

# 対訳対と協調フィルタリングを用いた商品推薦

柴田 翔平 (東京農工大学 工学部 情報工学科)  
古宮 嘉那子 (東京農工大学 工学研究院)  
小谷 善行 (東京農工大学 工学研究院)

## Product Recommendation using Translation Pairs and Collaborative Filtering

Shohei Shibata (Department of Computer and Information Sciences  
Faculty of Engineering, Tokyo Agriculture and Technology)  
Kanako Komiya (Institution of Engineering, Tokyo Agriculture and Technology)  
Yoshiyuki Kotani (Institution of Engineering, Tokyo Agriculture and Technology)

### 1. はじめに

近年、アニメや映画といった日本のメディア作品が海外で人気となり、それに関連する商品を外国人が購入する機会が増えている。しかし、日本語と外国語の間に存在する言語の壁から目的の商品を検索するのは難しい。特に、作品内の登場人物や地名は機械翻訳でも対応できない場合が多く、このような言語の壁が妨げとなって、外国人が目的の商品を購入できない場合がある。

本稿では、商品タイトルの日本語と外国語が対となった対訳対と、日本ユーザと外国ユーザの商品購入情報を基にして協調フィルタリングを用いることで、言語の壁を越え、外国人が目的とする商品を推薦することを目的とする。協調フィルタリングとは、一般的なインターネットのショッピングサイトでも利用される、ユーザの嗜好情報に基づいて推薦を行う方法であり、この方法を用いれば、日本と外国ユーザの嗜好の類似点を見つけ出し、言語をまたいで商品推薦を行うことができると考える。

### 2. 関連研究

商品推薦に関しては、これまで様々な研究が行われている。その中でも、類似しているユーザを選択するための類似度に関する研究が多くある。

ソーシャルネットワークサービス (SNS) 上でのつながりをグラフで表現した情報と、商品購入情報の二種類を用いて類似度を定義する研究 (Symeonidis and Tiakas and Manolopoulos (2011)) や、SNS 上のつながりの強さまで考慮した類似度を定義する研究 (Symeonidis and Tiakas and Manolopoulos (2010)) などがある。

また、類似度には時事性を含ませることが難しいため、時事性を含まない類似度を用いると、ユーザの嗜好の変化に合った推薦を行うことが困難にある。そのような問題に対応する研究も行われている。

意外性のある Web ページをリコメンデーションするため、ユーザの Web ページのブックマーク情報と Wikipedia のコンテンツを照らし合わせる研究 (Chang and Quiroga(2010)) や、ユーザ間の関係と嗜好の時間による変化を反映させた協調フィルタリングによる推薦の研究 (川前, 坂野, 山田, 上田(1997)) などがある。

しかし、我々の調査した限り、これまでの研究では言語をまたがる商品推薦は考えられていない。そこで本稿では、商品推薦に一つの国のユーザの情報を用いるだけでなく、二つ目の国のユーザの情報を用いることで、商品推薦の結果に幅を持たせ、言語をまたいだリコメンデーションが行えるようなシステムを提案する。

### 3. 対訳対を用いた商品推薦

ある外国ユーザを対象に、日本ユーザの商品購入情報から商品推薦を行うことを考える。外国ユーザと日本ユーザ間の類似度を計算し、協調フィルタリングによる商品推薦を行いたいが、商品購入情報をそのまま用いるだけでは言語の違いが存在するために、外国ユーザと日本ユーザの間で情報の共有ができず、類似度の計算が行えない。そこで、双方の言語の違いを埋めるため、商品タイトルの日本語と外国語との対関係を蓄積した「対訳対」を作成する。商品タイトルの対関係が存在している場合には、日本と外国で異なる商品タイトルでも、同じ商品を購入していると扱うことができる。

提案する商品推薦システムにおいて、外国ユーザと日本ユーザの商品購入情報は、ベクトル化して類似度計算に用いる。このベクトルの素性は商品であり、素性値は商品への評価値である。このベクトルと対訳対を用いて、同じ商品を購入して評価している外国ユーザと日本ユーザ間でコサイン類似度を計算し、協調フィルタリングによる商品推薦を行う。

