

OWL を用いた音楽嗜好データの表現と音楽情報推薦への応用

武内 裕一[†] 青野 雅樹[†]

[†]豊橋技術科学大学 情報工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: [†]{ytake, aono}@kde.ics.tut.ac.jp

あらまし インターネットの急速な発展により、ネット上には多くの情報が溢れている。しかし、その莫大な情報の中から個人が欲する情報を獲得することは困難である。そこで、多くの情報の中から、個人の目的に沿った情報を得ることができる情報フィルタリングが注目されている。その手法の一つに協調フィルタリングがあるが、個人の嗜好に合わない情報を推薦する場合が多々ある。本稿では、個人の嗜好データに応じた情報推薦を目的とし、個人個人の音楽嗜好データを収集し、OWL での表現を試みた。それを基に、収集した人同士の OWL 構造を活かした類似度を定義して音楽情報推薦を行った。あわせて、既存手法である協調フィルタリングによる情報推薦と本手法を比較したので報告する。

キーワード OWL, オントロジー, 情報推薦, 類似度

OWL Representation of Personal Music Preference Data and The Application to Music Information Recommendation

Yuichi Takeuchi[†] Masaki Aono[†]

[†] Toyohashi University of Technology 1-1 Tenpaku-cho, Hibarigaoka, Aichi, 441-8580 Japan

E-mail: [†]{ytake, aono}@kde.ics.tut.ac.jp

Abstract With the rapid spread of the Internet, more and more massive dataset is available on-line. However, it is difficult to acquire a valuable piece of information that a person really wants. Information filtering has been one of the techniques that extract important pieces of information for individuals. Among many information filtering techniques, we will focus on the recommender system for an individual by taking music data as an example. In this paper, we will propose a new method for recommending songs by using a personalized data structure with an OWL, Web Ontology Language, that reflects person's preference, and by introducing a similarity measure, which we call "OWL structural similarity", between two OWLs. We also present preliminary experimental results including the comparison between collaborative filtering, Jaccard coefficient-based method, and proposed OWL structural similarity-based method.

Keyword OWL, ontology, information recommendation, similarity

1. はじめに

インターネットの急速な発展に伴い、インターネットから得られる情報は年々増えつつある。その膨大な情報の中から、利用者が所望する情報を適切にかつタイムリーに抽出して提示する研究が注目を集めている。そのような技術の中でも、個人個人が持つ様々な嗜好に応じた情報推薦技術は、電子商取引をはじめ多くの応用分野を有する。

個人が所望する情報を獲得するもっとも一般的に利用される手段には、Yahoo!やGoogle等に代表されるポータルサイトの検索エンジンを用いて、これにクエリーを入力し、検索エンジンにより提示されたホームページを通して獲得する手段がよく用いられる。一方、個人の嗜好に応じた情報を獲得することに関しては、

個人の嗜好に基づく情報推薦システムの研究やアマゾンに代表されるeShop等への実用化が盛んである[1]。しかし、どちらの場合においても推薦されるものが必ずしも自分の欲するものでないことが多々ある。

本稿では、個人の嗜好データに応じた情報の獲得を目的として、個人の音楽嗜好データを取り上げ、これをオントロジー記述言語として代表的なOWLで表現する。そして、OWL間の類似度を定義し、その類似度に基づく情報推薦の適用実験例として、音楽情報の推薦を報告する。本稿で提案する類似度としては、OWLの構造的な類似度を積極的に利用する手法を用いる。またOWLでは、一般に任意の属性値をもつノードを“DataType Property”ノードとして付加できる機能があり、これと構造的な類似度を組み合わせることも可能

である。本手法の有効性を確認するために、本手法と協調フィルタリング、および個人の音楽嗜好データを集合論的に扱う類似度(Jaccard係数)に基づく音楽情報推薦との比較実験を行った。

以下、2節では本研究に関連する研究について、3節では研究のアプローチについて述べる。4節では音楽嗜好データのOWLによる表現例、5節では2つのOWL間の類似度の算出方法の提案、6節では音楽情報の推薦について説明する。7節では音楽情報推薦の実験と評価について述べ、最後にまとめと今後の課題を8節で行う。

