

オンラインショップで購入できない土産を提示するシステムの構築 Construction of a System to Suggest Souvenirs which are not Available at Online Shops

長尾 哲志†
Noriyuki Nagao

安藤 一秋‡
Kazuaki Ando

1. はじめに

旅行中の悩みの 1 つに土産の購入がある。その理由は、観光に時間を費やしたいので土産を選ぶ時間が確保できない、相手に喜ばれるものを買いたい却不知道のかわからないなど様々である。また、近年ではオンラインショップの増加により、多種多様な土産が時間と場所を選ばずに購入できるようになった。この点も土産の購入を難しくしていると考えられる。

現地でしか購入できない土産情報は、QA サイトや口コミサイト、ブログなどに散在している。しかし、現地でしか購入できない土産を一元的に提供しているサイトやサービスは存在しないため、土産情報を容易に入手できない。また、観光情報を対象とする研究[1-5]はいくつか存在するが、土産情報に特化したシステムは存在しない。

そこで本研究では、オンラインショップでは購入できない土産情報(品名、店名、評判など)を Web 上から収集・整理して、オンラインショップで購入できる土産と共にユーザーに提示するシステムの構築を目的とする。本研究では、オンラインショップで購入できない「飲食物」を現地でしか購入できない土産として定義する。通常、現地でしか購入できない土産でも、デパートなどの物産展などで購入できる場合があるが、そのような土産は対象外とする。

本稿では、まず、ブログに記述されている商品情報(品名やレビューなど)の量や内容などについて調査した結果を述べる。次に、現在検討中の土産情報源の構築と土産提示システムの概要を述べる。最後に、QA サイトから品名を抽出する手法を検討するための予備調査の結果を述べる。

2. ブログに記述される土産の調査

現地でしか購入できない土産に関する情報は、QA サイトや口コミサイト、ブログなどに散在している。我々の先行研究[6]で、QA サイトが土産情報源として有用であることを確認した。本稿では、Yahoo!ブログを対象として、ブログに記述されている土産情報(品名、レビューなど)の量や内容について調査する。なお、調査対象として、Yahoo!ブログの「菓子、デザート」カテゴリに対し、キーワード「土産」で検索した上位 100 件のブログを利用する。

調査の結果、100 件中 24 件(24.0%)の記事に品名や感想などの土産情報が記述されていた。残りの 76 件には、手作りお菓子やコンビニスイーツの紹介など、目的とする土産情報以外が記述されていた。そのため、ノイズを除去する手法の検討が必要である。

次に、土産情報が記述されていた 24 件の記事に対して、詳細を調査した。まず、会社名と地域名の出現傾向を調査

した結果、品名よりも前に会社名(店舗名)が出現する記事が 9 件(37.5%)存在した。そのため、会社名を手掛かりとして品名を抽出できる可能性がある。また、記事中に土産の画像を提示している記事が 22 件(91.6%)存在した。ブログからは土産画像も同時に取得できると可能性がある。次に、品名の出現有無について調査した結果、ブログの本文中に品名が提示されている記事が 18 件(75%)存在した。この内、16 件は土産に対する感想が述べられていた。したがって、ブログからは品名と共に土産に対するレビューが取得できる可能性がある。最後に、土産を渡す側と受け取る側のどちらの視点で記述されているかを調査した。その結果、商品を渡す側の視点で書かれた記事は確認できず、受け取る側は 8 件(33.3%)、残りの 16 件(66.7%)は、自分用に購入した視点であった。

以上の結果から、ブログには土産情報が埋もれていることが確認できた。また、QA サイトでは少なかった、画像やレビューが取得できる可能性があることも確認できた。さらに、ブログには、相手に渡すために購入した内容よりも自分用に購入した内容が多いことがわかった。

