

平成 30 年 5 月 15 日現在

機関番号：17102

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00451

研究課題名(和文)利用者投稿型コンテンツの多様性動向分析

研究課題名(英文)Contents diversity analysis for consumer generated media contents

研究代表者

伊東 栄典 (Ito, Eisuke)

九州大学・情報基盤研究開発センター・准教授

研究者番号：90294991

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究はCGM(Consumer Generated Media)における(1)コンテンツの多様性動向の定量的評価と、(2)利用者のコンテンツ選択モデルの確立を目的とした。動画、小説、学術論文を対象にした。(1)について、多様性を定量的に評価する指標を3つ提案した。類似度総和指標を用いた結果、動画と小説については、多様性減少傾向を確認でき、学術論文は多様性あることも確認した。(2)について、コンテンツ閲覧数分布調査で、全体的に対数正規分布であること、ジャンル毎に分布が異なることを確認した。回帰分析で閲覧数を推定した結果、良い推定精度を得たものの利用者のコンテンツ選定を決定づける要因は確立できなかった。

研究成果の概要(英文)：This research aimed to (1) quantitative evaluation of content diversity trends in CGM (Consumer Generated Media), and (2) establishment of content selection model of users. Targeted CGMs are movies, novels and academic papers. In (1), I proposed three metrics for quantitative contents diversity. Using sum of similarity metric, I found that diversity are decreasing for movies and novels, but academic papers keep diversity. In (2), I found that pageview distribution is close to the log-normal distribution, but the distribution differs for each genre. As a result of regression analysis to estimate pageviews of contents, estimation accuracy was good, but establishment of the user's content selection model was incomplete.

研究分野：情報検索 情報統合 テキストマイニング

キーワード：コンテンツ CGM 多様性 類似度 回帰分析 深層学習

### 1. 研究開始当初の背景

サービス開始から数年経過した CGM (Consumer Generated Media) では、コンテンツの多様性喪失(画一化)が進んでいる。画一化は文化の発展を阻害する。

### 2. 研究の目的

本研究では「ニコニコ動画」の楽曲、「小説家になろう」の小説、学术论文を対象に多様性の動向を分析する。閲覧回数上位のコンテンツを対象に、タグやキーワードの種類数を時系列で見ることで、多様性喪失(画一化)を定量評価する。また、閲覧回数の分布から多様性喪失の遠因となる利用者のコンテンツ選択モデルを確立する。利用者の選択モデルは、新たなランキング手法の提案や、外部刺激によるコンテンツの多様性確保に寄与できる。比較対象として学术论文も分析する。

### 3. 研究の方法

コンテンツの画一化を測るためには、多様性の指標が必要である。そこで3つの指標を援用した。

- ・ 情報エントロピー
- ・ ジニ係数
- ・ Cos 類似度の総和

比較した。これらをニコニコ動画の動画、「小説家になろう」の小説、学术论文について適用し、推移を見る。

次に利用者のコンテンツ選択モデルの確立を試みた。まず動画と小説について閲覧回数の分布を調査する。次に、回帰分析で、その分布となるモデル構築を試みる。クラスタリングの適用により、どの分野に偏るため多様性が減少するのか調査する。

### 4. 研究成果

「ニコニコ動画」の楽曲、「小説家になろう」の小説について、タイトルおよびタグに出現する単語について、情報エントロピー、ジニ係数、Cos 類似度の総和の3つの指標の値を算出した。その結果、単語あたりの情報

エントロピーは値の変化が乏しく使えないことが分かった。経済格差を示すジニ係数を単語の頻度に適用した所、値の変化が見えないため、単語の多様性分析には使えないことが分かった。

Cos 類似度の総和指標を用いて多様性の動向を測定した。月ごとの新規投稿コンテンツについて測定した結果、多様性の減少を定量的に確認できた。図1にニコニコ動画の Cos 類似度総和、図2に Cos 類似度総和の推移を示す。

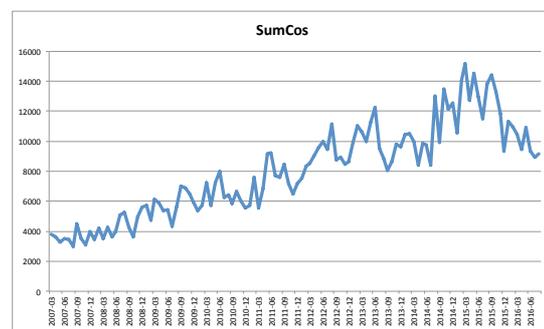


図1：ニコニコ動画 各月に投稿された動画のタグ単語の SumCos 値

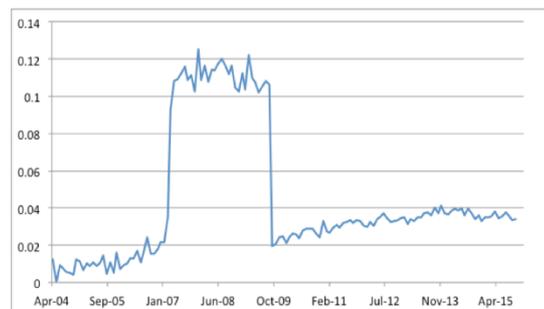


図2：小説家になろう 各月に投稿された小説のキーワード単語の SumCos 値水

学术论文についても同様の分析を適用した所、学术论文では多様性減少の傾向は見られなかった。次に LDA による潜在トピック分析を学术论文集合に適用した所、論文数が増えるとトピック数が増えるという、当たり前の結果が得られた。学术论文では多様性減少の傾向が無いことがわかったため、この後の分析から学术论文は除外した。