2. 関連研究

関連研究として、個人のプロフィールをシステム間で共有することで、システムごとの適応精度の差異を軽減する“A3”がある[2]。これは、本研究と同じく個人の嗜好データをOWLで記述し、プロトタイプとして本の推薦システム:PDLを紹介している。A3では、木の深さに応じたべき乗類似度を使用しているが、評価実験は行っていない。

また、選択した商品と同じ商品を購入した他人の購買商品を利用者に推薦するアマゾンがある[1]。しかし、アマゾンの推薦の場合、購買履歴のみを参照するために、その人が欲していない商品が推薦される場合が多々ある。アマゾンでは、情報を推薦する手法に情報フィルタリングで代表的な協調フィルタリングを用いている。協調フィルタリングは、複数の利用者の情報を使う必要があるため、利用者数が少ない初期の段階では類似する利用者を発見できず、有効な推薦が期待できないコールドスタート問題が発生する[3]。また、アマゾンの情報推薦システムの場合、クレジットカードをもつ本人が、家族用の商品(書籍やCDなど)を購入した履歴とカード保持者本人の購買履歴が混在してしまい、正しい情報推薦からかけ離れてしまうという状況も問題となる。

音楽データを対象とする推薦システムでは、岩濱ら[4]の研究がある。この研究では、音楽データ(MIDI)から特徴量を抽出し、その特徴量と個人の音楽データに対する評価値から決定木を構築する。その決定木と評価用データの特徴量を比べ、その人に音楽データの推薦を行う。しかし、音楽データから特徴量を抽出するのは困難である。

3. 研究のアプローチ

本研究では、A3で行われなかった評価実験を行うとともに、本手法で定義するOWL間の構造的な類似度の有効性を確認するために、協調フィルタリングと音楽データを集合論的に扱った類似度(Jaccard類似度)を用いた音楽情報推薦の結果との比較を行う。

3.1. 研究の進行方法

本研究では、以下に示す手順で研究を進めた。

3.1.1. 音楽データの収集

本研究では、個人の音楽嗜好データを用いるため、アンケートによる音楽嗜好データの収集を行った。アンケート方法として、アンケート用紙による収集、電子メールによる収集、WebページのFORM入力からの収集など様々な手法でデータ収集を行った。アンケート内容は、アーティスト名、曲名、アルバム名、その曲のCDまたはデータを持っているかどうか、その曲に対する嗜好の5段階評価(すごく好き、好き、普通、興味なし、嫌い)の5つの項目について記述してもらった。合計81名から音楽嗜好データを収集した。

3.1.2. 音楽嗜好データのOWLでの表現

アンケートにより収集した個人の音楽嗜好データとしての表現形式としてOWLを採用した。これは、OWLで表現されたデータへの新規属性の拡張容易性や再利用性、推論等の研究への適用を意識したためである。OWL表現の補助ツールとしてStanford大学で開発されたオントロジー作成支援ツールであるProtégé[5]を、本手法で用いるOWL作成のフレームワークとして利用し、個人の音楽嗜好データ用のOWLのテンプレートを出力した。このテンプレートOWLを基に、アンケートで得られた嗜好データをOWLファイルで表現するプログラムを作成し、81名の音楽嗜好データのOWLファイルの作成を行った。

3.1.3. 類似度の算出プログラムの実装と音楽情報の推薦

OWLファイルから構造的な類似度と、データの集合論的な類似度を与えるJaccard係数による類似度を求めるプログラムの実装を行った。一方、強調フィルタリング手法も実装し、これら3つの手法による音楽情報の推薦を行った。

3.1.4. 本手法の評価

本手法の有効性を検証するために、前記3つの手法での音楽情報の推薦を、推薦適合率(推薦された曲のうち、その人が好んだ曲の割合)を定義して、比較評価を行った。

4. 音楽嗜好データのOWLによる表現例

本提案手法で用いる個人の音楽嗜好データをOWLで記述した例を図1(深さ4)に示す。根ノードからgenre, sub-genre, artist, titleの階層構造とした。葉ノードであるtitleには、OWLの特徴であるデータタイププロパティ属性ノードを利用して、その曲を持っているかを表す“have”(boolean)、その曲に対しての興味を示す“like”(非負の実数値)のデータタイププロパティを持

たせるように工夫した. OWLの作成は3.1.2節で述べた手法で行った.

