

食品の履歴情報とブログ上の評判情報の統合検索システムの開発

中畑将吾[†] 室井淑栄[‡] 谷本美穂[†] 市橋剛[‡] 中野裕介[†] 安藤一秋[‡] 垂水浩幸[‡]

[†]香川大学大学院工学研究科 信頼性情報システム工学専攻

[‡]香川大学工学部 信頼性情報システム工学科

消費者が特定の商品について詳しく知りたい場合、生産履歴を調べたりブログ記事から他の消費者がどのように思っているかを参考にしたりする。情報を発信しているのは立場の違う「生産者」と「第三者」であるものの、どちらの情報も消費者にとって貴重な情報源である。そこで我々は、履歴情報とブログ上の評判情報を統合し、これらを検索システムのソースとして活用するための枠組を開発している。本枠組により、原材料の生産農家から加工業者さらに消費者までがそれぞれの立場で情報を発信できる。また、消費者は立場の異なるものが発信した情報を横断的に検索することができる。本稿では、履歴情報と評判情報を統合する意義について議論し、本枠組について述べる。

Development of a Search System Integrating Food Traceability Information and Reputation from Blog Entries

Shogo NAKAHATA[†] Yoshie MUROI[‡] Miho TANIMOTO[†] Tsuyoshi ICHIHASHI[‡]
Yusuke NAKANO[†] Kazuaki ANDO[‡] and Hiroyuki TARUMI[‡]

[†]Graduate School of Engineering, Kagawa University

[‡]Faculty of Engineering, Kagawa University

When consumers want to know detailed information of a product, they search traceability information or word of mouth information. Senders are in different position “producer” or “third party”, but information that they sent is invaluable sources for consumers. We integrate traceability information and reputation from blog entries and develop a framework that deals with both of these sources. With this framework, various people including producers of raw materials, food processors and consumers can transmit information from each viewpoint. Consumers can collect information from various sources. In this paper, we discuss the merit of integration of traceability information and reputation, and describe our framework.

1. はじめに

近年の相次ぐ食品事故、偽装表示問題を受け、食品の安全と安心が強く求められている。そこで一部のメーカ（例えば石井食品株式会社の OPEN ISHII[1]）や団体（例えば財団法人食品流通構造改善促進機構の SEICA[2]）では自社商品や業界商品の生産履歴情報などを Web 上に発信し、商品の安全性をアピールする動きがある。

このような背景から我々はこれまで、履歴情報

（添加物、アレルゲン物質、またそれらの原料はさらにどこから…）をデータ化するためのフォーマットを定義し、これらを発信・収集するシステムを提案してきた[3]。本システムの目的は、高品質商品を生産する優良な生産者がデータベーススキーマに定義されているデータ項目に縛られず自由に情報を発信できる、また農家のような一次生産者が二次生産者や流通業者を介さずに直接消費者に情報を発信できる枠組みを実現し、流通する情報の量・質を向上させることである。

さらに本年度から我々は、生産者による履歴情報に加えてブログ上の評判情報も統合可能な枠組を開発している。そして、これらを商品検索システムのソースとして活用する。本稿では、履歴情報と評判情報を統合する意義について議論し、これを実現するための枠組について述べる。

2. 履歴情報と評判情報の統合

2. 1. 統合する意義

消費者が特定の商品について詳しく知りたい場合、メーカーや団体の Web サイトを閲覧したり Web 上のブログ記事から他の消費者がどのように思っているかを参考にしたりする。情報を発信しているのは立場の違う生産者と第三者であるものの、どちらの情報も消費者にとって貴重な情報源である。そこで我々は、立場が異なるという点に注目し、もし両方の角度から見た情報を横断的に取得することが可能であれば、より効果的に商品の検索や比較をすることが可能であると考えた。そこで、我々は一般消費者が公開しているブログ上の評判情報をテキストマイニング技術を用いて分析する。そして、履歴情報と統合して商品検索システムのソースとして活用する。

この試みは、より広範囲でのトレーサビリティ—原料の生産から食した後の感想にいたるまでの情報共有—の実現を期待し提案するものである。本提案では、原材料の生産農家から加工業者さらに消費者までがそれぞれの立場で情報を発信できる (図1)。また、消費者はこれらの情報を横断し

