RO-011

Web 上のライフログの統合利用に向けたライフログリポジトリの構築
Build the Life-log Repository for Integrated Use of Life-logs on the Web

出口 貴也† 奥野 拓‡
Takaya Deguchi Taku Okuno

1. はじめに

近年，Web 上では CGM やソーシャルメディアと呼ばれる，ユーザ発信型のサービスが人気を集めている．代表例として，ソーシャルネットワーキングサービスの mixi (http://mixi.jp/) や facebook (http://www.facebook.com/)，マイクロブログサービスの Twitter (http://twitter.com/)，ソーシャルブックマークサービスのはてなブックマーク (http://b.hatena.ne.jp/) などが挙げられる．これらのサービスを利用することで，ユーザは Web 上に自身のライフログを残すこととなる．この Web 上のライフログの有効活用方法が近年注目を集めている[1].

他方で，Web 上では情報が爆発的に増加しており，ユーザが Web から有益な情報を見つけ出すことは困難になってきている．このような状況に対して，情報推薦や情報フィルタリングが，Web 上で広く利用されるようになった．代表例として，Google のパーソナライズド検索や，Amazon.com (http://www.amazon.com/) の商品推薦が挙げられる．これらは，ユーザの行動履歴を何らかの手段で取得し，サービスそのものを各々のユーザに特化させている．近年では，このようなパーソナライズされたサービスが多数提供されている．

本研究では Web 上に蓄積されたユーザのライフログを，情報推薦手法に適用し，サービスのパーソナライズを行うことを目指す．Web 上の様々なサービスに散在するユーザのライフログを集約・蓄積したものを，本研究ではライフログリポジトリと呼ぶ．

ライフログリポジトリの活用方法としては，過去の振り返り，情報の再アクセス (リファインディング) 支援，情報推薦などが考えられる．このような，ライフログを集約することで成り立つ利用法を，本研究ではライフログの統合利用と呼ぶ．

本研究で構築するライフログリポジトリは，このような様々な統合利用の基盤となるものであるが，特に情報推薦への応用に主眼を置いている．本稿では，情報推薦を目的としたライフログリポジトリの要件について検討し，構築手法を提案する．

2. ライフログリポジトリの必要性

2.1 ライフログの分散

現在ユーザは CGM やソーシャルメディアを通して，Web 上で自身の考えや，日常生活，趣味嗜好に関わる情報を発信している．そして，それがライフログとして蓄積されている．図 1 のように，ユーザの発言，撮影した動画や写真，音楽視聴履歴，読書履歴，位置情報，プロフィール

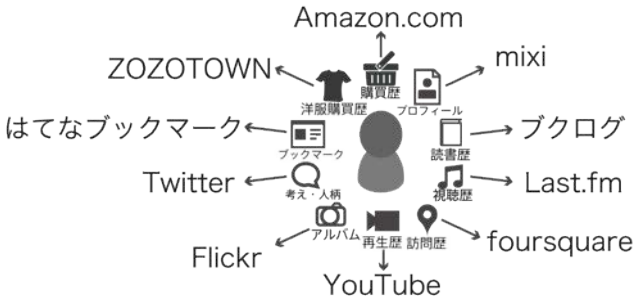

図 1 ライフログの分散

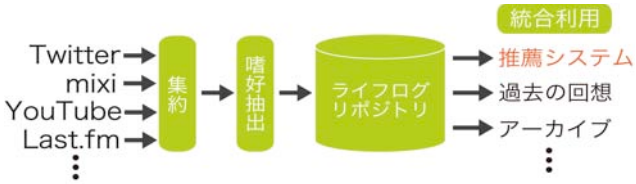

図 2 ライフログリポジトリ

情報などは，サービスそれぞれが個別に保有している．そのため，提供サービス内で閉じた形でのみ利用されている．したがって，ユーザ自身のライフログが Web 上に分散してしまっている状況にある．それにより，ライフログの統合利用を行うのは難しい現状にある．

