

## 履修プラン策定のためのレコメンダシステム

藁輪拓† 寺澤卓也†

東京工科大学 メディア学部†

## 1. はじめに

大学生は、履修科目のほとんどを自らの選択で決定しなければならない。この科目選択は悩ましい作業である。

一方、近年、インターネット上ではお勧め機能が付随したサービスが増えている。これによりユーザーの嗜好や状況に合った情報選択が容易になってきている。そこで、本研究では履修計画にレコメンダシステムを導入することで、学生の科目選択を支援する手法を検討した。

本研究では、livedoor がオープンソースとして公開しているレコメンダエンジン「Cicindela [1]」を利用してシステムの構築を行った。そして筆者が在籍する東京工科大学メディア学部での運用を想定し、実際の履修データを元にその有用性と改善点について考察を行った。

## 2. レコメンダエンジン

レコメンダシステムとして利用可能なエンジンは多数存在する。本論文ではその中から 2 つのエンジンについて触れておく。

「Mahout」は、オープンソースエンジンとして Apache より公開されている機械学習を行うためのライブラリである [2]。アルゴリズムが複数用意されており、プログラムを作らずとも大規模なデータを処理できる。

「Cicindela」は livedoor が公開するオープンソースのエンジンである。こちらも複数の設定を試すことができる。http ベースの WebAPI が用意されているので簡単にさまざまなシステムに組み込むことができる [1]。

本研究では導入が比較的容易であり、応用性に期待が持てた Cicindela をエンジンとして利用することにした。

## 3. 想定する運用状況

本研究で想定したのは、筆者が在籍する東京工科大学メディア学部在学生の履修状況に対する適用である。実験を行うにあたり、本学学務課の協力を得て、2010 年度入学者の中から匿名の 200 人をランダムに抽出し、彼らの 4 年間の

履修データを個人情報に配慮した上で提供していただいた。提供されたデータからは、必修科目や試験等でクラスが決まるような科目は除外し、学生が自ら選択した科目に対してのみ処理を行えるような形にした。

実際のレコメンダシステムには、買い物履歴を利用するなど、レコメンダされるユーザーの嗜好情報を同時に取得し、その都度データベースの精度を上げているものもある。しかし今回はすでに出来上がっているデータベースに対して、システムの適用を行う。インターフェース的側面は実験に必要な部分だけにとどめ、結果を元にそれらについて考察していくことにする。

## 4. 実装

本研究では 3 で提示した状況に従い、実際にデータをエンジンに処理させた。また Cicindela には http ベースの入力・出力用 WebAPI が導入されているのでそれらもあわせて活用した。データの入出力の流れを図 1 に示す。

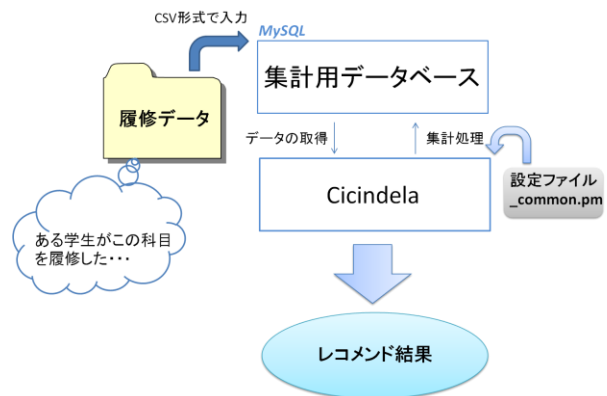


図 1 入出力の流れ

### 4.1 データベースについて

データベースには Cicindela と互換性のある MySQL を用いた。これにより、Cicindela 側の機能で自動的に集計用のテーブルを生成することができるようになった。入力は WebAPI を経由でも行えるが、今回は CSV ファイルの形式で直接データベースに流し込み、WebAPI 経由で格納した場合と同じ状態のデータベースを再現した。