ながら次のような検索が可能である。

2. 2. 消費者のメリット

まず、従来ではできなかった「小麦アレルギーを患っていても安心で、なおかつ人気のある商品は？」などのような履歴情報と評判情報を統合した検索ができる。また、検索結果を効率的に閲覧することができる。例えば、生産者が有機農産物あるいは低カロリーであることをアピールしている商品についての詳細な履歴情報と、「有機農産物だから買った」あるいは「ダイエットのために」などの第三者の意見を合わせて閲覧することができる。もし手作業で調べようとした場合、それぞれのページ (十以上のメーカーや業界、百以上のブログ記事) からそれぞれ情報を閲覧しなければならない。

さらに、両サイドの意見を比較することが可能である。商品特性によっては、生産者サイドは自社商品の印象や味について紹介し、消費者もブログ上に個人の主観による印象や味の評価をしている。例えばワインでは、味、香り、全体の印象について花や果実などにたとえたり、「洗練された」「高貴な」「力強い」などの表現をしたりする。これらを統合すると、両サイドの評価が共通する部分、あるいは共通しない部分に分かる。



図1. それぞれの立場で情報を発信

2. 3. 生産者のメリット

我々の提案は、高品質商品を生産する優良な生産者にも大きなメリットがある。なぜならば、本枠組は生産者に詳細な履歴情報と（ブログを書いた）消費者の感想を（本検索システムを使う）消費者に伝える機会を与える。積極的な情報公開は高く評価されるため、生産者は企業努力を積極的に発信すればよい。また価格以外の情報（例えば口コミ情報）の流通が促進されるため、他企業との差別化が図れる。

次章から本枠組を実現するための手法について述べる。

3. 履歴情報の発信

商品は農家のような一次生産者や加工業者などの複数の生産者によって作られている。我々は、これを商品のヒエラルキーと捉え、XML のリンクを繋ぐことでの履歴情報を表現することにした。本章では生産者による履歴情報の発信方法について述べる。

3. 1. データ化フォーマット

本節ではまず、履歴情報をデータ化するためのフォーマットについて述べる。本フォーマットは RSS 1.0 (RDF Site Summary) [4] を流用しており、これを RSSPI (RSS-based Product Information) と呼ぶ。生産者は RSSPI に従って履歴情報を記述し Web 上に発信する。

RSS1.0 は Web サイトのページタイトルとそのページがリンクするトピックスを記述するための XML ベースのフォーマットである。RSS ドキュメントは channel, image, item, 追加要素の4つの要素を持ち、item 要素は別の RSS ドキュメントにリンクすることができる。リンクされた RSS ドキュメントも自身の item 要素でさらに別の RSS ドキュメントにリンクすることができるので、ヒエラルキーを実現できる。そこでこの仕組みを商品のヒエラルキーに当てはめる。

図2にサンドイッチを例に挙げて説明する。サンドイッチをつくる加工業者はサンドイッチの RSSPI ドキュメントの channel 要素に詳しい情報を記述する。次に、item 要素でジャムやバナナなど材料の RSSPI ドキュメントにリンクを張る。

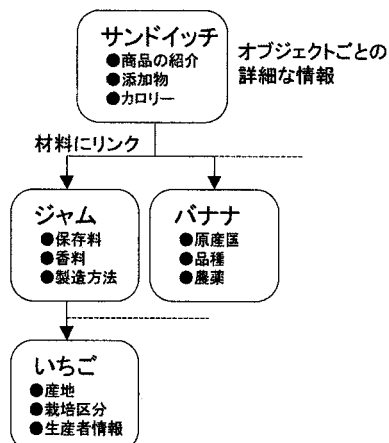


図2. サンドイッチのヒエラルキー

一つの RSSPI ドキュメントに全て記述するのではなく、個々の情報をオブジェクトとして扱うため、サンドイッチをつくる加工業者はジャムにリンクを張るだけで、ジャムの情報（ジャムをつくる農家が公開している詳細情報）を表現したことになる。

なお RSS はあくまでサイトサマリーを記述するためのフォーマットであるため、商品名や概要などを記述するための基本的なタグは使用できるが、生産履歴を詳しく記述するためのタグはもともと用意されていない。そこでモジュールを作成し名前空間を拡張して記述する。モジュールは XML Schema を用いて定義している。

3. 2. RSSPI ドキュメントの作成と発信

生産者は、RSSPI ライタを用いて RSS ドキュメントを作成する。生産者は、テキスト、チェックボックス、リストボックスなどの GUI を操作して履歴情報を記述する。もしくは HTML から RSSPI 形式に変換する。作成した RSSPI ドキュメントは自社の Web サーバに公開する。もし個人レベルで Web サーバを管理できない場合は、農協などのグループ単位でまとめて公開すればよい。そして、更新したことを知らせるために、5章で説明する Ping サーバに更新 Ping を送信する。