2.2 限定的な嗜好に基づく情報推薦

1 で挙げたようなサービスにおける情報推薦の多くは，サービス上で獲得したユーザのライフログに基づき，ユーザの嗜好を予測し，情報の推薦を行っている．例えば，ショッピングサイトであれば，ショッピングサイト内での購入履歴や，商品閲覧履歴を元に，推薦する商品を決定している．しかしこの場合，ユーザがショッピングサイト上で購入・閲覧しないジャンルの商品に関しては，たとえユーザが好む商品であったとしても推薦することができない．現在，図 1 のように多くのサービスは扱うコンテンツが専門化しており，それぞれのサービスは，ユーザの嗜好の一面のみを扱っている．したがって，サービス上で情報推薦を行う場合には，ユーザの多様な嗜好のうちの一部を対象とした情報推薦しか実現できていない．これは，ショッピングサイトのような，サイト内の商品という限定的な情報を推薦する場合には大きな問題にはならないと考えられる．しかしながら，Web ページ推薦のような，ユーザに対して網羅的な情報を推薦する場合には問題になる．

2.3 リポジトリ構築による問題解決

2.1 および 2.2 で述べた問題を解決するためのアプローチとして，図 2 のように，ユーザが Web 上で残すライフログを，複数のサービスから集約する．また，情報推薦への応用を考慮し，ライフログから嗜好の抽出を行う．さら

† 公立はこだて未来大学大学院, Graduate School of Future University Hakodate
‡ 公立はこだて未来大学, Future University Hakodate

表 1 ライフログが蓄積される代表的なサービス

	サービス名	概要	ライフログ
ユーザの嗜好がコンテンツ	delicious, はてなブックマーク	ソーシャルブックマークサービス．ユーザはブックマーク，タグ，コメントを共有．	ブックマーク，コメント
	YouTube	ユーザは動画をアップロード，閲覧，評価．	アップロードした動画，高評価リスト，お気に入り，再生リスト，再生履歴
	Last.fm	音楽視聴履歴を元にした SNS．常駐ソフトウェアがユーザの iTunes 等の再生履歴をアップロード．	視聴履歴
	Flickr	ユーザは，写真をアップロード，共有，管理．	アップロードした写真
	ブクログ	仮想本棚作成サービス．ユーザは自身の読書歴や蔵書を管理．	仮想本棚に登録した書籍，ゲーム，DVD 等
文章がコンテンツ	facebook, mixi	SNS．ユーザはプロフィールや日記の記述，Web ページを共有．	個人が記述する日記，コメント，つぶやき，プロフィール等
	Twitter	マイクロブログサービス．ユーザは 140 文字以内の短文を投稿．	つぶやき，プロフィール等
	アメーバブログ，はてなダイアリー	一般的なブログサービス．文字数等の制限は無く，ユーザは自由に記事を投稿．	ブログ記事

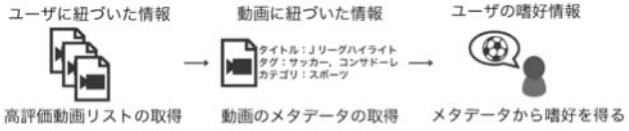

図 3 ライフログのメタデータによる嗜好抽出 (YouTube)

に二次利用性を考慮し，標準化された形式でリポジトリ上にライフログを蓄積する．

このようにして，Web 上の個人のあらゆるライフログと，そこからの嗜好情報を併せて蓄積したものがライフログリポジトリである．つまり，リポジトリをベースとした情報推薦システムを構築することにより，長期間に渡って蓄積されたユーザの多様な嗜好を扱うことができる．それにより，2.1 および 2.2 で示した問題が解決可能であると考えられる．

3. ライフログリポジトリ構築手法

2.3 で述べたような，統合利用の基盤となるライフログリポジトリを構築する上で，満たすべき要件の検討・考察を行った．特に，情報推薦への適用に主眼をおき，ライフログの収集，ライフログからの嗜好抽出，嗜好分類，二次利用性を考慮した蓄積の四段階を経て構築を行う．以降，それぞれの段階について詳細に述べる．

3.1 ライフログの収集

ライフログには，実世界で蓄積するログと，前述した Web 上で蓄積するログがあるが，本研究では後者に限定する．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，多くのサービスは，JSON や XML といった標準化された形式でサービス内のデータを公開している．これを利用することで，ユーザに紐づいたデータを収集することができ，これがユーザのライフログとなる．しかしながら，サイト上で公開されているデータでありながら，Web API が提供されていない場合は，対象サービスの HTML を解析し，必要なデータを抽出する処理 (スクレイピング) を行う．