日本ユーザの商品購入情報は「楽天株式会社」、外国ユーザの商品購入情報は「GroupLens Research」より提供していただいた情報を用いた。対訳対は、Wikipedia より配布されているダンプデータからタイトルの対応関係を抽出することで作成した。

なお、ユーザの商品購入情報や対訳対に存在する商品タイトルには、英語の大文字と小文字やバージョンの違いなど、表記の揺れが存在するため、それを削除した状態の情報も用いることとする。この情報を、商品タイトルを整形した情報と呼ぶ。

## 4. 実験

### 4.1 商品推薦システムの評価実験

商品推薦システムの出力となる推薦結果の評価は、推薦された商品がユーザの意図するものであったかという判断になる。しかし、その判断は主観的なものであるため、システムへの評価が集まったとしてもその性能について議論することは難しい。

そこで、システムの定量的な評価を行うため、外国ユーザの商品購入情報に存在する商品を対象に商品購入情報をマスキングした上で商品推薦を行った。商品を推薦した数のうち、商品推薦結果に現れるマスキングした商品の割合を「適合率」とし、また、商品推薦結果に現れるマスキングした商品の順位を用いた「平均逆順位 (MRR)」を定義し、商品推薦システムの評価実験の指標とした。適合率と MRR は、以下の式で計算される。

$$\text{適合率 } P = \frac{\sum_{e \in E} \sum_{j \in J} C_{ej}}{\sum_{e \in E} \sum_{j \in J} N_{ej}} \quad \text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}(i)}$$

適合率  $P$  において、 $C_{ej}$  はマスキングされた商品の数、 $N_{ej}$  は商品推薦数、 $e$  は外国ユーザ、 $E$  は外国ユーザの集合、 $j$  は日本ユーザ、 $J$  は日本ユーザの集合を表す。

MRR において、 $N$  はテストデータ数を表し、本稿では類似度が計算された外国ユーザと日本ユーザの組み合わせ数の 2 倍である。また、 $\text{rank}(i)$  は、 $i$  に対する推薦結果中、マスキ

ングされた商品の最高順位を表す。商品推薦の結果に正解が含まれなかった場合には、 $\text{rank}(i) = \infty$  とする。MRR が高いほど、推薦結果の上位にマスキングされた商品が出現しているということになる。なお、評価実験は、外国ユーザの購入している商品のうち、マスキング対象の商品を情報を二つに分割して、二分割交差検定によって行った。商品タイトルの整形有無も考慮に入れたため、計四種類の評価実験を行っている。また、評価実験結果だけでなく、実際のリコメンデーション結果についても示す。

## 4.2 実験結果

外国ユーザー一人あたり推薦される商品数を 1 から 5 まで変化させたときの適合率のグラフを図 1 に、MRR のグラフを図 2 に示す。また、実際の商品推薦結果の一部を表 1 に示す。