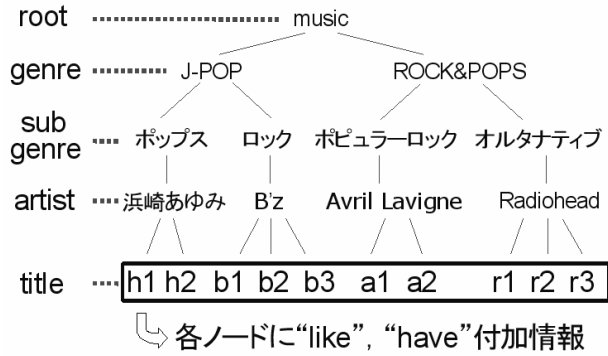


図1 OWLで記述した例 (Xさん, 深さ4)

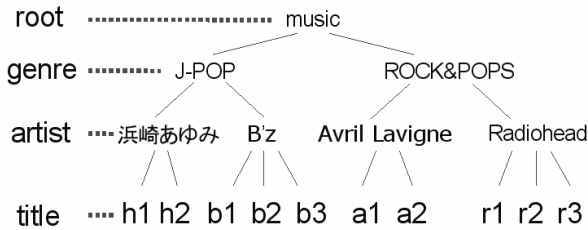


図2 OWLで記述した例 (Xさん, 深さ3)

また, 深さを変化させた場合の類似度を求めて音楽情報の推薦を行う実験も行った. Xさんの深さ4のOWL表現は図1の構造になるが, 深さ3のOWLは図2のように sub genreが無い構造とした.

5. 類似度算出法

OWLで表現された任意の2人の音楽嗜好データの類似度計算には, 以下に示す2種類の類似度算出法を用いた.

5.1. Jaccard 係数

Jaccard係数は, データを集合論的に扱った場合の類似度を求める一手法である. XさんとYさんの音楽嗜好データがあった場合, XとYが持つノードの和集合とXとYの持つノードの積集合の比で以下の式(1)で求めることができる (musicのノードは含まない).

$$sim_{Jacc}(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}. \quad (1)$$

Jaccard係数を求める例として, 深さ4の場合の図1のXと図3のYのJaccard係数を実際に求めてみる. XとYのノードの和集合は28, 積集合は9となり, Jaccard係数は32.1%となる. Jaccard係数を用いる場合, OWLの構造は一切利用していない

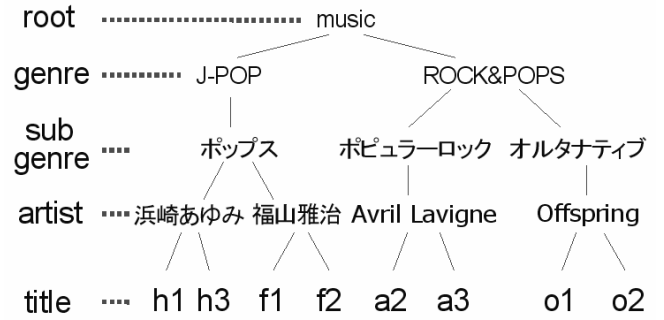


図3 OWLで記述した例 (Yさん, 深さ4)

5.2. 一般化コサインの拡張「OWL 構造類似度」

一般化コサイン類似度(GCS: Generalized Cosine Similarity)は木構造データ間の類似度を表現するひとつの手段として提案された[6]. 類似度を求める2人の音楽嗜好データ(Xのtitle: l_1 , Yのtitle: l_2)に対する一番近い共通の親(LCA: Lowest Common Ancestor)を用いてtitle間の類似度 $\vec{l}_1 \cdot \vec{l}_2$ を求める. 我々は, [6]で定義された類似度 ($\vec{l}_1 \cdot \vec{l}_2$) に対して, 深さによって変化する重みD(depth)を乗ずるような拡張を行った(式(2)).