3.2 嗜好の抽出

3.2.1 サービス毎のライフログの性質

どのようなライフログが蓄積されるかという観点で，代表的なサービスの性質について表 1 にまとめた．各サービスを，コンテンツの性質という観点で二種類に分けた．一つ目が，はてなブックマークや，YouTube，Last.fm といった，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直接的に表れていることが多い．二つ目が，facebook や mixi，Twitter といった，ユーザが記述する文章をコンテンツとするサービスである．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．

サービスのコンテンツには，そのメタデータとして，タイトル，カテゴリ情報，タグ情報，コメント，日時情報等が存在する場合が多い．前者のサービスの特徴は，このメタデータが，サービスのコンテンツを説明する情報となっている点である．前者のサービスは，表 1 のように，扱うコンテンツが，動画や写真といったマルチメディアであり，テキストではないことが多い．したがって，コンテンツの管理や，識別のためにメタデータが充実している場合が多い．また，コンテンツを説明するデータであるメタデータ自体が，ユーザの嗜好となり得ると考えられる．一方，後者のサービスは，コンテンツに嗜好の他にも雑多な情報を多く含むため，メタデータが付与されていても，コンテンツを十分に説明できていないことが多い．また，マイクロブログのような短い文章を扱うコンテンツには，タイトルやタグ，カテゴリといった管理のためのメタデータが付与されていないことが多い．したがって，このようなサービスが持つライフログから嗜好を抽出するためには，コンテンツのテキストマイニング処理を行う必要がある．

以上のように，ライフログは，メタデータのみで嗜好抽出できるものと，嗜好を抽出するためにコンテンツの解析が必要となるものに分けられる．本稿では，対象を前者のライフログに絞っている．

3.2.2 タグ情報を用いた嗜好抽出

本研究では，メタデータの中でも特にタグ情報に着目する．タグ情報はコンテンツを説明する情報であり，一般に単語で表される．また，タグ情報は，コンテンツ提供者が付与するものと，ユーザ自身が付与するものがある．一般にタグは複数付けられていることが多く，カテゴリやタイトルといった情報よりも，情報量は多いと考えられる．本研究では，このタグ情報をユーザの嗜好とみなし，タグの抽出と分類を行う．

タグ情報を用いた嗜好抽出の概要を図 3 に示す．例えば，YouTube であれば，ユーザに紐づいた情報として，高評価動画リストを取得できる．そして，そのリスト内の各動画にはメタデータとして，タイトルやタグ，カテゴリといった情報が付与されている．特に，タグには，その動画の内容や，出演者，ジャンルを表すキーワードが設定されていることが多い．したがって，それらからユーザが高評価を

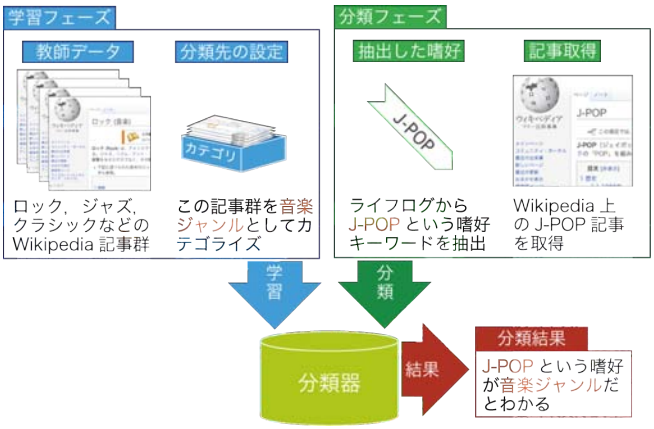

図 4 文書分類技術を利用した嗜好分類

与えている動画の特徴を推測でき，ユーザの嗜好に繋がる．また，Last.fm であれば，ユーザの曲の視聴履歴が取得できる．そして，その楽曲データに付与されているタグから，ユーザの好きな音楽ジャンルを判断できる．