動画と小説について閲覧回数の分布を調査した。統計ソフトRを用いた非線形回帰分析で分布を調べた。その結果、動画と小説の両方が閲覧回数の分布は対数正規分布であることが分かった。別件で調査した学内電子メールの容量分布も対数正規分布となり、人間社会における普遍的な分布である可能性が見えた。

閲覧回数が対数分布となるコンテンツ選択のモデルを構築するため、回帰分析を適用した。ニコニコ動画の閲覧回数を線形回帰分析および深層学習（6層のマルチパーセプトロン）で分析した。その結果、高い精度で閲覧回数を推定出来たものの、コンテンツの選択モデル構築には至らなかった。多様性が減少しているならば、どこかの分野に偏っているはずである。クラスタリングの適用により、偏りを分析した。ニコニコ動画の動画タイトル・説明文・タグからなる文書をDoc2Vecでベクトル化してクラスタリングを適用した。しかしながら、おおまかな傾向は分かるものの、詳細な偏り分析まで至らなかった。

#### 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 6件）

- ① Eisuke Ito, Yuya Honda, Sachio Hirokawa: Empathy factor mining from reader comments of e-manga, Proc. of eKnow2018, IARIA, pp.107-112, Mar., 2018. (ISBN: 978-1-61208-620-0)
- ② Eisuke Ito, Yuya Honda: Keyword diversity trend of consumer generated novels, Proc. of ICES2017., pp.140-147, 2017.
- ③ Kazuhisa Noguchi, Tomoya Iida, Eisuke Ito: An analysis of CGM contents pageview using SIR Model and GBM, Proceedings of ICCTD2017 (2017 8th International Conference on Computer Technologies and Development) pp.19-21, 2017.
- ④ Kyohei Kamihata, Eisuke Ito: A quantitative contents diversity analysis on a consumer generated media site, Proc. of AROB 21st 2016 (The Twenty-First International Symposium

on Artificial Life and Robotics 2016), pp.436-440, 2016.

- ⑤ Zechen Wu, Eisuke Ito: Correlation Analysis between User's Emotional Comments and Popularity Measures Proc. of IIAI-AAI 2014/ESKM 2014, pp.280-283, September 3, 2014.
- ⑥ Xiao Lin, Eisuke Ito, Sachio Hirokawa: Chinese Tag analysis for foreign movie contents, Proc. of IEEE/ACIS ICIS2014, pp.163-166, June 4, 2014.

〔学会発表〕（計 9件）

- ① 柴田知親, 伊東栄典:回帰分析によるCGM動画再生回数推定, 第10回データ工学と情報マネジメントに関するフォーラム(DEIM2018), 日本データベース学会, C5-2, 2012.
- ② 佐嘉田悠樹, 伊東 栄典:クラスタリングによる利用者投稿動画の偏在性分析, 第10回データ工学と情報マネジメントに関するフォーラム(DEIM2018), 日本データベース学会, G4-2, 2012.
- ③ 飯田委哉, 伊東栄典, 佐嘉田悠樹: クラスタリングによるオンライン小説の多様性動向分析, 情報処理学会 火の国情報シンポジウム2018, B1-4, Mar.1, 2018.
- ④ 本田優也, 伊東栄典, 廣川佐千男: 読者コメント連動型漫画における人物の登場頻度分析, 人工知能学会 SIG-SW0-043-02, pp.02\_1-02\_6, Nov.25, 2017.
- ⑤ 佐嘉田悠樹, 伊東栄典: CGM 百科辞典を用いた利用者投稿動画クラスタリング, 平成29年度電気・情報関係学会九州支部連合大会, 11-2P-06, pp.544-545. 2017.
- ⑥ 柴田知親, 伊東栄典: 回帰分析を用いた動画再生回数の推定, 平成29年度電気・情報関係学会九州支部連合大会, 11-2A-08, pp.367-368. 2017.
- ⑦ 本田優也, 伊東栄典: 利用者投稿型小説サイトにおけるキーワードの多様性分析, 情報処理学会 火の国情報シンポジウム2017, A2-4, Mar.1, 2017.
- ⑧ 上畑恭平, 伊東栄典: タイトルと説明文に着目した利用者投稿サイト動画の多様性分析, 情報処理学会 火の国情報シンポジウム2016, 4A-3, Mar., 2016.
- ⑨ 上畑恭平, 伊東栄典: タグの類似度に着目した利用者投稿サイト動画の多様性分析, 信学技報, vol.115, no.381, AI2015-40, pp.83-88, 2015.

〔図書〕（計 0件）

〔産業財産権〕

○出願状況（計 0件）

○取得状況（計 0件）

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

伊東 栄典 (ITO, Eisuke)

九州大学・情報基盤研究開発センター・准  
教授

研究者番号：90294991

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし