レビューテキスト間の類似度を用いた協調フィルタリング

岡田 瑞穂† 藤井 敦‡
† 東京工業大学工学部情報工学科
‡ 東京工業大学大学院情報理工学研究科計算工学専攻

1 はじめに

近年，ショッピングサイトや映画の評価サイトなどでは，ユーザにとって有用と思われる商品やコンテンツを提示する推薦システムが実用化されている．推薦システムは，推薦対象となるユーザがアイテムを好む度合いを予測する．予測値の計算には多くのユーザに関する嗜好情報を利用することが有効で，協調フィルタリング (Collaborative Filtering: CF) に関する研究が行われている．

CF ではユーザの行動履歴を〈ユーザ，アイテム，ユーザがアイテムに与えた評価値〉の組として収集し，各セルの値を評価値とするユーザ×アイテム評価値行列を作る．CF は大きくユーザ間型 CF とアイテム間型 CF に分類でき，どちらも機械学習の観点からは k-近傍法と見ることができる．ユーザ間型 CF [1] では，ユーザの各アイテムに対する評価値の傾向から嗜好パターンが似ている他のユーザを見つける．すなわち，評価値行列の行ベクトル同士の類似度によりユーザ間の類似度を計算し，近傍ユーザとの類似度と評価値に基づいて各推薦候補アイテムの評価値を予測する．アイテム間型 CF [2] では，推薦対象となるユーザが過去に評価したアイテムと推薦候補アイテムの類似性の観点から評価値を予測する．アイテム間型 CF ではアイテム同士の類似度を評価値行列の列ベクトルの類似度で表し，近傍アイテムへの評価値と類似度を考慮して評価値を予測する．

一般に CF に用いられる評価値行列では未評価のセルが全体のほとんどを占め，評価値を正確に予測するために十分なデータ量とは言えず，少ない評価値データから正確に近傍を測ることは難しい．そのため，評価値の予測に評価値行列以外の情報を用いることは有効である．近年においては，評価値と共にレビューを投稿できるシステムが増えてきている．レビューテキストにはアイテムの内容に関する情報とユーザの意見が含まれており，これはユーザの嗜好を測るための有用な情報源である．

似たようなレビューを書くユーザ同士は嗜好が類似していると考え，近傍ユーザと見なせる．また，ユーザが特定のジャンルの映画に興味を持っていれば，近傍アイテムとして同じジャンルの映画を選ぶことはごく自然な流れである．さらに,「怖い」や「わくわくする」といった感想に関する類似性からも近傍アイテムを選ぶ手法が考えられる．本研究は，ユーザによるレビューテキストがユーザーの嗜好やアイテムの特徴に関する情報を共によく表していると考え，テキストをユーザやアイテムの特徴ベクトルとする手法を提案し，映画評価サイトのデータセットを用いて評価する．評価実験では，レビューテキスト以外の外部情報も用いて様々な観点から考察する．

2 関連研究

中辻ら [3] は，ユーザ間型 CF におけるユーザ間の類似度をレビューテキストを用いて測った．中辻らは，ユーザがレビューを記述しつつアイテムに評価値を与えたときの形容表現を抽出し，各形容表現を感性クラスとして捉えてユーザ間の興味の一致度を感性クラスに対する興味の一致度で測った．その後，アイテムに対する評価行為の類似度と統合し最終的なユーザ間の類似度とし，ユーザ間型 CF を行いその有効性を示した．しかし，中辻らの実験では評価実験の規模が小さく，比較が不十分という問題がある．

Cane ら [4] は，映画レビューテキストの感情極性値を計算し，CF における評価値として利用する手法を提案した．Cane らは感情語の表す感情クラスとして Positive, Neutral, Negative を定義し，ファジーセットの概念を用いて同じ語に複数の感情クラスへの所属値を持たせることで，予測精度を向上させた．それぞれの語は，訓練データにおける評価値つきレビューテキストから評価値との出現頻度に基づき，各クラスへの所属度が求められる．Cane らの研究では，ユーザやアイテムをレビューテキストを用いた文書ベクトルで表す手法は検討されていない．

Niklas ら [5] は，レビューテキストを活用して CF における特徴に加えることで，映画評価値の予測精度を向上させた．Niklas らの手法では，映画の観点を表す語を特定し，各観点の意見語の極性値を平均して全体の極性値とした．比較実験により，評価値に加えてジャンル情報のみを用いたベースラインと比較して，意見語の極性と総量を考慮すると良い精度が得られることを示した．本研究では，語の極性は考慮せずに，頻度のみに基づき重みを与える手法について比較する．