3.3 嗜好の分類

3.3.1 情報推薦の過程

一般に推薦システムでは，嗜好データの獲得，嗜好の予測，推薦の提示の三つの段階を経て推薦を行う．特に，推薦データ獲得の段階では，嗜好のモデル化を行い，ユーザプロファイルを作成する[2]．ライフログに基づく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ァイルの作成を行う．プロファイ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．

本研究では，ライフログから嗜好を獲得し，プロファイリングを行う．一般の推薦システムで扱う嗜好は，2.2 で示したように限定的 (嗜好のジャンルが多様でない) である一方，ライフログリポジトリを用いる場合は，扱う嗜好が多様となる．ショッピングサイトの商品推薦を例にすると，扱う嗜好はショッピングサイト上の嗜好に限定されており，商品に関する嗜好であることが明確である．一方，ユーザに関する多様な嗜好が蓄積されたライフログリポジトリにおいては，嗜好が限定されていないため，何に関する嗜好なのかが不明確である．

したがって，例えばライフログからジャズとロックという嗜好を抽出した場合は，それが音楽嗜好であると明確にした状態で蓄積しておく必要がある．3.4.3 にて詳しく述べるが，本研究では蓄積の際に，このような嗜好に関するメタデータを XML 形式でタグ付けして表現する．先の例では，音楽が要素名であり，その内容がジャズとなる (<music>Jazz</music>)．ライフログリポジトリを用いて有効な推薦を行うためには，抽出した嗜好を，より適切なカテゴリに分類する方法を検討する必要がある．

3.3.2 ナイーブベイズ分類器による嗜好分類

本研究では，嗜好分類のために，ナイーブベイズ分類器による文書分類[4]を行う．提案の概要を図 4 に示す．学習フェーズにおいて，本研究では Wikipedia 上の記事群を教師データとして分類器を作成する．例えば，ジャズやロックといった Wikipedia 記事を用意し，これらが音楽ジャンルカテゴリの記事であるということを学習させる．

・Twitter

```
<status>
  <created_at>Wed Sep 22 19:20:26 +0000 2010</created_at>
  <id>2519980103</id>
  <text> 五稜郭でラーメンなう </text>
  <source>web</source>
  <user>
    <name>deguchi</name>
  </user>
</status>
```

・ブクログ

```
<item rdf:about=" http://booklog.jp/users/..." >
  <title> ライフログのすすめ </title>
  <dc:date>
    2010-08-09T15:28:57+09:00
  </dc:date>
  <dc:subject> 研究 </dc:subject>
  <dc:creator>deguchi</dc:creator>
</item>
```

図 5 ライフログのデータ構造の違い

次に分類フェーズに移る．本研究では，抽出した単語を嗜好として扱う (以降，嗜好キーワードと呼ぶ)．しかし，単語であるため，文書分類を行うには情報が不足している．そこで，本研究では嗜好キーワードを表題とする Wikipedia 記事でそれを補う．

図 4 の例では，抽出した嗜好キーワードが J-POP であった場合に，Wikipedia 上の J-POP 記事を取得し，その記事を分類器にかける．

この例では，Wikipedia 上の J-POP 記事の分類結果が音楽ジャンルであれば，J–POP という嗜好自体が音楽ジャンルだとわかり，これが嗜好の分類となる．以上のようにして， Wikipedia 記事の分類を行うことにより，嗜好キーワードの自動分類を行う．

3.4 二次利用性を考慮した蓄積

3.4.1 ライフログのデータ構造

本研究では，主に Web API を用いてデータの収集を行う．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．図 5 は，実際に Twitter とブクログ (http://booklog.jp/) から Web API を用いて取得したデータの一部である．この図では，Twitter はデータ生成日時が世界標準時で表されているが，ブクログは日本標準時で表されている．このように，同じ概念であっても，データ形式に相違がある場合がある．したがって，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．ライフログの蓄積を行う際には，二次利用性を考慮し，標準化処理を行うべきである．それによって，各サービスの仕様を意識せずに，ライフログを統合利用 (二次利用) できる．

3.4.2 ライフログのメタデータの標準化

ライフログを統一的に扱うために，中村ら[5]，Shimojo ら[6]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．

