

Web上のライフログの包括的利用に向けた ライフログリポジトリの構築

出口貴也* 奥野拓

(はこだて未来大)[†]

1 はじめに

ライフログとは、人間の行動をデジタルデータとして記録に残すことを指し、近年注目を集めている分野である。近年では特に、ブログや mixi (<http://mixi.jp/>)、Twitter (<http://twitter.com/>) に代表される、CGM (Consumer Generated Media) と呼ばれるユーザ発信型のサービスが Web 上で多数提供され、人気を集めている。こうしたサービスをユーザが利用する中で、ユーザが発信し、蓄積した情報を Web 上のライフログと呼ぶ。この Web 上のライフログが、本研究の対象である。

ライフログは自身の行動記録であるため、集約することで、よりパーソナライズしたシステムの構築が可能となる。例えば、ライフログからの過去の回想や、ライフログを使った情報の再アクセスへの支援、ライフログをベースとした推薦システムといった利用が可能である。このような利用法が、ライフログの包括的利用である。

Web 上に存在するライフログは、ユーザ自ら発信した情報であるために、ユーザの嗜好を多く含むという特徴がある。例えば、ユーザの音楽視聴履歴を扱う Last.fm (<http://www.lastfm.jp/>) からのライフログならば、ユーザが好む音楽嗜好が見えてくる。また、読書管理を行うブックログ (<http://booklog.jp/>) であれば、ユーザの好きな本や興味分野がわかる。このように、扱うコンテンツが専門化しているサービスが持つライフログを集約することで、ユーザの多様な嗜好が見える。

本研究では、嗜好情報も含めたライフログの包括的利用のベースとなる、ライフログリポジトリの構築を目的とする。そのためには、ライフログの収集、嗜好抽出、二次利用性を考慮した蓄積の三段階を経る必要がある。本稿では、この三段階の実現方法について提案する。

2 ライフログのコンテンツを考慮した集約

現在、ライフログの多くは、提供サービス内に閉じた形で利用されている。これにより、各サービスがライフログを内包し、ライフログは Web 上で分散している。そのため、それらを包括的に利用することは困難である。

この問題に対し、ライフログを集約するサービスが近年提供され始めている。代表例として、ライフログをアルバム形式で時系列に並べる Life-X (<http://life-x.jp/>) や、プロフィールとして集約する iddy (<http://iddy.jp/>) が挙げられる。前者は、ライフログのメタデータに含まれる作成日時を利用して並べているものであり、後者は

ライフログを一つのページにまとめ、一覧表示しているものである。つまり、これらのサービスでは、ライフログのメタデータを利用した集約を行っており、ライフログの内容 (コンテンツ) を考慮した集約は行われていない。

ユーザの嗜好情報を多く含むライフログの特性を活かすためには、メタデータのみを利用した集約では不十分である。コンテンツを解析し、抽出した嗜好情報をメタデータと併せて集約すべきである。

3 ライフログの二次利用性

1 で挙げた各サービスは Web API を公開しており、ユーザに紐づいたデータの収集を行うことができる。しかしながら、各 Web サービスの仕様が異なるため、得られるデータの内容や、構造には差異がある。よって、収集したライフログをそのまま蓄積するのでは、ライフログを統一的に扱うことはできないと考えられる。つまり、ライフログの包括的利用 (二次利用) が困難となる。ライフログの蓄積を行う際には、二次利用性を考慮し、データの差異を埋める処理などを行うべきである。こうすることで、各サービスの仕様を意識せずに、ライフログを扱うことができると考えられる。

ライフログを統一的に扱うために、中村ら [1]、下條ら [2] は、ライフログの標準データモデルを提案している。この提案では、ライフログが記録された時刻、記録した人、記録した場所、記録に用いたデバイスといった、各ライフログが共通に持っているであろうメタデータの標準化を行っている。この標準データモデルによって、ライフログのメタデータに関しては二次利用性を考慮した蓄積を行うことができると考えられる。しかしながら、ライフログのコンテンツを統一的に扱うことは、この標準データモデルでは困難である。

4 ライフログリポジトリの構築手法

4.1 ライフログの収集

ライフログの収集には、各サービスが公開している Web API を用いる。現在、多くのサービスが内包するデータを、JSON や XML といった形式で公開している。これを利用して、ユーザに紐づいたデータを収集する。なお、一般に Web API が公開されていない mixi のようなサービスに関しては、スクレイピングを行う必要がある。

4.2 ライフログからの嗜好抽出

2 で挙げた課題を踏まえて、本研究では収集したライフログに嗜好抽出処理を行う。Web 上のライフログは、

* b1007241@fun.ac.jp

[†] 函館市亀田中野町 116-2 公立はこだて未来大学

嗜好に関して二種類に分けられる。一つ目が、Last.fm やブックログのような、ユーザの嗜好そのものをコンテンツとして扱うサービスである。このようなサービスでは、その性質から、嗜好がライフログにも直接的に表れていることが多い。二つ目が、Twitter や mixi の日記といった、ユーザが記述する文章をコンテンツとして扱うサービスである。このようなサービスでは、嗜好だけでなくユーザに関わる様々な情報がライフログに含まれている。