4. ブログ上の評判情報のマイニング

本章では、現在検討を進めているブログ上の評判情報のテキストマイニングについて、実現可能性を含めて述べる。

4. 1. マイニングのターゲット

我々はまず、消費者にとっての有用性が向上するためには、どのような商品のどのような情報をマイニングすればよいのかを検討することから始めた。なぜならば、商品特性によってブログに書かれやすい業界と書かれにくい業界があると考えられるためである。例えば、日常商品よりも特別な意味を持って食べた商品の方が、その時の状況や、誰と食べたか、どのように調理したか、味はどうだったか、など詳細に記述するであろう。また、商品特性によって出現頻度の高い用語は異なる（例えば「新鮮」は果物には使用されるがレトルト食品には使用されない）上に、仮に同じ用語が使用されていても意味合いや感覚が微妙に違う場合もある（例えば「辛い」はカレーではポジティブ、味噌汁ではネガティブ）。そこで我々は、まずは商品の対象をしばり精度の高いマイニング手法を実現してから、徐々に商品の対象を広げていくことにした。

我々は今回ワインを題材とした。なぜならば、他の商品に比べてブログの件数が多く、また味や香りの印象について表情豊かな表現をするためである。そして種類が多く「タンニン」などの専門用語も難しいため、初心者にとって他の消費者の生の声は貴重な情報源になると考えた。なお、マイナーなワインは少数ロットの輸入で売り切ってしまう在庫がないことが多い。従って、安定供給される大手の200種類程度のワインを対象とすることを検討している。

4. 2. マイニングの流れと得られる結果

図3にマイニングの大まかな流れを示す。

- ①Web クローラを用いて一般消費者によるブログ記事を収集する。
- ②感想などの評判を特徴抽出する。
- ③テキストマイニングにより、商品ごとの評判を分析する。

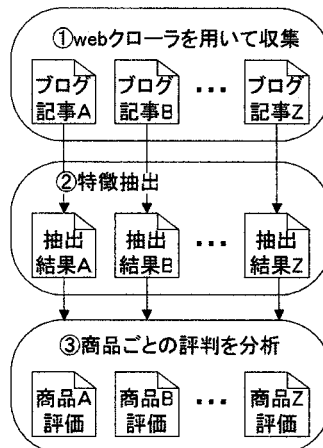


図3. マイニングの流れ

具体的な手法については現在検討中であるが、得られる結果として次に示すものを想定している。

- ・ 酸味・甘味・タンニン・果実味などの味の属性に関して、それぞれ何件書き込みがあったか。また、それぞれの「強い・弱い」の評価値の平均値
- ・ 香りの印象
- ・ 一緒に食べた料理
- ・ 注目度、話題度

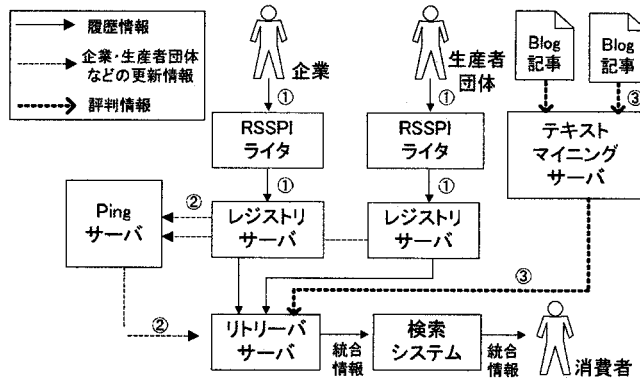


図4. 統合検索システムの構成図

5. 統合システムの構成

本章では、履歴情報と評判情報を統合するシステムの構成と流れについて述べる。図4に構成図を示す。本統合システムは、次の四種類のサーバで構成される。

- ・ レジストリサーバ
RSSPI ドキュメントを公開するためのサーバ（自社あるいは農協などのグループ単位で管理する Web サーバ）
- ・ Ping サーバ
更新情報を管理するためのサーバ
- ・ テキストマイニングサーバ
ブログ記事を収集し、商品別の評判を分析するサーバ
- ・ リトリバサーバ
履歴情報と評判情報を統合するサーバ