3.4.3 嗜好情報を併せた標準化

3.4.1 で示したとおり，収集したライフログは，二次利用性を考慮し標準化処理を行う必要がある．本研究では，ライフログのほかに抽出した嗜好情報も蓄積する．したがって，ライフログのメタデータを標準化するとともに，嗜好情報を統一フォーマットにより蓄積する必要がある．

本研究では，メタデータの標準化に関しては，3.4.2 で挙げた先行研究によって提案されている標準データモデル

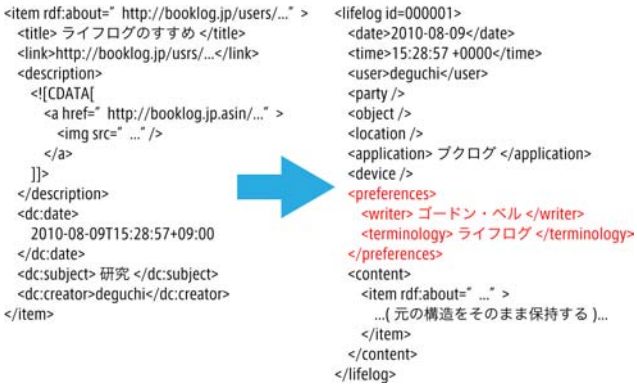

図 6 嗜好と併せたライフログの標準化

図 7 嗜好キーワードへの評価画面

を踏襲する．そして，抽出した嗜好情報を記述するために，新たに preferences タグを導入する．具体例としてブクログのライフログを提案する形式に変換した例を図 6 に示す．図の右側が，二次利用性を考慮した変換結果である．分類結果（カテゴリ名）を preferences タグ内の要素タグとして記述し，分類対象であった嗜好キーワードを，その内容として記述する．図では，ゴードン・ベルという嗜好キーワードが，writer という，その上位概念によってタグ付けされている．

また，この例では，ブ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．ライフログの作成日時が日本標準時で表現されていた場合は，世界標準時に変換するなど，データ型の統一も含まれる．

以上のように，収集したライフログと抽出した嗜好情報を標準的な形式で蓄積することで，統合利用が容易になると期待できる．

4. 実験と評価

提案手法のうち，嗜好抽出と嗜好分類に関する実験を行った．4.1 では，ライフログから抽出した嗜好が，ユーザの嗜好に合致しているかを評価することを目的としている．4.2 では，作成した分類器が嗜好分類として妥当であるかの評価を目的としている．

4.1 抽出結果の妥当性

4.1.1 実験概要

抽出した嗜好がユーザの嗜好と合致しているかを評価するため，被験者 6 人を対象とした実験を行った．対象としたサービスは，YouTube とはてなブックマークの 2 つである．被験者全員に，YouTube を最低 1 週間以上使用してもらった．また，はてなブックマークには被験者が常用するブックマークをインポートし，タグ付けをしてもらった．いずれのサービスも，使い方の方針などは指示せず，被験者の意思で自由に使用してもらった．

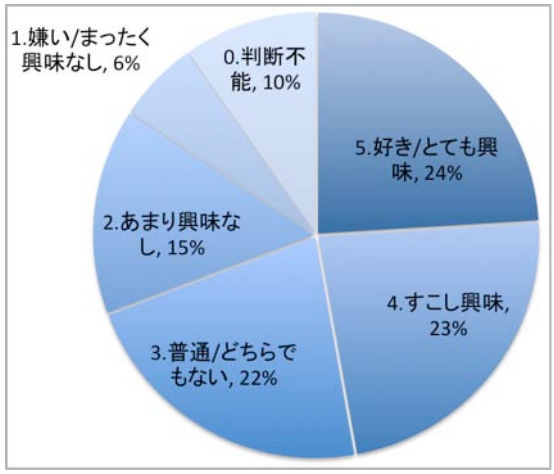

図 8 抽出した全嗜好に対する被験者からの評価の割合

このようにして，被験者の YouTube の再生履歴と，はてなブックマークのタグ付け済みブックマークを収集し，3.2.2 の手法を用いて嗜好抽出処理を行った．ここで，抽出した嗜好キーワードが，被験者の嗜好と合致しているかを評価するために，被験者にアンケートを行った．図 7 は，実際に用いたアンケート画面の一部である．抽出した嗜好キーワードを列挙し，それぞれの嗜好キーワードについて嗜好の度合いを，「5.好き/とても興味あり」，「4.すこし興味あり」，「3.普通/どちらでもない」，「2.あまり興味なし」，「1.嫌い/まったく興味なし」の 5 段階で評価をしてもらった．また，嗜好キーワードが何を意味しているかが判らず，嗜好の度合いが判断できない場合は，「0.判断不能」を選択してもらった．