3 協調フィルタリングの概要

3.1 ユーザ間型 CF

ユーザ間型 CF は，評価値ベクトルの行ベクトルをユーザの特徴ベクトルとする手法が一般的である [1]．ユーザ間の類似度は Pearson 相関係数や Cosine などで計算する．以下に Cosine で測った場合の手法を説明する．ユーザ u と v の特徴ベクトルをそれぞれ $\vec{u}$，$\vec{v}$ とする．ただし，評価値行列の未評価セルに対応する次元は無視し，同じユーザから共通に評価値を与えられているアイテムの次元だけでベクトルを構成する．すると u と v の類似度 $sim_{u,v}$ は式 (1) で計算する．

$$sim_{u,v} = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \ \ \vec{v}}{\|\vec{u}\|\|\vec{v}\|} \tag{1}$$

$sim_{u,v}$ を用いてユーザ u のアイテム i への予測評価値 $\hat{r}_{u,i}$ を式 (2) で計算する．

$$\hat{r}_{u,i} = \frac{\sum_v sim_{u,v} r_{v,i}}{\sum_v |sim_{u,v}|} \tag{2}$$

ここで，$r_{v,i}$ は他のユーザ v の i に対する実際の評価値であり，i を評価済みのユーザのみを考慮する．

3.2 アイテム間型 CF

アイテム間型 CF は，アイテム間の類似度を評価値行列の列ベクトル同士の類似度で計算する [2]．ユーザ間型と同様に Cosine で測る場合を説明する．アイテム i と j の特徴ベクトルをそれぞれ $\vec{i}$，$\vec{j}$ とすると，類似度 $sim_{i,j}$ は式 (3) で計算する．ただし，ユーザ u から共通に評価値を与えられている次元だけでベクトルを構成する．

$$sim_{i,j} = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \ \ \vec{j}}{\|\vec{i}\|\|\vec{j}\|} \tag{3}$$

u による他のアイテム j に対する実際の評価値 $r_{u,j}$ を用いて，i への予測評価値 $r_{u,i}$ は式 (4) で求める．

$$\hat{r}_{u,i} = \frac{\sum_j sim_{i,j} r_{u,j}}{\sum_j |sim_{i,j}|} \tag{4}$$

4 提案手法

本論文では，3 章で述べた CF における特徴ベクトルを，評価値ではなくレビューテキスト中の語に基づくベクトルで表す方法を提案する．ユーザ間型 CF の場合は，あるユーザが投稿した全てのレビューを，アイテム間型では，映画に投稿された全てのレビューをまとめて一つの文書とし，bag-of-words で表現する．この際，不要語は WordNet の stoplist により除去した．ユーザまたはアイテム a に対して文書 D_a がある状況を考える．D_a に含まれる索引語 t の重み $w(t, D_a)$ を tf-idf 法を用い式 (5) で定義する．ここで，n_{t,D_a} は単語 t が文書 D_a に出現する回数を表し，f_t は全文書数 $|D|$ のうち単語 t が出現する文書数を表す．

$$w_t^{D_a} = \left(\frac{n_{t,D_a}}{\sum_t n_{t,D_a}}\right) * \log\left(\frac{|D|}{f_t}\right) \tag{5}$$

また，文書 D_a に含まれる単語の異なり数を L としたとき，L 次元のベクトル $\vec{D_a}$ を式 (6) で定義し，a の特徴ベクトルとして用いる．

$$\vec{D_a} = (w_{t_1}^{D_a}, w_{t_2}^{D_a}, \ldots, w_{t_L}^{D_a}) \tag{6}$$

そして，二つの文書 a,b 間の類似度を式 (1)，式 (3) と同様に Cosine で測り，CF に適用する．

$$sim_{a,b} = \cos(\vec{D}_a, \vec{D}_b) \tag{7}$$

5 評価実験

5.1 方法

評価値を予測するタスクでは映画推薦システムの先駆的な研究グループである MovieLens[1] のデータセットが有名である．しかし，今回はユーザと映画それぞれに関するレビューテキストを利用したいため，独自にデータを集めた．その際，多くの研究で用いられる MovieLens 100k Data set と比べて遜色のない量のデータを収集した．英語の映画評価サイトである The Internet Movie Database[2](IMDb) から，128,030 の ⟨ ユーザ，映画，評価値 ⟩ の組を抽出した．その際，各ユーザはそれぞれ 20 以上の映画を評価し，各映画は 20 以上のユーザから評価されるように集めた．評価値の値域は 1 から 10 の 10 段階である．また，評価値を伴わないレビューテキストは除外した．MovieLens 100k Data set との比較を表 1 に示す．ここからランダムに 10,000 組のデータを抽出してテストデータとし，残りの 118,030 組を用いてテストデータの評価値を予測する．

表 1: データセットの比較

	ユーザ数	映画数	評価値数	密度
実験データ	834	3065	128,030	5.01%
MovieLens	943	1682	100,000	5.95%

実験では，評価値のみを利用した従来の手法とテキストを用いた手法とを比較する．その際，レビューテキストだけでなく，さまざまな代理テキストを用いた手法と比較することで提案手法の有効性を評価する．