システムの流れは次の通りである。

- ① 生産者は RSSPI ライタを用いて生成した RSSPI ドキュメントをレジストリサーバに公開し ping サーバに更新 ping を送信する。
- ② Ping サーバは、更新情報を管理しリトリバサーバに報告する。Ping サーバは、RSSPI ドキュメントのやり取りを以下の点において効率的にする。一点目としてリポジトリ管理の簡易化が挙げられる。Ping サーバが常に最新の RSSPI ドキュメントの URL を取得しているため、リトリバサーバはレジストリサーバの URL を把握し管理する必要がない。

二点目としてサーバ間トラフィックの軽減が挙げられる。リトリバサーバはどの RSSPI が更新されているかを Ping サーバから受け取るため、レジストリサーバに定期的なアクセスする必要がない。

- ③ テキストマイニングサーバは、ブログ記事を収集し商品ごとの評価を分析する。そして分析結果をリトリバサーバに送信する。更新頻度は一日に一度程度を想定している。
- ④ リトリバサーバは、Ping サーバから受け取った更新情報をもとにレジストリサーバから RSSPI ドキュメントを収集する。そしてヒエラルキーを辿り商品別に履歴情報としてまとめる。また、テキストマイニングサーバから商品別の評判情報を受け取る。
- ⑤ 商標または登録商標に基づいて、各情報を統合しリレーショナルデータベースに登録する。リトリバサーバに登録された情報は、次章で述べる検索システムのソースとして活用される。

6. 検索システム

我々は、消費者が店頭で商品の検索や比較が行えるよう、モバイルデバイス上で動作する検索システムを開発している。モバイルデバイスを用いて買い物の手助けをする試みは他にもある。例えば、ショッピングメモから売り場をナビゲーションし詳細な情報を提供する試み [5] やマルチモーダルインタフェースとして操作でき、購入履歴を参考に情報を提供する試み [6] がある。しかし



図5. 検索システムの画面イメージ

ながら本検索システムは2. 2節で述べた消費者のメリットを実現するものであるため、買い物を手助けするための効果的な手法の一つであると考えている。

本検索システムは、例えば「原料のブドウが〇〇で、なおかつ酸味が強いと評判されているワインは？」のような履歴情報と評判情報を統合した検索ができる。さらに図5に示すように、味や香りの印象について、生産者サイドの紹介とブログ上の評判を比較することができる。

7. おわりに

本研究では、生産者による履歴情報と、一般消費者が公開しているブログ上の評判情報を統合し、商品検索システムのソースとして活用するための枠組を提案した。本枠組は、原材料の生産者から消費者までがそれぞれの立場で履歴情報を発信できること、消費者のより効果的な商品検索が実現できることを期待し提案したものである。

今後は、ブログ記事の信頼性を観点に含めたマイニング手法を検討していく。なぜならばブログ記事は必ずしも質が高いものばかりではない。商品について詳しい人とそうではない人の意見は信頼度が異なる。また一般消費者の本音とステルスマーケティング(ロコミによる宣伝効果を狙って、商品とは直接無関係な第三者を装いヤラセ情報の発信、伝播を図る行為)による記事は信頼度が異なる。関連研究として、即時性および重要性の観

点から信頼度の高いブログ記事を取得する試み[7]があるが、今後我々はこの手法を応用するなどして、良質のブログ記事を見分ける手法と集めたデータの分析手法を検討していく。

謝辞

ワインビジネスについて有益な助言をいただいた株式会社久本酒店の佐藤哲也社長に深く感謝いたします。

参考文献

- [1] OPEN ISHII, <http://open.ishiifood.co.jp/>
- [2] 青果ネットカタログ SEICA, <http://seica.info/>
- [3] 中畑将吾, 中野裕介, 佐川裕一, 垂水浩幸: RSSを用いた商品情報発信・収集システムの提案, 情報処理学会論文誌, Vol. 48, No.1, pp.98-109, 2007.
- [4] “RSS(RDF Site Summary)によるサイト情報の要約と公開”, <http://www.kanzaki.com/docs/sw/rss.html>.
- [5] P. Kourouthanassis and G. Roussos: Developing Consumer-Friendly Pervasive Retail Systems, IEEE Pervasive and Ubiquitous Computing, Vol.2, NO.2, pp.32-39, 2003.
- [6] Z. Davis, M. Hu, S. Prasad, M. Schuricht, P. M. Melliar-Smith and L. E. Moser: A Personal Handheld Multi-Modal Shopping Assistant, IEEE, International conference on Networking and Services (ICNS'06).
- [7] 中島伸介, 田中克己: 信頼度に基づくblog情報フィルタリング, 日本データベース学会 Letters Vol.3, NO.2.