

リアルタイム興味解析に基づく地域情報最適化フレームワークの提案 (132301008)

The real-time interest analyzing framework for regional information networks

研究代表者

川村 秀憲 北海道大学大学院情報科学研究科
Hidenori Kawamura Hokkaido University

研究分担者

鈴木 恵二[†] 平田 圭^{††}
Keiji Suzuki[†] Kei Hirata

[†]はこだて未来大学 ^{††}北海道大学

[†]Future University Hakodate ^{††}Hokkaido University

研究期間 平成 25 年度～平成 26 年度

概要

地域に密着した祭りやイベント、広告等の地域情報を興味解析技術を駆使した配信の最適化によって数万人規模のユーザに配信すると同時に、そこから収集される閲覧履歴、閲覧者の属性の動態調査を総合的に分析するアルゴリズムを開発し、魅力向上と課題改善を支援する情報配信分析システムの開発を行った。具体的には、amazon aws による情報配信基盤、イベント情報収集システム、ディープラーニングを利用した推薦アルゴリズム、情報配信スマートフォンアプリケーションの開発とサービスの社会実装を行った。

1. まえがき

本研究では、地域に密着した祭りやイベント、広告等の地域情報を興味解析技術を駆使した配信の最適化によって数万人規模のユーザに配信すると同時に、そこから収集される閲覧履歴、閲覧者の属性、口コミ、イベント参加者の動態調査を総合的に分析するアルゴリズムを開発し、魅力向上と課題改善を支援する情報配信分析システムを開発することを目的とする。

そして、ニーズ調査と改善施策のサイクル期間を最小化することでイベント開催者等が自律的にサービス改善を行い、事業強化が促進される環境を実社会に提供し、サービスの魅力が向上することで市民の外出を誘発し、地域振興に結びつけるのが最終的な狙いである。

2. 研究開発内容及び成果

この目的の実現へ向けて、本研究では大きく分けて次の3つの研究テーマに沿って研究を実施した。

(1) ビッグデータに対応したクラウド分散処理による情報配信と分析の実用化

(2) 情報の時間による価値変動を前提とした情報配信と分析手法の研究開発

(3) 地域の事業者や観光産業関係者と協力したサービス普及と地域産業への定着

以下、それぞれの詳細について述べる。

2. 1. ビッグデータに対応したクラウド分散処理による情報配信と分析の実用化

サービスへのアクセス数として10万～100万人規模、情報の購読ユーザ数で1万～10万人規模のユーザに研究成果を投入し、地域情報の流通を促して研究成果による地域活性化に貢献できるシステムを想定し、ビッグデータを処理できるシステムをamazon awsをベースに開発した。

具体的には、アクセスユーザ数に応じたスケーラブルなウェブサーバ、データ量に応じたスケーラブルなSQLサーバ、ユーザのデータを処理する際の計算環境分散化、それらの安定運用を支える技術を開発した。

さらに、ウェブ上の様々な公式サイトで更新されるイベント情報を収集するために、クローラと呼ばれる一連のソフトウェア群を開発した。クローラはサイトごとのHTML構成を解析して設計されており、イベント情報の更新があった際、イベントの日付やタイトル、内容について自動的にテキスト処理を施してイベント情報の元を抽出する。研究開発期間内において、およそ250のイベント情報サイトを個別に解析し、それぞれに対応した250個のクローラを実装した。

抽出されたイベント情報はそのままでは内容、口調、さまざまなデータ項目などが整形されていないので、収集された情報は人手によってイベント情報として利用可能な形に整形・修正する必要がある。そこで、クローラによって収集されたイベント情報を人手で整形するシステムについて開発を行った。

このシステムは、クローラが収集したイベント情報や情報元サイトを一つ一つ確認し、タイトル、日時を切り出しつつ適切な内容文に加工するためのシステムである。1件あたりの作業時間が1秒でも短くなるように、スクロール量、マウス移動量、クリック数が最少となるように、試行錯誤を繰り返しながらユーザインターフェースを最適化した。これによって、作業に慣れた入力作業者はおよそ1分程度で一つのイベント情報を処理できるようになった。

2. 2. 情報の時間による価値変動を前提とした情報配信と分析手法の研究開発

クローラと人手で処理された大規模イベント情報データベースを適切に処理する際、前処理としてテキストから重要な特徴量を抽出することができれば、ジャンル推定や情報推薦などの精度を高めることが可能となる。事前の研究では、テキストを形態素解析にかけて単語の出現頻度による文章ベクトルを作成し、TFIDFによってテキスト間の類似度を利用することを想定していたが、推薦や分類などに利用した際あまり高い精度が得られなかった。そこで、各イベント情報に対してディープラーニングと呼ばれる機械学習方法を適用することで教師なし学習による特徴量の自動抽出について研究を行った。

さらに、サービスに登録したユーザにイベント情報を届ける際、ユーザの過去の利用履歴によって数多くの候補から興味を示しそうなものをピックアップして届けるアルゴリズムの研究開発を行った。ここでは、下記の8つのフィルターアルゴリズムを開発した。