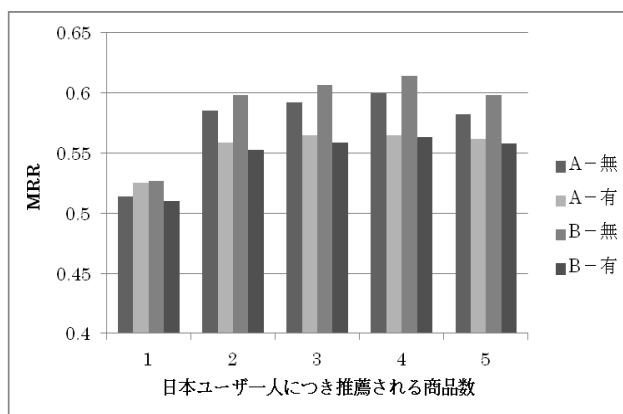


図 1 システムの評価実験における適合率

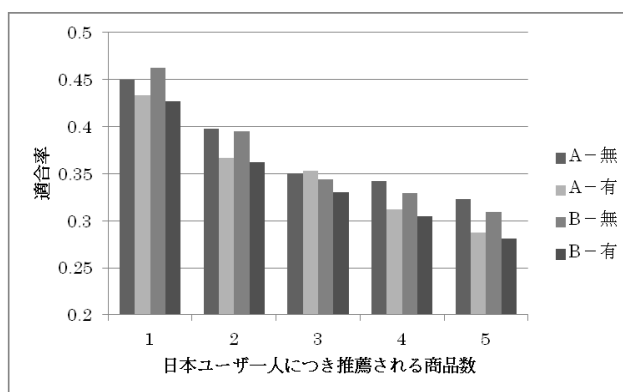


図 2 システムの評価実験における MRR の値

表 1 システムによる実際の商品推薦結果

英語ユーザ ID	日本ユーザ ID	商品番号	商品名
204269	68437	208	pinocchio
204269	68437	260	cinderella
204269	68437	76476	さるかにばなし
204269	68437	76477	三びきのこぶた
252953	16352	11025	Howl's Moving Castle
252953	16352	94061	となりのトトロ新装版

図 1 と図 2 において，凡例の A はマスクング対象を対訳対に存在する商品，B は外国ユーザの購入している商品を示す．また，有無は商品タイトルの整形を施したかどうかを示す．

図 1 と図 2 を見ると，マスクング対象 A,B ともに商品タイトルの整形を行わない方が適合率，MRR ともに高い値を示している．このことは，商品タイトルの整形によって商品タイトルの対応関係が増加したため，商品推薦の数自体は増加したものの，類似度が計算されるユーザも増えることでその推薦内容も多様になり，推薦結果にマスクングした商品が現れにくくなっていると考えられる．

しかし，表 1 に示したシステムによる実際の商品推薦結果を見ると，「pinocchio」（ピノキオ）や「cinderella」（シンデレラ）から「さるかにばなし」が推薦され，童話によるつながりから推薦が行われていると考えられる結果や，「Howl's Moving Castle」（ハウルの動く城）から「となりのトトロ」が推薦され，ジブリ作品のつながりから推薦が行われていると考えられる結果が存在した．このことから，評価実験とは別に，商品推薦システムは本稿の目的に沿った推薦を行っているといえる．

#### 4. 3 まとめと今後の展望

商品推薦システムの評価において，適合率は，商品タイトルを整形していない情報を用いた実験で，外国ユーザー一人あたりに推薦される最大商品数を 1 に設定したとき，適合率 46% を得た．また，MRR は，商品タイトルを整形していない情報を用いた実験で，外国ユーザー一人あたりに推薦される最大商品数を 1 に設定したとき，0.61 という値を得た．

また，表 1 に示した実際の商品推薦結果から，商品推薦システムは，本稿の目的に沿った結果を出力することができていると考えられる．

しかし，本稿の対訳対のみでは日本と外国の商品の対応関係が少ないため，今後は日本ユーザと外国ユーザとの間をより広く取り持てるような条件の拡張を行っていく必要がある．

#### 謝 辞

本研究を行うにあたり，楽天株式会社と国立情報学研究所が協力して提供している『楽天データセット』を利用させて頂いた．また，GroupLens Research より提供していただいたデータも利用させて頂いた．利用を快諾して下さった各社に謹んで御礼申し上げる．

#### 文 献

- Pei-Chia Chang and Luz M. Quiroga (2010). "Using Wikipedia's Content for Cross-Website Page Recommendations that Consider Serendipity". *Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence*, pp293-298.
- Panagiotis Symeonidis and Eleftherios Tiakas and Yannis Manolopoulos (2011). "Product Recommendation and Rating Prediction based on Multi-modal Social Networks" *Proceedings of the ACM Conference Series on Recommender Systems 2011*, pp61-68.
- Panagiotis Symeonidis and Eleftherios Tiakas and Yannis Manolopoulos (2010). "Transitive Node Similarity for Link Prediction in Social Networks with Positive and Negative Links" *Proceedings of the ACM Conference Series on Recommender Systems 2010*, pp 183-190.
- 川前徳章，坂野鋭，山田武士，上田修功 (1997). "ユーザの嗜好の時系列性と先行性に着目した協調フィルタリング". 電子情報通信学会論文誌D Vol.J92-D No.6,pp.767-776.