以降, これを「OWL構造類似度」と呼ぶことにする.

$$\vec{l}_1 \cdot \vec{l}_2 = \frac{2 * depth(LCA_U(l_1, l_2)) * D(depth)}{depth(l_1) + depth(l_2)} \quad (2)$$

次に, OWL特有の拡張機能の一つであるデータプロパティとしてtitle(曲目)ノードに追加した“like”(式(3)の a_i と b_j)と式(2)を一般化した $\vec{l}_i \cdot \vec{l}_j$ との積の和を以下の式で求めた.

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \vec{l}_i \cdot \vec{l}_j \quad (3)$$

AはXのtitle(l_i)集合, BはYのtitle(l_j)集合を表す. 式(3)を正規化した値を「OWL構造類似度」とする(式(4)).

$$sim(A, B) = \frac{\vec{A} \cdot \vec{B}}{\sqrt{\vec{A} \cdot \vec{A}} \sqrt{\vec{B} \cdot \vec{B}}} \quad (4)$$

LCAの例として, 図1のXのtitle“h1”と図3のYのtitle“f2”のLCAを求める. まず, “h1”の親ノードは「浜崎あゆみ」, “f2”の親ノードは「福山雅治」となるため, artistの段階ではLCAは求まらない. 次に, 「浜崎あゆみ」の親ノードは「ポップス」, 「福山雅治」の親ノードも「ポップス」となり, sub genreの段階でLCAが求まることになる.

6. 音楽情報の推薦

前節で類似度を求める2つの手法について説明したが、本節ではこれら2つの類似度を用いた場合の音楽情報の推薦を行う手順、及び協調フィルタリングを用いた場合の音楽情報の推薦を行う手順を説明する。

6.1. 類似度を用いた場合の音楽情報推薦

前節で定義した類似度を用いた場合の音楽情報推薦の手順を説明する。音楽情報を推薦してほしい人Aさんを基準とし、他の人との類似度の計算を行う。類似度の計算を行った人数の上位数%(実験では10%)の人々のOWLに現れる「曲目」をマージする。マージした音楽データで重複する曲目があれば削除し、AさんのOWLに含まれる曲目を引く。以上の手順より出来上がったOWLに含まれる曲目において、データタイププロパティ“like”の評価値が大きい順に、上位の適当な数までの曲をAさんに提示し、評価をしてもらう。評価には式(5)に示す推薦適合率を用いた。

$$\text{推薦適合率} = \frac{\text{like評価の曲数}}{\text{推薦全曲数}} \quad (5)$$

式(5)におけるlike評価とは、Aさんに推薦された曲を“すごく好き”、“好き”または“曲を知らないが聞いてみたい”(注：これはアンケート項目にはない)と思った場合の評価である。この推薦適合率を用いることで、音楽情報推薦の各手法の評価を行った。

6.2. 協調フィルタリングを用いた音楽情報推薦

OWLで表現された音楽嗜好データに協調フィルタリングを適用した場合の音楽情報推薦は、音楽情報を推薦する人の曲に対する評価(データタイププロパティのlikeの値)を手がかりにして、音楽嗜好の似ている人を見つける。そしてまだ視聴していない曲や評価を行わなかった曲の推薦度を推定することで、その人へ音楽情報推薦を行う。協調フィルタリングの式は、たとえば[7]を参照されたい。

7. 本手法の評価実験

本手法による音楽情報推薦の性能を評価するために、2つの類似度(Jaccard係数(式1)、OWL構造類似度(式4))による音楽情報推薦、および協調フィルタリングによる音楽情報推薦の比較実験を行った。使用する音楽嗜好データは、アンケートにより収集した81人分のデータである。7.1.では2つの類似度を求める手法における類似度の比較を行い、7.2.では2つの類似度による音楽情報推薦の結果を示す。7.3.では協調フィルタリングによる音楽情報推薦の結果を示し、7.4.において類似度による音楽情報推薦の結果と協調