4.1.2 結果と考察

図 8 は，抽出した全嗜好キーワードに対する被験者の評価を集計したものである．被験者が肯定的評価 (5, 4) を与えた嗜好キーワードは，全嗜好キーワードのうち 47%であった．反対に，被験者が否定的評価 (2, 1) を与えた嗜好キーワードは，全体の 21%であった．このことから，抽出した嗜好キーワードの半数近くは，ユーザの嗜好として妥当であったと言える．この結果は，タグ情報だけでなく，カテゴリやタイトルといった他のメタデータを併用することで，改善できると考えられる．

次に，嗜好キーワードを，抽出元のサービス別に分けて考察を行う．図 9 は，被験者が与えた評価毎に，それぞれのサービスが占める割合を示したものである．この図から，「5.好き/とても興味」の評価が与えられた嗜好キーワードの，73%は YouTube のライフログから抽出したものであり，35%ははてなブックマークのライフログから抽出したものだと分かる．はてなブックマークに関する各評価の割合を比べると，否定的評価 (2, 1) よりも肯定的評価 (5, 4) が，YouTube に対して高い割合を占めている．

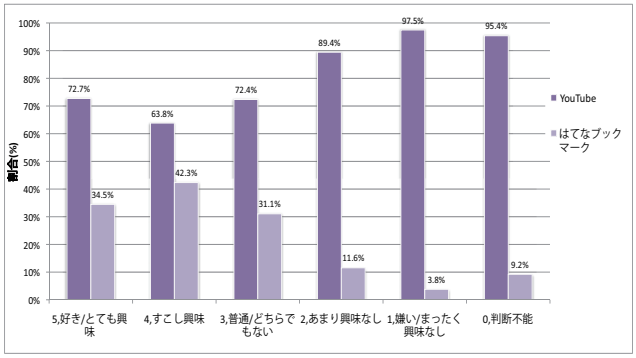

図 9 各評価におけるサービス別の割合

これは，YouTube のタグ情報はコンテンツ提供者が，他人に向けて付与したものであるのに対し，はてなブックマークのタグ情報は，サービス利用者が自分自身のために付与したものであることが影響していると考えられる．

コンテンツへのタグ付けの意図・目的や，ひとつのコンテンツに付けるタグの数はサービスによって変わる．例えば，YouTube では，動画をより多くの人に見てもらうために，大量のタグ付けが行われる．そのため，動画との関連性が低いタグまでもが付与され，ユーザの嗜好と合致しない場合が多くなる．一方，はてなブックマークでは，ブックマークを管理するために，カテゴライズ目的でのタグ付けを行われる．そのため，ユーザの嗜好に合致し易いものの，そのタグの数は少数となる．

このように，サービスの性質に影響を受けるため，サービス毎にタグ情報の扱いや重み付けを変える必要がある．

4.2 嗜好分類の妥当性

4.2.1 実験概要

実験では，映画，スポーツ，音楽，本，テレビ番組，ゲーム，食べ物，IT の 8 カテゴリに対して分類を行った．この 8 カテゴリは，抽出した嗜好を網羅的に分類するという粒度で設定した。まず，この 8 カテゴリに，教師データとして各カテゴリに含まれる 4,500 件の Wikipedia 記事 (合計 36,000 件) を学習させた．ここで用いた Wikipedia 記事群を，教師データセットとする．

次に，教師データセットとは別に，テストデータセットを準備する．これは，教師データセットとは全く別のデータ群であり，重複する記事は含まれない．テストデータセットの記事数は，教師データセットの 10%とした[7]．したがって，8 カテゴリの各々に含まれる 450 件の Wikipedia 記事 (合計で 3,600 件) を，テストデータセットとして準備した．

