

場所に焦点を当てた複数ブログの自動要約

Location-based Automatic Summarization of Multiple Blog Entries

前田剛 難波英嗣 竹澤寿幸
Tsuyoshi Maeda Hidetsugu Nanba Toshiyuki Takezawa

広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

1. はじめに

旅先の情報を収集するための情報源として、旅行ブログエントリが注目されている。旅行ブログエントリの利点として、最新の情報や地元の人しか知らないローカルな情報を入手することが可能である点が挙げられる。そこで我々は、旅行ブログエントリを自動的に検出し、各エントリを地図上にマッピングして提示するシステムを構築している。しかし、複数の旅行ブログエントリがある地点に集中した場合、これら全てを閲覧するには多大な時間と労力を要する。そこで本研究では、ある地点について述べられた複数の旅行ブログエントリの要約を行い、旅行者にとって有益な文と画像を提示するシステムを構築する。これにより、旅行者にとって、ある地点での特徴を迅速に捉えることができると考えられる。

2. 複数旅行ブログエントリの要約

本研究では、ある地点について述べられた旅行ブログエントリ中の文と画像をクラスタリングすることで要約を生成する。要約の生成手順は、以下の4つのステップから構成される。

(ステップ1) 文のクラスタリング

旅行ブログエントリ中のテキストを一文ごとに分割し、文単位でクラスタリングを行う。クラスタリングの際、各文に出現する名詞、動詞、形容詞の $tf*idf$ 値を用いる。

(ステップ2) 画像のクラスタリング

旅行ブログエントリに存在する画像の抽出し、画像のクラスタリングを行う。クラスタリングには、BoVW および Color Histogram により表現された画像特徴量を用いる。なお、ステップ1と2では、クラスタリングのツール bayon を用いる。

(ステップ3) 文クラスタと画像クラスタの対応付け

旅行ブログエントリ内のある一文の前後に画像が存在した場合、文にこの画像が属する画像クラスタを対応付ける。そして、文クラスタ内の一文と対となる画像クラスタのうち、最も多かった画像クラスタを文クラスタに対応付ける。図1は、文クラスタと画像クラスタの対応付けを図式化したものである。なお図1の場合、文クラスタに画像クラスタ1が対応付く。文の前後に存在する画像が、必ずしも関連しているとは限らない。そこで、クラスタ単位で対応付けを行うことにより、より関連の深い文と画像を表示させることが可能となる。

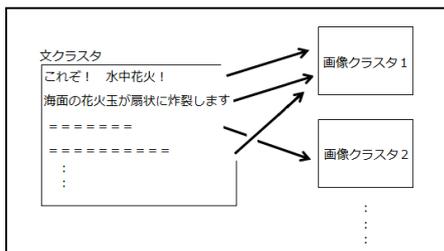


図1: 文クラスタと画像クラスタの対応付け

(ステップ4) 要約の生成

対応付いた文および画像クラスタから代表文、代表画像を選出し、要約を生成する。ある地点に関する旅行ブログエントリに多く出現する単語は、その地点を象徴する単語であり有益であると考えられる。有益な単語は、旅行ブログエントリに多く出現するため、これを含む文が属する文クラスタのサイズは大きくなる。したがって、最終的に表示する代表文は、最もサイズの大きい文クラスタの一文とし、代表画像は、その文クラスタに対応付いた画像クラスタの画像とする。

3. 実験

実験には、「厳島神社」について書かれた旅行ブログエントリ49件を用いた。厳島神社に関する代表文、代表画像の抽出例を図2に示す。



図2: 厳島神社に関する代表文と代表画像

本研究では、要約を生成することが目的であるが、そのためには、ステップ3において文クラスタに適切な画像クラスタを対応付けることが必要である。そこで、ステップ3についての評価を行うため、対応付けた文クラスタと画像クラスタ対の正解率を求める。実験の結果、全ての対応付いた70対のうち、正しい対は27対で、正解率は0.386であった。正解率を下げている要因として、対応付け以前に文のクラスタリングが適切でない場合が考えられる。本研究では、bayonを用いてクラスタリングを行ったが、文クラスタ70件中28件適切でないものが含まれていた。クラスタリングを正確に行った上で、提案手法のクラスタの対応付けを行えば、より高い正解率を得られると考えられる。そこで理想的なクラスタリングが得られた場合の正解率を計算したところ、0.643という結果が得られた。今後は、ステップ4において生成された要約に対する評価を検討していく予定である。

4. おわりに

本研究では、複数の旅行ブログエントリの要約を行う手法を提案した。代表文及び代表画像を表示することで、視覚的に旅行ブログエントリの内容を容易に理解することが可能になった。

発表論文

前田剛, 河野有希, 石野亜耶, 難波英嗣, 竹澤寿幸: 場所に焦点を当てた複数ブログの自動要約, DEIM 2014, 2014.