フィルタリングによる音楽情報推薦の結果を比較する。

7.1. 類似度の比較

音楽情報を推薦する特定の個人に対して、他の80人のOWLとの類似度の比較を行った。5節で示した2種類の類似度算出法を用いて、OWLの深さが3と深さ4の場合の類似度、およびJaccard係数による類似度結果を図4に示す。

OWL構造類似度はJaccard係数よりも大きな値となった。また、OWL構造類似度においては、深さ3より深さ4の類似度が大きくなる結果となった。

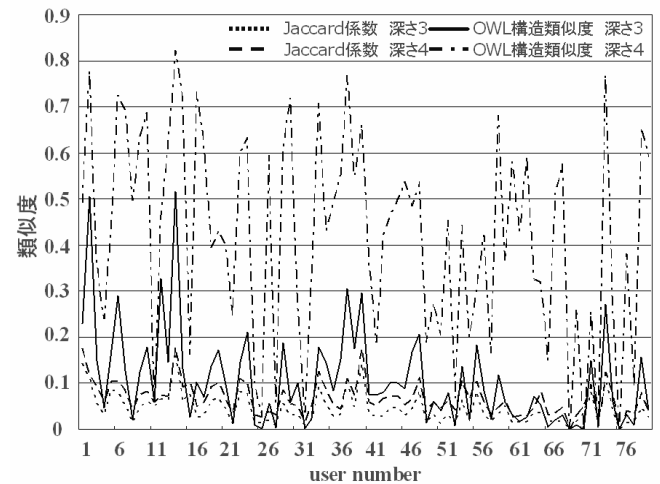


図4 Jaccard係数とOWL構造類似度結果の比較

OWL構造類似度の深さの異なる表現による類似度の差は、深さ3に比べて深さ4のOWLの方がsub genreの階層が増えている分、より近い親がLCAの値となるため、式(2)の値が大きくなり、深さ4のOWL構造類似度の値の方が大きくなったものと推察される。

7.2. 類似度による推薦の比較

Jaccard係数とOWL構造類似度を用いた場合の音楽情報推薦を比較した。類似度を求めるためにOWLを比較する人数を40人、80人の場合でそれぞれ実験を行い、さらにOWLの深さ3と深さ4の場合における各手法による音楽情報推薦の推薦適合率を求めた。類似度の大きかった上位10%の人数の音楽嗜好データを用いて音楽データの推薦を行い、求めた推薦適合率を表1に示す。

表1 類似度による推薦適合率

	40人 深さ3	80人 深さ3	40人 深さ4	80人 深さ4
Jaccard係数	44%	44.7%	40%	42%
OWL構造類似度	35%	44.7%	58%	62.5%

表1より、Jaccard係数による推薦適合率は人数や

OWL の深さを変えても大きな変化は見られなかった。これは、Jaccard 係数を求めるには 2 人の OWL のノードの積集合/和集合で計算されるため、OWL の構造が深さ 3 または深さ 4 であっても、sub genre があるかないかの違いであるから、Jaccard 係数の値に特に大きな変化は見られなかったためと考えられる。

OWL 構造類似度の場合、比較人数や OWL の深さを変えることで変化が見られた。特に深さが 3 から 4 になった場合の推薦適合率が大きく向上していることが表 1 から分かる。この理由として、OWL の構造が sub genre の階層を追加し深さ 4 になることで、音楽情報を推薦するある人の嗜好に近い別の人の OWL との類似度が OWL の構造の深さ 3 の場合に比べて高くなるためと考えられる。つまり、OWL の構造を拡張し、より詳細な音楽情報を付与することで、より個人の音楽嗜好に近い曲が推薦されるので、推薦適合率が向上したと思われる。人数を 40 人から 80 人に増やした場合の推薦適合率の向上は、音楽情報を推薦する人の嗜好に近い人の OWL が増えることから、より好まれる曲が推薦されたためであると考えられる。