3. 土産情報源の構築

本研究では、QA サイトとブログを土産情報源として利用する。QA サイトとブログを併用することで、互いに不足する情報を補完できると考える。また、オンラインショップからも土産情報を収集し、QA サイトとブログから収集してきた土産情報と統合することで、土産情報 DB を構築する。図 1 に土産情報 DB を構築する処理の概要を示す。

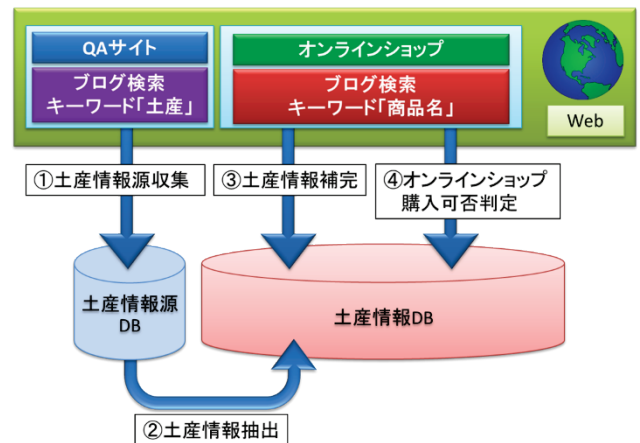


図 1: 土産情報 DB を構築する処理の概要

以下、土産情報 DB の構築手順を示す。

① 土産情報源の収集

QA サイトとブログから、キーワード「土産」を含む記事を収集し、土産情報源 DB に登録する。

† 香川大学 大学院工学研究科 Kagawa University Graduate School of Engineering

‡ 香川大学 工学部 Kagawa University Faculty of Engineering

② 土産情報の抽出

①で収集した記事から品名とその他の土産情報(会社名/店舗名, 評判など)を自動抽出して, 土産情報 DB に登録する。

③ 土産情報の補完

①で収集した記事に含まれていなかった土産情報(販売場所, ジャンル, 個数情報, 土産の写真, 賞味期限など)を検索エンジンや Web 上のショッピングモールなどを利用して補完し, 土産情報 DB に登録する。

④ オンラインショップで購入可否の判定

検索エンジンや Web 上のショッピングモールなどを利用して, 対象の商品がオンラインショップで購入可能であるかを判定する。ここでは, オンラインショップで購入できない商品を現地でしか購入できない土産とみなし, オンラインショップでの購入可否情報を付与し, 土産情報 DB に登録する。図 2 に判定処理の概要を示す。

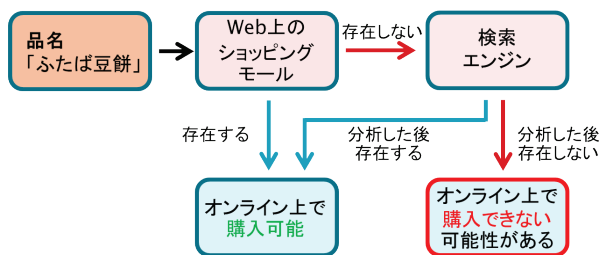


図 2: オンラインショップで購入できない土産判定

4. 土産提示システムの概要

提案システムは, 任意の場所情報に基づいた様々な土産情報(オンラインショップにおける購入可能割合や評判など)をユーザにランキング形式で提示する。

検索/提示方法として, 都道府県名や品名などから検索する方法と, 地図を利用して検索する方法の 2 つを実装する予定である。地図を使用した方法では, 現在地や任意の場所の位置情報を受け取り, 該当場所近辺の商品情報を地図上に提示することで, 視覚的な商品検索を可能とする。

提案システムでは, オンラインショップにおける購入可能割合を提示するため, オンラインショップで購入できない可能性の高い商品から提示することが可能になる。

5. QA サイトからの品名抽出調査

土産情報源構築における土産情報の抽出手法を検討するため, QA サイト (Yahoo!知恵袋) を対象に, 市町村名と会社名を利用した品名抽出の調査について述べる。