評価方法として，評価値を予測するタスクにおいて一般的な指標である絶対平均誤差 (MAE) を用いる．予測評価値の総数を N とすると，式 (8) により算出される．

$$MAE = \frac{\sum_N |\hat{r}_{u,i} - r_{u,i}|}{N} \tag{8}$$

[1] http://www.grouplens.org/node/73
[2] http://www.imdb.com/

5.2 前提手法

CF の精度を検証するにあたり，前提手法としていくつかの指標から見た平均値で予測する．

AllAverage　全訓練データ 118,030 組の平均評価値を予測値とする．

UserAverage　予測データのユーザの評価値を予測値とする．一般に, 同じ評価値ばかり投稿するユーザに対してはこの手法でも良い精度が得られる．

ItemAverage　予測データのアイテムの平均評価値を予測値とする．推薦候補アイテムに対するユーザからの評価値にばらつきが小さいほど，予測精度は良くなる．

5.3 従来の CF と代理テキストを利用する手法

ユーザ間型 CF とアイテム間型 CF それぞれにおいて，以下に示す特徴量を用いて実装する．レビューテキストだけではなく，他の代理テキストを用いた手法と比較することで，それぞれの要素が CF における評価値の予測にどのように関わってくるのかを比較する．

Rating　評価値行列のベクトルを特徴ベクトルとする，従来の CF である．

Review　4 章に示した提案手法を用い，ユーザまたはアイテムに紐づく全てのレビューテキストを合わせた文書の bag-of-words で表す．レビューテキストには映画の内容を表す語と，ユーザの意見や感想を表す語が混在している．この手法では全ての語を索引語として採用する．

Subj　Review における文書の中からユーザの主観的な表現だけを抽出して索引語として用いる．感情語辞書として MPQA Subjectivity Lexicon[3] を利用した．これは感情極性値付きの単語リストで，今回は中間の感情極性値を持つ語を除いた全ての掲載語を感情語として採用した．

Summary　映画の内容を良く表しているあらすじ情報を用いる．IMDb の各映画に投稿される Plot Summary というユーザ投稿型のあらすじテキストを各映画ごとに集めて用いた．

MetaData　IMDb では，各映画についてジャンル，監督，脚本，出演者に関するメタデータが付与されている．このメタデータを各映画について抽出して文書と見なした．

5.4 結果と考察

実験結果を表 2 に示す．5.3 節における各手法を用いてユーザ間型 CF に適用したものには U. と，アイテム間型 CF に適用したものには I. とした．結果を見ると，評価値のみを用いた 3 つの前提手法と U.Rating と I.Rating を比較すると U.Rating が最も精度が良く，10 段階の評価値予測において MAE が 1.473 となった．I.Average が 2 番目に良く，I.Rating を上回った．一般に，ユーザ数に対してアイテム数が多いときは CF はうまく働かないという報告がある [6]．

表 2: 実験結果

	手法	MAE	t 検定 U.Subj	t 検定 I.Review
前提手法	AllAverage	1.741	***	***
	UserAverage	1.692	***	***
	ItemAverage	1.478	***	
ユーザ間型 CF	U.Rating	1.473	***	
	U.Review	1.473	***	
	U.Subj	1.467	-	
	U.Summary	1.476	***	
	U.MetaData	1.475	***	
アイテム間型 CF	I.Rating	1.517	***	***
	I.Review	1.468		-
	I.Subj	1.485		*
	I.Summary	1.504	*	***
	I.MetaData	1.475	*	*

代理テキストを利用した手法においては，ユーザ間型 CF では U.Subj が最も精度が良く，アイテム間型 CF では I.Review の精度が最も良かった．特にアイテム間型 CF においては，テキストを利用したいずれの手法も，評価値のみを利用した I.Rating と比べて精度が向上した．

ユーザ間型 CF とアイテム間型 CF のうちそれぞれ最も精度が良かった U.Subj と I.Review に注目し，この二つの手法と他の全ての手法を両側 t 検定により評価した結果を表 2 に示した．ここで，*は有意水準を表し，*は 5%以下，**は 1%以下，***は 0.1%以下で有意な差があったことを示す．空欄は有意な差がなかったことを示す．U.Subj は前提手法のいずれの手法とも有意水準 0.1%以下で有意な差があった．さらに，U.Subj は U.Review，U.Summary，U.MetaData に対しても有意な差があったので，ユーザを表すのには主観的な感情表現を表す語が有効だとわかった．I.Review はアイテム間型の他の全ての手法と有意な差があった．しかし，I.Average とユーザ間型の各手法と比べて有意な差は見られなかった．