- ・ イベント情報をジャンル、情報源でセグメント化し、ユーザごとの反応を確率モデルで表現した「ジャンルフィルター」「情報源フィルター」
- ・ イベント情報の人気度合を確率的に表現した「人気フィルター」
- ・ 本研究で開発したディープラーニングによって得られる特徴量を利用し、閲覧イベント情報と類似性の高い情報を選別する「deep フィルター」
- ・ ユーザ群のイベント情報閲覧履歴を利用し、ピアソンの相関係数を利用してイベント情報の類似性を測り推薦に利用する「Item-CF フィルター」
- ・ 対応するイベントの「緯度経度」と過去の閲覧履歴からノンパラメトリック手法の一つであるカーネル密度推定を利用し、確率分布化することで推定される分布を利用した「緯度経度フィルター」
- ・ 各フィルターで得られるスコアリングの結果をユーザ全体で同じ加重和によって融合する「加重和フィルター」
- ・ 各フィルターで得られるスコアリングの結果をユーザ個別で加重和によって融合する「個別加重和フィルター」

各フィルターの計算方法についての詳細は省くが、これらのフィルターの中でも「deep フィルター」と「緯度経度フィルター」は本研究期間中に開発されたオリジナル性の高い独自の手法であり、イベント情報特有の特徴量や緯度経度情報を推薦の計算の中に組み込むことができ、高い性能を発揮することが期待できる新手法である。

2. 3. 地域の事業者や観光産業関係者と協力したサービス普及と地域産業への定着

アルゴリズムの開発、およびイベント情報収集の実証実験を経て、得られたデータを社会実装し、一般の人々に便利に利用してもらうこと、およびそれによるデータ収集をめざし、スマートフォンアプリケーションの開発、一般公開を行った。スマートフォンアプリの開発にあたって、Amazon aws 上にアプリケーションにデータを受け渡す API を作成した。アプリケーション本体は、iPhone、Android のウェブ view をベースとし、JavaScript、jQuery mobile によって開発を行った。このアプリケーションでは、ユーザが登録することで日々閲覧したイベント情報の履歴がサーバに蓄積され、そのデータをもとに新たなイベント情報のスコアリングが決定され、推薦順にイベント情報が表示される。図が開発したスマートフォンアプリケーションのスクリーンキャプチャである。

現在、iPhone、Android に対応しており、札幌版、横浜版、福岡版、名古屋版、神戸版を公開している状況である（公開サイト <http://bemall.jp/sapporo/app/banner>、<http://bemall.jp/yokohama/app/banner>）。それぞれ、月におよそ 1500 件程度のイベント情報が掲載されており、日々、利用状況のデータが収集されているので、実際の市民の方々の利用履歴に基づいてアルゴリズムの改善や有用なデータ提供の実証実験を開始した段階である。

3. 今後の研究開発成果の展開及び波及効果創出への取り組み



図 1 情報発信サービス（イメージ）

現在、北海道大学発ベンチャーである株式会社調和技研と協力し、実際にイベント情報配信サービスの社会実装を継続中である。情報の入力整形には障害者自立支援施設である「札幌チャレンジド」に作業を依頼してイベント情報の確保を行っている。2015 年 8 月のサービス利用者（ユニークユーザ）は 25 万人程度であり、多くの人々の外出を創出していると思われる。サービス自体はまだまだ試験運用の段階であるので、地域の広告代理店やイベント主催者、百貨店等と協力し、ビジネスとして自立できる段階を目指すとともに地域振興へ向けて研究成果の PR と開発されたサービスの普及に取り組んでいる。

4. むすび

本研究では、「知っていたら行ったのに」の社会的機会損失をなくし、地域の人々の生活の質向上、および地域のイベントや商業活動の活性化を達成するために、地域情報の配信最適化サービスの研究開発、および社会実装に取り組んできた。SCOPE によってスピーディに研究開発を行うことができ、社会実装への第一歩を踏み出した。

【誌上発表リスト】

[1] 小野 良太、山下 晃弘、川村 秀憲、鈴木 恵二、“イベント情報推薦のためのスコアリングの検討”、観光情報学会論文誌「観光と情報」 Vol. 11 No. 15 pp. 23-34 (2015/5)

[2] Ryota Ono, Kei Hirata, Hidenori Kawamura, Keiji Suzuki, “Scoring Algorithm for Event Notice Recommender System”, The 2nd International Conference on Serviceology (CServ2014) (2014/9)

[1] 瀬川 晋作、川村 秀憲、鈴木 恵二、“Deep Auto-Encoder によるイベント記事の分析”、第 9 回 SIG-DOCMAS 研究会 (2015/3)

【受賞リスト】

[1] 小野 良太、観光情報学会 研究発表会奨励賞、“地域イベント情報配信サイトにおける利用者の閲覧履歴を用いた情報推薦” (2014/6)

【報道掲載リスト】

[1] “個人の興味、自動解析”、北海道新聞、2014 年 2 月 15 日

【本研究開発課題を掲載したホームページ】

<http://bemall.jp>