5.1. 市町村名を利用した品名の抽出

土産は, 場所と共に記述される可能性がある。そこで, 香川県を対象とした Yahoo!知恵袋の質問回答情報 (856 件) から, 市町村名 (語尾に市町村名あり/なしの 2 種類 (例: 高松市と高松)) を手がかり表現として品名候補の抽出調査を行う。本調査では, <市町村名>+ (にある | の) + <品名候補>+ (が | も) + (人気 | おすすめ | オススメ | お勧め | 有名) という規則を利用して, 品名候補を抽出する。なお, 市町村名は香川県の市町村名を利用する。

調査の結果, 語尾に市町村名があるものを使用した場合, 76 件 (8.87%), 語尾に市町村名がないものを使用した場合,

11 件 (1.28%) について品名候補 (文字列) が抽出できた。なお, 品名候補が品名のみでの事例は確認できなかった。

本調査で利用した規則では, 品名のみを抽出できなかったことから, 品名を記述する際, 品名と市町村名が 1 文中に同時に記述されにくい可能性があると考えられる。

5.2. 会社名を利用した品名の抽出

土産について記述する際, 会社名と品名が 1 文中に共起する可能性がある。そこで, 香川県を対象とした Yahoo!知恵袋の質問回答情報 (856 件) から, 会社名 (店舗名) を手がかり表現として, 品名候補の抽出調査を行う。調査に利用する規則は, <会社名>+<品名候補>+ (が | も) + (人気 | おすすめ | オススメ | お勧め | 有名) などである。また, 抽出した品名候補を手で正解, 準正解, 不正解の 3 つに分類する。使用する会社名は, 香川県の会社名 (335 社) である。なお, 事前調査より, 856 件には, 409 個の品名が出現することを確認している。

調査の結果, 正解が 62 件, 準正解が 26 件, 不正解が 321 件となり, 抽出精度は 15.2% (62/409) となった。準正解の 26 件を正解に加えた場合, 精度は少し向上して 21.5% (88/409) となった。次に, 856 件の回答情報から会社名を含む 256 文を抽出して, 会社名と品名が 1 文中で共起する割合を確認した結果, 35.9% (92/256) であった。以上から, 会社名も品名と同一文中に多くは記述されないと考えられる。

6. おわりに

本稿では, まず, ブログを対象とした土産情報の調査を行い, 100 件のブログ記事のうち 24 件で土産情報が記述されていることを確認した。次に, 土産情報源を構築するための処理と土産提示システムの概要について述べた。最後に, 市町村名または会社名を利用して, QA サイトから品名を抽出する調査を行い, 品名と市町村名, または会社名が 1 文中に共起する形で記述されにくい可能性を確認した。

今後は, Yahoo!知恵袋と Yahoo!ブログからノイズを除去しつつ, 効率よく土産情報を抽出するための手法について検討を行い, 土産提示システムの開発を行う。

参考文献

- [1] 石野他, “旅行ブログエントリからの観光情報の自動抽出”, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, (2010).
- [2] 奥他, “地域限定性スコアに基づく位置情報付きコンテンツからの地域限定語句の抽出”, IPSJ 論文誌 データベース, 5(3), pp. 97-116, (2012).
- [3] 上原他, “Web 上に混在する観光情報を活用した観光地推薦システム”, IEICE 技術研究報告, 112(367), pp.13-18, (2012).
- [4] 遠藤他, “特定地域に限定しない観光キーワードの自動抽出”, DEIM Forum 2014, E9-2, (2013).
- [5] 河野他, “小説や旅行記からの地理情報の自動抽出と編纂”, IPSJ 全国大会講演論文集, Vol.73, No.1, pp.785-786, (2011).
- [6] 長尾他, “オンラインショップで購入できない土産を提示するシステムの検討”, IEICE 2015 年総合大会論文集 (情報・システム 1), p.36, (2015).