5.5 索引語数を揃えた比較実験

代理テキストを用いる手法において文書を構成する索引語数の差による影響を除外するため，索引語の数を揃えて各手法の精度を比較した．各手法における文書の索引語を重みが大きい順に N 件まで採用した場合の比較について，N を 10 から 300 まで 10 ずつ変えた実験結果を図 1 と図 2 に示す．なお，特徴量を小さくしすぎると近傍が測れず予測ができない場合がおきる．その場合は従来手法の CF を用い，Rating により予測をした．

図 1 では，いずれの N においても U.Subj の精度が最も良かった．これにより，ユーザを表すには主観的な語を用いることが良いことがわかった．さらに，U.Subj では N を小さくするほど精度が上がる結果になった．図 2 では，N が大きくなるほど I.Review の精度が良くなっていった．N が小さい範囲では I.Review，I.Subj，I.Summary の精度が悪い結果になった．これにより，映

[3] http://www.cs.pitt.edu/mpqa/subj_lexicon.html

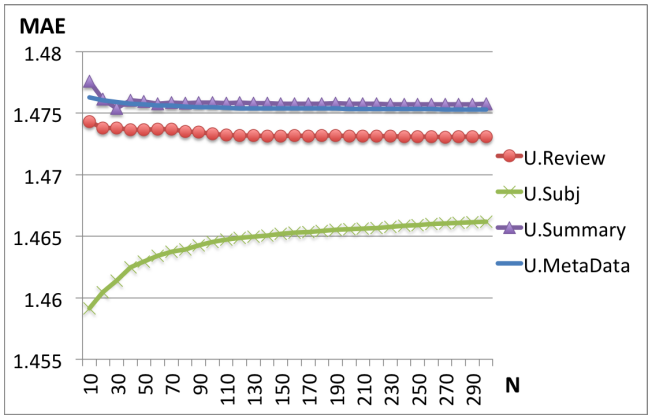

図 1: ユーザ間型 CF における索引語数ごとの MAE

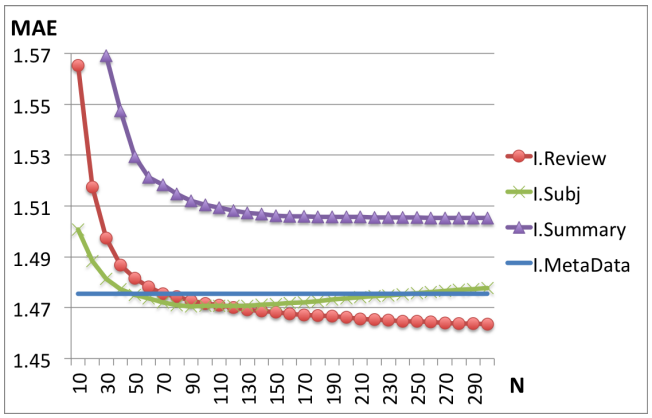

図 2: アイテム間型 CF における索引語ごとの MAE

画を表すには重みの大きい少数の語では精度が悪く，多くの語を必要とすることで精度が良くなることがわかる．I.Review はある所から I.Subj よりも精度が良くなる．このことから，レビューテキスト内のユーザの主観的な語だけでは不十分で，それ以外の語により精度が上がることがわかる．

6 おわりに

CF において，ユーザやアイテムをレビューテキストを用いて表す手法を提案し，その有効性を評価した．ユーザ間型 CF ではユーザの主観表現を，アイテム間型 CF ではレビューテキストの全ての語を用いた場合に顕著な精度の向上が見られた．今後の課題としては，索引語を出現頻度のみからではなく，レビュー著者の信頼度などを考慮した新たな重み付け手法が考えられる．

謝辞

本研究の一部は，科学研究費補助金基盤研究（B）（課題番号 22300050）によって実施された．

参考文献

[1] Breese, J. S., Heckerman, D., and Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering, *Proceedings of the 14th conference on Uncertainty in Arti cial Intelligence, Vol.461, pp. 43–52, 1998.*

[2] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms, *Proceedings of 10th International World Wide Web Conference, pp.285–295, 2001.*

[3] 中辻 真, 近藤 光正, 田中 明通, 内山 匡: アイテムに係る形容表現を用いたユーザ類似度測定, 人工知能学会全国大会論文集, *Vol. 24, 3C4-02, 2010.*

[4] Cane, W. L., Stephen, C. C., Fu-lai, C.: Integrating Collaborative Filtering and SentimentAnalysis, *Proceedings of the ECAI 2006 Workshop on Recommender Systems, pp. 62–66, 2006*

[5] Niklas, J., Stefan, H. W., Mark, C. M., Iryna, G.: Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations, *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, pp.57–64, 2009*

[6] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T.: Evaluating Collaborative Filtering Recommender Systems, *ACM Trans. on Information Systems, Vol. 22, No. 1, pp. 5–53, 2004.*