以上のように，教師データセットを用いて学習させた分類器を用いて，準備したテストデータセット 3,600 件の分類を行った．

なお，本研究で用いている，ナイーブベイズ分類器は確率論に基づくアルゴリズムである．したがって，分類の際には各々のカテゴリの中から，最も確率が高いカテゴリを選び分類を行う．つまり，適当なカテゴリが無いにもかかわらず，他のカテゴリよりは確率が高いという判断から，いずれかのカテゴリに分類されてしまう．そこで，本研究では，分類結果のスコアがある一定の閾値以下であるなら

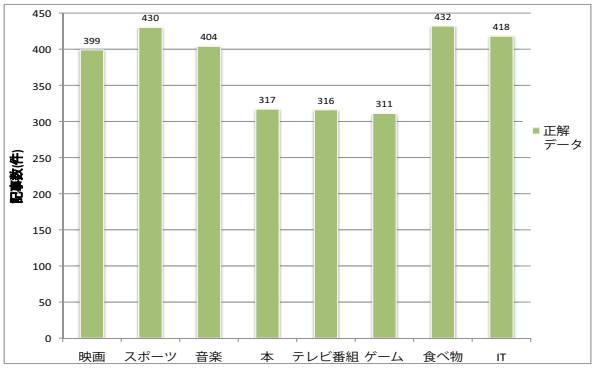

図 10 各カテゴリにおける正解データの数

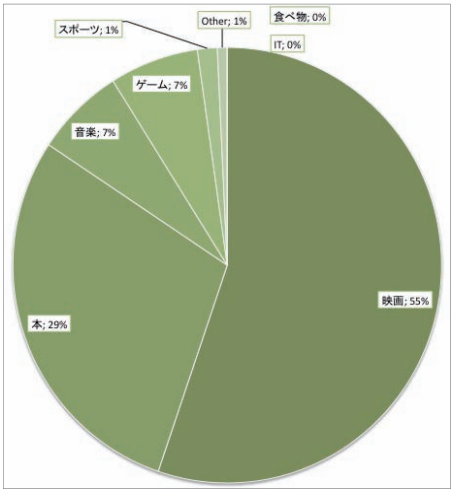

図 11 テレビ番組カテゴリの誤分類の割合

ば，分類結果の信頼度が低いとみなし，Other カテゴリに分類している．

4.2.2 結果と考察

テストデータセットの分類を行い，正解データの数を集計した結果が，図 10 である．この図は，例えば，映画カテゴリのテストデータセット内の Wikipedia 記事の分類結果が，映画カテゴリに分類された場合は，それを正解データとしてプロットしてある．つまり，値が大きいほど分類が精確であることを表す．この図より，映画，スポーツ，音楽，食べ物，IT カテゴリに関しては，テストデータセットの 9 割近くを正しく分類できているとわかる．よって，この 5 カテゴリについては，十分な精度で分類できているといえる．

しかしながら，本，テレビ番組，ゲームのカテゴリに関しては，正しく分類できたのは全体の 7 割弱程度である．この原因を分析するために，この 3 カテゴリについて，分類に失敗した Wikipedia 記事が，どのカテゴリに誤分類されているかを調べた．図 11 は，3 カテゴリのうち，テレビ番組カテゴリにおける誤分類の割合である．

図 11 より，Other カテゴリへの分類された割合は低いことがわかる．つまり，誤分類でありながら，分類結果としては信頼性のある分類ができていたケースが多いということになる．例えば，テレビ番組カテゴリ用のテストデータセットに含まれていながら，テレビ番組カテゴリ以外のカテゴリに分類されたケースのうち，約 5 割が映画カテゴ

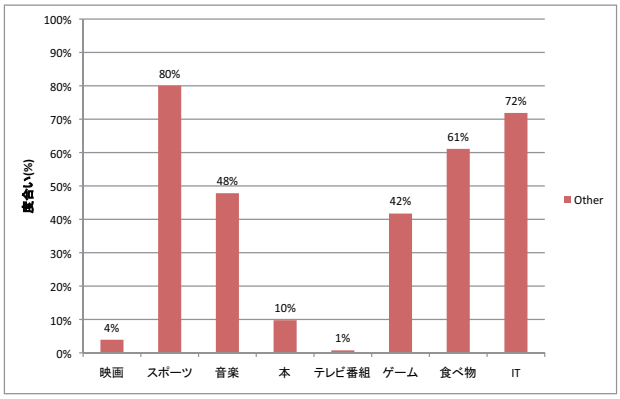

図 12 各カテゴリにおける Other に分類されたデータ数

リに分類されている．このことから，この映画カテゴリに誤分類された Wikipedia 記事は，正解データとしてはテレビ番組カテゴリであるが，映画カテゴリにも属する．例えば，映画化されたドラマなどはこれに該当する．