このように Web 上のライフログを分類し、どのようなデータがユーザの嗜好になり得るのかについて考察した。嗜好が直接的に表れるライフログは、メタデータを利用することでユーザの嗜好が見える。代表例として、Last.fm であれば、ユーザの好むアーティストがわかり、またそのアーティストのメタデータを辿ることで、好きな音楽ジャンルがわかる。YouTube であれば、ユーザが高評価している動画リストを取得でき、その動画のメタデータからユーザの好みがわかる。

一方で、嗜好の他にも様々な情報を含むライフログは、コンテンツそのものを解析しなければ嗜好が抽出できない。よって、形態素解析等のテキストマイニング処理を行う必要がある。特に、mixi のプロフィールは、SNS の性質上詳細にユーザの好みについて記述されていることが多く、嗜好抽出の際には重要なソースになると考えられる。Twitter のつぶやきも、ブログ記事に比べて更新量が多く、端的に好みについて語られている場合があるため、多くの嗜好が凝縮されていると考えられる。

以上のように、ライフログには、メタデータのみで嗜好抽出できるものと、コンテンツの解析から嗜好が抽出できるものが存在する。コンテンツからの嗜好抽出の具体的な方法については、今後検討していく。

4.3 二次利用性を考慮した蓄積

3 で挙げた課題を踏まえ、収集したライフログと抽出した嗜好情報を、二次利用性を考慮して蓄積する必要がある。本研究では、3 で挙げた中村ら [1]、下條ら [2] の標準データモデルを拡張する形で蓄積する。

1 で挙げたライフログの包括的利用法の中でも、推薦システムは、ユーザの嗜好情報が重要となる利用法である。一般に推薦システムでは、ユーザの嗜好情報を獲得し、嗜好のモデル化を行い、ユーザプロフィールを作成する [3]。ライフログをベースとした推薦システムの構築を行う際には、ライフログから抽出された嗜好情報を用いてユーザプロフィールの作成を行う。

プロフィールの作成のためには、関心や好みの度合いを数値化する必要がある [3]。そのためには、クラスタリングや重み付けといった処理が必要となる。その過程では、異なるライフログから抽出した嗜好情報であっても、同じ嗜好であれば、同一のものとして扱う必要がある。

そこで本研究では、ある嗜好キーワードを抽出した後、その広義語をシソーラス辞書等を用いて取得し、データ

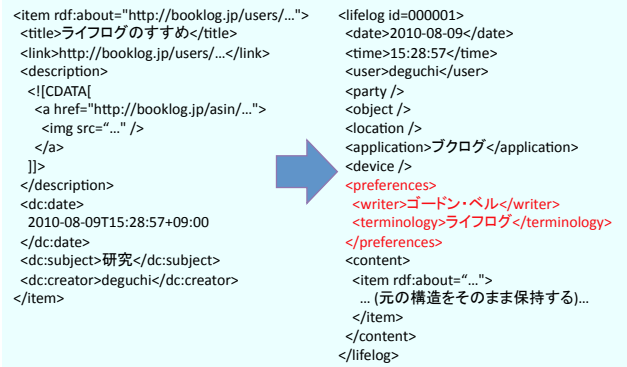


Fig. 1 ライフログの蓄積例

モデルに埋め込む。例えば、ブックログからユーザがある作家 A を好むと分かった場合は、<作家>作家 A</作家> の形で蓄積する。また、Twitter から、ある作家 B を好むと分かった場合は、<作家>作家 B</作家> の形で蓄積する。こうすることで、プロファイリングを行う際には、作家タグの重み付けを行い、ユーザがどの作家をより好んでいるかを数値化することが可能となる。

Fig. 1 に、ブックログから取得したデータを、本研究の蓄積方式に変換した例を示す。この例では、ブックログのメタデータを標準データモデルに沿って変換している。そして、ブックログの読書データより抽出した嗜好情報を、<preferences> タグ内に広義語でタグ付けして埋め込んでいる。このようにして、収集したライフログと抽出した嗜好情報を蓄積することで、二次利用が容易になると期待できる。

5 まとめ

本稿では、ライフログの包括的な利用を行う際の問題点の議論と、ライフログリポジトリ構築手法の提案を行った。今後は、コンテンツからの嗜好抽出手法や嗜好の分類手法の調査、本手法の実装を行う。また、有効性を評価するために、ライフログリポジトリを用いた推薦システムの構築を行うことを計画している。

参考文献

- [1] 中村匡秀, 下條彰, 井垣宏: 異なるライフログを集約するための標準データモデルの考察, 電子情報通信学会技術研究報告, 第 109 巻, pp.35-40, (2009).
- [2] 下條彰, 福田将之, 井垣宏, 中村匡秀: 異なるライフログをマッシュアップするためのデータ変換・集約アクセス API の実装, 電子情報通信学会技術研究報告, 第 109 巻, pp.85-90, (2010).
- [3] 土方嘉徳: 嗜好抽出と情報推薦技術, 情報処理, vol.48, no.9, pp.957-965, (2007).
- [4] 相澤清晴: ライフログの実践的活用: 食事ログからの展望, 情報処理, vol.50, no.7, pp.592-597, (2009).