多種テキストの位置関係を明らかにし、テキスト特徴量への影響要因を調査した。また、フィールドでのインタビュー、アンケートを行い、コミュニケーションの基礎的特徴を検討した。機械学習を適用することで、コミュニケーション類型ごとのテキス

トの特徴を明らかにし、テキスト特徴量とコミュニケーション・スタイルの関係を明らかにした。社会学、言語学、情報学的、社会情報学的見地から分析、解釈を行い、方法論的、実証的知見を整理した。

5. 主な発表論文等

〔雑誌論文〕(計4件)

1. Yuichiro Kobayashi, Misaki Amagasa, Takafumi Suzuki. Investigating the Chronological Variation of Popular Song Lyrics Through Lexical Indices, 2(1), 90-107, 2017 [査読有り].
2. 吉元涼介・芳鐘冬樹・鈴木崇史. 質問紙調査を用いた商業集積地区とショッピングモールの利用行動の比較分析：コミュニケーションを中心として, 情報メディア研究, 15(1), 2016 [査読有り].
3. Yui Arakawa, Ryosuke Yoshimoto, Fuyuki Yoshikane, Takafumi Suzuki. Analyzing the content and text of tweets by Japanese academic researchers, Journal of Japanese Association for Digital Humanities, 1(1), 9, 2015 [査読有り].
4. Takafumi Suzuki, Mai Hosoya. Computational stylistic analysis of popular songs of Japanese female singer- songwriters, Digital Humanities Quarterly, 8(1), 2014 [査読有り].

〔学会発表〕(計10件)

1. Ryosuke Yoshimoto, Chizuko Takei, Takafumi Suzuki, Fuyuki Yoshikane. Network Structure Analysis of the Articles in the Japanese Civil Code, ALEAP 2017 (Bangkok, 2017.11.3).
2. Gento Kato, Takanori Fujiwara, Christian Collet, Tetsuro Kobayashi, Takafumi Suzuki. Threatening Event, National Identity and Network Dynamics of Motivated Information

Communication: Exploring Japanese Twitter during the Rise of Territorial Disputes, April through October 2012, APSA 2017 (San Francisco, 2017.9.1).

3. Yuichiro Kobayashi, Misaki Amagasa, Takafumi Suzuki. Investigating the chronological variation of lyrics of popular songs through lexical indices, QUALICO 2016 (Universitaet Trier, Trier, 2016.8.26).

4. Takafumi Suzuki. Iterative application of random forests to extract less frequently used but important features from texts, PACLING2015 (The Stones Hotel, Legian, Bali, 2015.5.19).

5. Takafumi Suzuki, Natsumi Yamashita. Analysis of perspectives in contemporary Japanese novels using computational stylistic methods, Digital Humanities 2014 (University of Lausanne, Lausanne, 2014.7.10).

6. Takafumi Suzuki, Maki Mizukami. Comparative analysis of purchasing behavior at physical and online bookstores by university students, ACTIS 2015 (Art Center of Kobe, Kobe, 2015.11.5-7).

7. Ryosuke Yoshimoto, Fuyuki Yoshikane, Takafumi Suzuki. A comparative analysis of shoppers' information behavior in commercial districts and shopping malls using a questionnaire survey, 2014 ANPOR Annual Conference (Toki Messe, Niigata, 2014. 11. 30).

8. 小林雄一郎・天笠美咲・鈴木崇史. 語彙指標を用いた流行歌の歌詞の通時的分析, 人文科学とコンピュータシンポジウム (同志社大学京田辺キャンパス, 京田辺, 2015年12月19日).

9. 吉元涼介・瀧澤啓太・宮田玲・芳鐘冬樹・鈴木崇史. 六法全書の参考条文を用いたネットワーク構造の分析, 第63回日本図書館情報学会研究大会発表論文集, 45-48(学習

院女子大学，東京，2015 年 10 月 17 日）.

10. 吉元涼介・芳鐘冬樹・鈴木崇史．質問紙調査を用いた商店街とショッピングモール利用行動の比較分析，日本行動計量学会第 42 回大会発表論文抄録集，376-380（東北大学川内北キャンパス，仙台，2014 年 9 月 5 日）.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

6．研究組織

(1)研究代表者

鈴木 崇史 (SUZUKI, Takafumi)

東洋大学・社会学部・准教授

研究者番号：70507037

(2)研究協力者

芳鐘 冬樹 (YOSHIKANE, Fuyuki)

COLLET, Christian

小林 哲郎 (KOBAYASHI, Tetsuro)

小林 雄一郎 (KOBAYASHI, Yuichiro)

宮田 玲 (MIYATA, Rei)

加藤 言人 (KATO, Gento)

竹井千寿子 (TAKEI, Chizuko)

吉元 涼介 (YOSHIMOTO, Ryosuke)

細谷 舞 (HOAOYA, Mai)

水上 麻希 (MIZUKAMI, Maki)

山下 夏実 (YAMASHITA, Natsumi)

瀧澤 啓太 (TAKIZAWA, Keita)

天笠 美咲 (AMAGASA, Misaki)