以上の結果より，分類の良し悪しを決めるのは，本研究に置いては，正解データ数ではなく，Other カテゴリに分類された数であると考えられる．分類結果の信頼度が低いカテゴリである Other に分類されるケースを減らすことにより，嗜好を漏れ無く，その性質に合ったカテゴリへと分類することが可能と考えられる．

そこで，各カテゴリにおける Other に分類されたデータの数を集計した，それをプロットしたものが図 12 である．この図より，スポーツや IT といった，正解データ数が多かったカテゴリであっても，Other に分類されるデータ数が多いことが分かる．実際に，スポーツカテゴリにおいて Other カテゴリに分類された Wikipedia 記事を調べると，「軟式野球」や「ロードレース」といった競技名に関する記事であった．このような問題は，教師データの少なさから起こる．今回の実験では，スポーツカテゴリとして，競技名，スポーツ選手名，チーム名に該当する Wikipedia 記事を学習させた．しかしながら，Wikipedia 上に存在する記事は，スポーツ選手名やチーム名に関わる記事が，競技名に関わる記事よりも圧倒的に多い．そのため，学習の際にも，選手名やチーム名の記事を，競技名の記事よりも多く用いたため，このようなことが起きた．この問題に対しては，カテゴリをより細分化したり，各カテゴリの教師データ数を均等にしたりすることが必要だと考えられる．

5. 結論

本研究は，ライフログリポジトリを構築することを目的としている．このリポジトリには，Web 上のライフログと，そこから抽出した嗜好情報を併せて蓄積する．また，このリポジトリを情報推薦へ応用することを想定し，それに伴う要件の検討を行った．そして，次の 4 段階からなる構築手法の提案を行った．

- Web API とスクレイピングによるライフログの収集
- タグ情報を利用した，ライフログからの嗜好抽出
- Wikipedia を用いた教師あり学習による，嗜好分類
- 二次利用性を考慮したライフログの標準化

これらのうち，嗜好の抽出と分類について，実験により評価を行った．抽出結果の妥当性についての評価では，抽出結果の半数近くは，ユーザの嗜好として妥当であった．また，分類の精度の評価については，ほとんどのカテゴリにおいて十分な精度で分類を行えることが確認できた．また，精度が低かったカテゴリについては，その原因の分析を行った．そこから，今後，嗜好をその性質に合ったカテゴリへと漏れ無く分類を行うための対策について検討した．

今後は，タグ情報だけでなく他のメタデータを活用し嗜好抽出手法を改善する．また，ライフログの性質によって重み付けを行う．この際には，今回の実験で対象にした YouTube，はてなブックマーク以外のサービスへの適用可能性を検討する必要がある．

このようにして，手法の改善を行った後には，より長期間に渡って本システムを用いた実験を行い，嗜好の抽出と分類を合わせた評価を行う．また，ライフログの統合による効果という観点からも評価を行う必要がある．

以上のような改善を行った後には，ライフログリポジトリを基にしたレコメンドサービスの開発，評価を行う．

参考文献

[1] 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).
[2] 神嶌敏弘，“推薦システムのアルゴリズム（１）”，人工知能学会誌，vol.22，no.6，pp.826-837，(2007).
[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] Toby Segaran，“集合知プログラミング”，オライリー・ジャパン，pp.127-153，(2008).
[5] 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).
[6] Akira Shimojo，Saori Kamada，Shinsuke Matsumoto，Masahide Nakamura，“On Integrating Heterogeneous Lifelog Services”，In The 12th International Conference on Information Integration and Web-based Applications & Services (iiWAS2010)，pp.261-268，(2010).
[7] Steven Bird，Ewan Klein，Edward Loper，“入門 自然言語処理”，オライリー・ジャパン，pp239-pp257，(2010).