

利用者の嗜好に適応した推奨システムに関する研究

映画推奨システムを題材に

2016SE079 竹内俊人

指導教員：沢田篤史

1 はじめに

近年、個人に適応した検索システムが数多く普及している。この検索システムは数多くのアプリケーションシステムで用いられている。その際、一般的にレコメンダシステムが使われる。レコメンダシステムとは、特定にユーザが興味をもつとされる情報のお勧めを提示するシステムのことである。しかし、レコメンダシステムにはユーザへの適応の精度が必ずしも高くないという問題点が存在する。

本研究の目的は推奨の履歴から得られる嗜好に関する知識の体系化により利用者へより高精度な適応を実現するレコメンダシステムのためのアーキテクチャ設計である。本研究では、映画のシステムを題材にする。本研究では、セマンティック web 技術で記述された普遍オントロジーと個人オントロジーの体系化をおこなう。体系化されたオントロジーを用いた協調フィルタリングを可能とするアーキテクチャ設計をする。その結果、協調フィルタリングにより嗜好が合うユーザ同士を比較することによりそのユーザの嗜好により適したものを推奨することができると考えている。

2 問題点

2.1 現在における推奨システムの問題点

個人に適応した検索システムの課題として個人への精度の高い適応が完全でない問題点が存在する。レコメンダシステムで一般的に用いられている技術手法に協調フィルタリング手法がある。協調フィルタリングとはユーザ同士を比較して類似度の高いユーザを推薦する方法論である。しかし協調フィルタリングで行われる推薦ではユーザ同士の嗜好が完全に一致しないという問題点がある。例として、映画を題材にこの問題について考える。ユーザ A とユーザ B が映画 z を見ていたとする。この二人のユーザの類似度が高いことからユーザ B に対してユーザ A が見ていた映画 x を推奨する。このときユーザ B は映画 x が自身の嗜好に合うものではなかった。この問題点を解決するために我々はオントロジーを用いることで解決できると考えている。オントロジーは個々の嗜好をメタデータを用いて管理、収集がおこなえるのでより個人の嗜好に適応することができるからである。

2.2 RDF

RDF(Resource Description Framework) とは、[2]web 上にあるリソースのメタデータを記述するための枠組みの一つである。RDF のメタデータのモデルでは、主語、述語、目的語の 3 つの要素でリソースに関する関係情報を

表現し、これをトリプル (triple) と呼ぶ。主語は記述対象のリソースを表現し、述語は主語の特徴と目的語の関係を示す。目的語は主語との関係のある物や述語の値を表現する。

2.3 協調フィルタリング

協調フィルタリングとは、多くのユーザの嗜好情報を蓄積し、あるユーザと嗜好の類似した他のユーザの情報を用いて自動的に推論を行うレコメンダ手法の一つである。協調フィルタリングには主にユーザの行動履歴からアイテム間の類似度を計算し、類似するアイテムをおすすめするアイテムベース型と、ユーザの行動履歴からユーザ間の類似度を計算し、類似するアイテムをおすすめするユーザベース型がある。今回、セレンディピティ (素敵な偶然に出会ったり、予想外のものを発見できること) が高くなると考えたためユーザベース型のアルゴリズムを用いることにする。

3 推奨履歴から得られる利用者の嗜好の体系化に基づく推奨システムの設計

本研究では、推奨履歴から得られる利用者の嗜好に関する知識と普遍的な知識をセマンティック web 技術により統合することで、ユーザの嗜好を精度高く取得し、推奨に利用できるシステムのアーキテクチャ設計を行う。以下、映画を推奨するシステムを例に説明する。

提案する映画推奨システムのユースケースは次の通りである。

1. ユーザが閲覧した映画について評価を心理情報、人的情報、その映画に 1~5 段階の評価をつけ、高評価の要素について Like プロパティを付与する。心理情報については Plutchik の「感情の輪」で定義されている 32 種類の感情表現を用いる。また、ジャンルの要素について LOD として利用する映画に関するデータベースの IMDB(Internet Movie Database)[1] で定義されている 23 種類で分類する。
2. 推奨システムのユースケースを示したデータモデルを図 1 に示す。システムがユーザが閲覧した映画の評価をデータセットで作成する。
3. 「Like」で評価された要素をオントロジーとして構築する際に、メタデータのモデルの特性で「Like プロパティ」として定義する。また、映画に関するオントロジーの構成要素として「人」、「ジャンル」、「感情」の 3 つの要素とユーザに対する情報を構成要素とする。定義したプロパティを含め RDF のメタデータモデルで

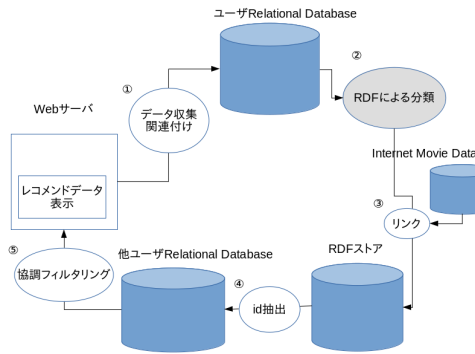


図1 データモデル図

あるトリプルを用いて個人オントロジーを構築する。例)見た映画=「Star Wars」 好き 心理的要素「驚嘆」,人的要素「Natalie Portman」。実装するRDFモデルの一部を図2に示す。このモデルに基づきRDF