### 4.2 Cicindela の設定について

エンジンの中に `_common.pm` というファイルがありこの中身を書き換えることでエンジンの

A recommender system to plan the college course selection

†Hiroshi Minowa and Takuya Terasawa

†Tokyo University of Technology, School of MediaScience

設定を行うことができる。今回の実験で用いた設定を図2に示す。

```
'pick' => {
  datasource => [ 'db:mysql:test01_db;host=localhost', 'root', 'password' ],
  filters => [
    'PicksExtractor',
    'InverseUserFrequency',
    'ItemSimilarities',
  ],
  recommender => 'ItemSimilarities',

  use_user_char_id => 1,
  use_item_char_id => 1,
},
```

⇒ 文字列を扱う場合のオプション

図2 Cicindela の設定

Filter 部分は「誰がどのアイテムを選択したか」「アイテム同士の関連性はどの程度か」を算出できるように設定した。またアイテム名、すなわち科目名には文字列を伴うことが予想されたので処理中に文字列を扱うためのオプションをあらかじめ加えておいた。

なお、出力に関しては WebAPI のうち「このアイテムを選んだ人はこのアイテムも選択しています」を経由して行った。

### 5. 評価実験

評価実験では、レコメンドの結果がどれぐらい学生にとって有意義なものになっているかを考えた。一つの科目に対し関連度の高い順に 20 件を出力し、それらの科目の分野、コース情報、その他科目特有の情報に注目し、データにどの程度の偏りが出るかを観察した。顕著にその偏りが現れたのは、三年次に履修することのできる専門分野の演習科目であった。それらの演習の基礎になる科目や、同じ教員によって開講されている科目などが結果として出力された。さらに、教養科目の履修にも特徴が現れた。例えば、映像分野の科目を履修している学生は、物理や数学などの理系科目の履修している傾向が見受けられ、ビジネス系の科目を履修している学生は、ディスカッションの多い科目を履修しているという結果がでた。これは学生自らの嗜好、得意分野が色濃く反映されているからであるといえる。その反面意外性のある傾向も見受けられた。例えば、音楽制作の科目を履修している学生がマーケティングの科目に興味を示しているケースがあった。恐らく、音楽分野を志す学生は自然とマーケティング分野にも目を向けがちなのだろう。少なくともこの結果はレコメンドシステム特有の、データ傾向に基づいた結果であった。その一部を図3に示す。

作曲演習 を取った人はこのような科目もとっています。

- 音楽の基礎
- コミュニティメディア論
- 造形の基礎
- 広告メディア論
- 数理基礎 I
- マーケティング論
- 芸術論
- 信号と情報
- 音楽製作技法の基礎
- サウンドストラテジー
- メディア産業論

Etc...

図3 レコメンド結果

これを踏まえたうえで、例えば、履修モデルの情報を付加させることができれば、習得したい資格や将来の目標のためにはどのような履修をすればいいかという指標にもなるだろう。Cicindela の場合、あるアイテムをあるカテゴリーに所属させる、という機能が備わっている。その機能を利用すれば、履修登録をする際、カリキュラム表を照合するわずらわしさも軽減できるだろう。

### 6. 終わりに

本研究では、筆者が所属する東京工科大学メディア学部での復習レコメンドシステムの運用を想定し、実際に 200 人分の履修データに対して実験、評価を行った。その結果、必然性と意外性の両方を兼ね備えたレコメンドが実現できた。今後の展望として、学部特有の科目別情報をうまく取り入れることで、さらに細かなレコメンドが実現できると考える。仮に実装するのであれば、大学側が履修データをデータベースに格納しやすいフォーマットを実現し、学生が気になる科目を選んだ際に、そのデータベースを元にレコメンドを行う形が一番理想的であると考えられる。Cicindela のように簡素な WebAPI が用意されているのであれば、それらの結果を授業科目の Web ページに表示することもできる。その他課題としては、授業科目に変更があった場合の対処や、別学年の科目が結果としてでる場合をどのように回避するかが挙げられる。

謝辞：データを提供していただいた本学学務課に感謝いたします。

### 参考文献

[1] Cicindela  
(<http://code.google.com/p/cicindela2/>)  
[2] Mahout  
(<http://mahout.apache.org/>)