Vol.2016-MPS-107 No.15
2016/3/9

商品間のネットワーク構造を用いた意外性を高める推薦モデルの提案

河合未夢 [1]　能上慎也 [2]　保坂忠明 [2]　安藤晋 [2]

概要：本研究では購買履歴に基づいて商品間のネットワークを作成し，その中で周辺ノードとなる商品を優先的に推薦するモデルによって意外性のある推薦システムの実現を目指した．ネットワークにおけるハブの分析は長い歴史をもつが，ネットワーク形成の過程を理解することが主要な目的であり，本研究におけるようなネットワーク構造を工学的に利用する着想は新規なものとなった．本研究が対象とした推薦システムの問題では繰り返し推薦されるポピュラーな商品を避ける，購入した商品と間接的な関係を持つ商品を探索するといった部分問題があり，それらを扱うための表現形式として商品のネットワークを用いるのが有効であった．

キーワード：推薦システム，意外性を志向した推薦，Strength of weak ties

A Proposal of Recommendation Model to Enhance the Unexpectedness Using a Structure of an Item Network

MIMU KAWAI[1]　SHINYA NOGAMI[2]　TADAAKI HOSAKA[2]　SHIN ANDO[2]

Abstract: In this paper, we made an item network on the basis of the history of purchasing and we tried to realize serendipity-oriented recommendation system by using the model that preferentially recommends items which are marginal nodes. Analysis of a hub on networks had a long history and the main purpose is to understand a process of the frame of networks, the idea of using a structure of networks for engineering was unique. In problems of recommendation systems against our paper, there were partial problems. One is to avoid recommending popular items and the other is to search items which have indirect relationship among purchasing items. Using an item network for a flame of expression was effective against these problems.

Keywords: Recommendation System, Serendipity-oriented recommendation, Strength of weak ties

1. はじめに

1)　推薦システムは，利用者に有益と思われる情報や商品を推薦するシステムであり，オンラインストアなどで広く利用されている．典型的なシステムの入力は利用者の購買履歴データ，出力は購買者や閲覧者に推薦する商品である．一般的に推薦システムの評価には入力とは別のテストデータを用意し，テストデータと推薦商品の一致する度合を用いる．一方で，推薦商品の新規性や意外性といった観点も，利用者の満足に繋がることが指摘されている[1]．逆に似通った商品ばかりを推薦すると新規性や意外性が小さくなる[2]．購買データとの一致を高めることと，推薦商品の意外性を高めることにはトレードオフの関係があると指摘されており[3, 4]，実用的には利用者の好みとの一致を維持しつつ，意外性を高める方法が求められる．

推薦商品の意外性は個人の経験に依存するため，購買履歴データから直接推定することは難しい．そこで本研究では社会的ネットワークにおいて新規かつ有益な情報が入手できる経路に関する研究[9]に基づき，相関の高い商品間にリンクを張った商品ネットワークを考え，ネットワークの構造を用いた有益性と意外性の高い推薦を行うモデルを提案する．

本論文の構成は次の通り．2 章で提案モデルの準備として関連研究を紹介する．3 章で商品ネットワークと提案モデルについて説明する．4 章で分析対象の説明と数値実験の結果を示す．5 章で本論文のまとめと今後の課題について述べる．

2. 準備

2.1 基本的な推薦モデル

基本的な推薦モデルである商品ベース協調フィルタリング(IBCF)[1]と，非負値行列因子分解モデル(NMF)[11]について述べる．まず，推薦モデルの入力として購買履歴データを定める．期間Tについて，利用者の集合をU，全利用者数を$|U|$，商品の集合をI，全商品数を$|I|$とする．また，時刻$t \in T$で利用者$u \in U$が商品$i \in I$を購入した個数を$n(t,u,i)$とする．購買履歴データを以下のように表す．

$$\{ n(t,u,i) \mid t \in T,\ u \in U,\ i \in I \} \tag{1}$$

次に，期間Tにおける利用者uが商品iを購入した総数をb_{ui}とする．

$$b_{ui} = \sum_{t \in T} n(t,u,i) \tag{2}$$

購買行列Bはb_{ui}を要素とした行列とする．

1　東京理科大学経営学研究科経営学専攻
Department of Management, Graduate School of Tokyo University of Science, Kuki, Saitama,346-8512, Japan
2　東京理科大学経営学研究科経営学専攻
Department of Management, Tokyo University of Science, Kuki, Saitama,346-8512, Japan

ⓒ 2016 Information Processing Society of Japan

$$B = [b_{ui}]_{|U| \times |I|} \tag{3}$$

まず IBCF では，Bの列である商品の特徴ベクトルから商品間の類似度を定める．たとえば，利用者が購入した商品毎に，類似度が大きい順に定められたN個の商品を推薦する．類似度としては，確率変数の関係の強さを表す相関係数，ベクトルの方向の近さを表す余弦類似度，集合の共起の度合いを示すジャッカード係数などが用いられる．