7.3. 協調フィルタリングによる推薦

協調フィルタリングを用いて、推薦する 1 人に対して推薦のために利用する人数を 40 人と 80 人の 2 つの場合で、音楽情報の推薦を行った。なお、評価に用いる指標は、類似度による音楽情報推薦でも用いた推薦適合率を用いた。表 2 に協調フィルタリングによる推薦適合率の結果を示す。

表 2 協調フィルタリングによる推薦の推薦適合率

	40 人	80 人
推薦適合率	20%	25%

表 2 より、40 人を用いた場合の協調フィルタリングの推薦適合率は 20% という結果になった。協調フィルタリングは、精度を増すためには母数の人数を増やす必要があるため、人数を 40 人の倍の 80 人に増やして推薦適合率を求めると 25% となり、多少の適合率の向上が観測された。

7.4. 類似度による推薦と協調フィルタリングによる推薦の比較

類似度を用いた場合の音楽情報推薦と協調フィルタリングによる音楽情報推薦の結果を比較するため、各手法により求めた推薦適合率を表 3 に示す。表 3 より、本手法である OWL の構造的な類似度を用いる OWL 構造類似度（深さ 4）の推薦適合率が大きな値となることが分かる。情報フィルタリングの一手法である協調

フィルタリングよりも推薦適合率が高くなったことで、本手法の有効性をある程度確認できた。

表 3 各手法による推薦適合率

	40 人深さ 4	80 人深さ 4
Jaccard 係数	40%	42%
OWL 構造類似度	58%	62.5%
協調フィルタリング	20%	25%

8. おわりに

本研究では、個人の嗜好データとして音楽嗜好データを取り上げ、アンケートにより収集した音楽嗜好データをオントロジー記述言語として代表的な OWL で表現した。OWL の類似度を Jaccard 係数と一般化コサイン類似度を拡張した OWL 構造類似度の 2 種類の方法で類似度を算出し、適用分野として、音楽情報の推薦例を報告した。また、協調フィルタリングによる音楽情報の推薦も行い、本手法との比較実験を行った。

今後の課題として、OWL の特徴である多種多様なセマンティックス（属性）を付与したノードをデータタイププロパティとして付加できる機能を活用して“like”, “have”以外のデータタイププロパティを使った音楽情報の推薦、Edit Distance 等、異なる類似度での比較実験、音楽情報の推薦システム（GUI）の作成、コールドスタート問題への対処、音楽以外のデータを本手法に適用した場合の有効性を示すこと、等が挙げられる。

謝辞 本研究は、電気通信普及財団の援助を受けて行いました。

文 献

- [1] Greg Linda, Brent Smith, and Jeremy York, “Amazon.com Recommendations Item-to-Item Collaborative Filtering,” IEEE Internet Computing, Jan-Feb.2003, pp.76-80, 2003
- [2] 宮上大輔, 河合由紀子, 田中克己, “A3:オントロジーの共有によるユーザ適応のためのフレームワークの提案,” DBWS 2004, no.I-3, 2004
- [3] 小原恭介, 山田剛一, 絹川博之, 中川裕志, “Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム,” The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005, no.2C2-02, 2005
- [4] 岩濱数宏, 土方嘉徳, 西田正吾, “決定木を用いた音楽情報フィルタリングシステムとその有効性の検証”, 電子情報通信学会論文誌, D-I Vol.J88-D-I, No.3, pp.642-656, 2005
- [5] Protégé, <http://protege.stanford.edu/>
- [6] P. Ganesan, H.G. Molisa, J. Windom, “Exploiting Hierarchical Domain Structure to Computer Similarity,” ACM. Trans. Inf. Sys. Vol.21, No.1, pp.64-93, 2003.
- [7] 古川貴志, 増田宏, “ユーザ操作可能な協調フィルタリング推薦に関する研究,” 東京大学教養学部修士論文, 2002