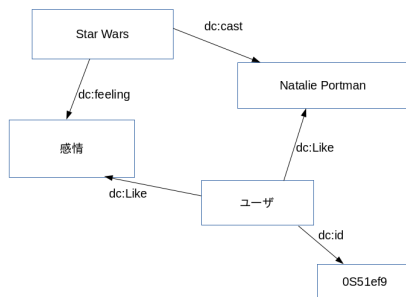


図2 実装するRDFモデル

を用いた構文を図3に示す。

```
<?xml version='1.0' encoding='UTF-8'?>
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:dc='http://purl.org/dc/elements/1.0/'
xmlns:ex='http://example.org/stuff/1.0/'
xml:lang='ja'>
<rdf:Description
rdf:about='https://www.imdb.com/user/ur110968605/?ref_nv_usr_prc'>
<ex:Likes>
<rdf:Bag>
<rdf:li>Natalie Portman</rdf:li>
<rdf:li>驚嘆</rdf:li>
</rdf:Bag>
</ex:Likes>
<dc:id>OS51ef9</dc:id>
</rdf:Description>
<rdf:Description
rdf:about='https://www.imdb.com/title/tt0121766/?ref_=rvl_tt'>
<dc:cast>Natalie Portman</dc:cast>
<dc:feeling>驚嘆</dc:feeling>
</rdf:Description>
</rdf:RDF>
```

図3 映画に関するRDF構文

4. 構成された個人オントロジーから SPARQL により、Like プロパティを持つ要素を含む他ユーザーの id を抽出する。rdflib という python ライブラリを用いて RDF の構文を読み込み、クエリでデータの抽出を行う。図4に結果を示す。
5. 抽出された id をもつユーザーをユーザーのデータセットより検索、その id をもつユーザーが見てる映画とユーザー A の映画で協調フィルタリングを行い、ユーザー A に推奨を行う。本研究では提案されるシステムで示さ

```
D:\Users\16se0794\Documents\kkyu folder>python
Python 3.8.0 (tags/v3.8.0:fa919fd, Oct 14 2019, 19:37:50) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> from rdflib import Graph
INFO:rdflib:RDFlib Version: 4.2.2
>>> g=Graph()
>>> g.parse('movie.rdf')
<Graph identifier=N1baeb7ea7e4b44ad6fccfa5c571111 (<class 'rdflib.graph.Graph'>)>
>>> from rdflib import Namespace
>>> cd = Namespace('http://purl.org/dc/elements/1.0/')
>>> cd.id
rdflib.term.URIRef('http://purl.org/dc/elements/1.0/id')
>>> for o in g.objects(subject=None,predicate=cd.id): print(o)
...
Traceback (most recent call last):
  File <stdin>, line 1, in <module>
TypeError: objects() got an unexpected keyword argument 'subject'
>>> for o in g.objects(subject=None,predicate=cd.id): print(o)
...
OS51ef9
>>>
```

図4 抽出結果

れる協調フィルタリングのプログラムによる実行について、ユーザーの評価を5段階として実行を行い、実行結果として評価したクラスが抽出された。実行結果を図5に示す。

```
C:\Users\16se0794\Documents\kkyu folder>python recommend_filtering2.py
OS51ef9のStar Warsの評価: 4
OS51ef9とEc67s3の類似度(ユーザーリッド距離) 0.5
OS51ef9さんとEc67s3さんの類似度(ピアソン相関係数) 0
Guda72に似た入ベスト 3 [(1.0000000000000004, 'Guda72'), (1.0, 'Kd142o'), (0, 'Sid103')]
Guda72におすすめのクラス ['驚嘆', 'Ironman']
C:\Users\16se0794\Documents\kkyu folder>
```

図5 実行結果

4 考察

本研究で提案される知識の統合を現実的に行った場合、ユーザーの数だけ嗜好情報が増え、システムの処理時間や処理数などが増える予想される。その状況を考慮すると推奨システムの有用性について確認する必要がある。そのためには、本研究で提案されるシステム全体の実装を行う必要があると考えている。よって、我々は全体の実装を行ってから現実的にシステムが機能できるかどうかをわかると考えている。

5 まとめ

本研究では推奨の履歴から得られる嗜好に関する知識の体系化により利用者へより高精度な適応を実現する推奨システムのためのアーキテクチャ設計を行った。結果として提案したシステム全体の実装を行うまでには至らなかった。しかし、提案したアーキテクチャ設計により嗜好に関する知識の体系化を行え、高精度な適応が見込めることは確認できた。今後は現実的にシステムが機能するかどうかを知るため、システム全体の実装を行う必要があると考えている。

6 参考文献

参考文献

- [1] Colin Needham : Internet Movie Database, http://www.imdb.com/?ref_nv_home.
- [2] W3C(MIT, ERCIM, Keio) : RDF Data Access Use Cases and Requirements, <http://www.w3.org/TR/rdf-dawg-uc/>.