次に NMF では，Bについて$|U|\times K$の行列$\acute{U}$と$K\times|I|$の行列$\acute{I}$を$\acute{U}\acute{I} \approx B$となるように学習する．ただし，$K$は購買行列の潜在次元数であり学習のパラメータである．利用者$x \in U$に対する商品$y \in I$の推薦スコアr_{xy}を，行列$\acute{U}$の第x行ベクトル$\boldsymbol{u_x}^{\mathrm{T}}$と行列$\acute{I}$の第$y$列ベクトル$\boldsymbol{i_y}$の内積とする．

$$r_{xy} = \boldsymbol{u_x}^{\mathrm{T}} \cdot \boldsymbol{i_y} \tag{4}$$

たとえば，利用者xに対して，推薦スコアの高い順に定められたN個の商品を推薦する．

基本的な推薦モデルの問題として，推薦した商品の結果を考慮していない点がある．この問題に対して，マルコフ決定過程を用いたモデル[13]が提案されている．

2.2 意外性向上のための推薦モデル

推薦商品の意外性を高めるために，推薦商品の多様性を高めるモデルが提案されている．文献[5]では，商品についてジャンル間の類似度を定義して，類似度の大きい異ジャンルの商品を推薦するモデルを提案している．文献[6]では，推薦商品の多様性を高めると利用者満足度が高くなると指摘している．

推薦商品の意外性を高める推薦モデルとしては，文献[3, 7, 8, 12]がある．文献[3]は，習慣モデルと嗜好モデルを導入して，2 つのモデルの推薦商品の差を意外性の高い商品と定義している．文献[7]は，ソーシャルメディアにおける利用者の行動履歴から隠れた興味の情報を推測して，これを意外性の高い情報と定義している．文献[8]は，推薦システムと利用者が対話的な操作を行い，推薦商品の意外性を高めるモデルを提案している．文献[12]は，推薦における個人性を向上させることで意外性を高めるモデルを提案している．

2.3 ネットワークを用いた情報推薦

文献[9]では，社会的ネットワークを用いて人同士の情報の伝搬を分析し，どこから新規で有益な情報を入手しやすいか研究している．社会的ネットワークは，人同士の強い関係と弱い関係で構成される．ネットワーク上で，強い関係から構成される割合が大きい密な部分グラフは，人同士に密な社会的関係があるので情報が共有化されている．自身がその密な部分グラフに属す場合，そこから得られる情報は既知の情報となりやすい．そのため，密な部分グラフ外の人との関係から得る情報の方が，新規で有益な情報となりやすい．

3. 提案モデル

3.1 提案モデル概要

推薦システムにおいて意外性が失われる要因として，1) 少数のポピュラーな商品が繰り返し推薦される，2)購入した商品と直接的な関係が高い商品のみが推薦される，といった点が考えられる．これを踏まえ，本論文は商品のネットワークを用いて意外性の高い商品を選択するモデルを提案する．ネットワークは相関の高い商品間にリンクを張ったもので，このようなネットワークでは重みの大きいリンクを多く持つハブ性が強い商品が存在する．ハブ性が強い商品は 1)のポピュラーな商品に相当するため意外性は低い．1)の問題に対処するため，ハブ性が弱いネットワーク上の周辺に位置する商品を推薦する．また 2) の問題に対処するため，ネットワーク上で相関が高い商品を辿ることで直接的な相関が高くない商品でも利用者の好みを考慮して推薦対象とする．次節以降は提案モデルついて具体的に説明する．

3.2 商品ネットワーク

商品をノードとする商品ネットワークを考える．ネットワークは重み付き隣接行列Aより表されるものとし，Aの要素a_{ij}は商品iとjの間にリンクがあるとき正の値，さもなくば 0 をとるものとする．さらに，リンクの有無は商品間の相関係数より定めるものとする．商品iとjの相関係数c_{ij}を購買行列B各々の商品に関する列から計算する．相関行列Cはc_{ij}を要素とした行列とする．

$$C = [c_{ij}]_{|I| \times |I|} \tag{5}$$

相関係数の閾値をthとし，th以上の商品間にリンクを張る．a_{ij}は，c_{ij}とtより以下のように定義される．

$$A = [a_{ij}] \tag{6}$$

$$a_{ij} = \begin{cases} 0 & \begin{array}{l} \text{if } i = j \\ \quad \vee\ c_{ij} < t \wedge c_{kj} \geq th\ \exists k \neq j \\ \qquad\qquad \wedge\ c_{ik} \geq th\ \exists k \neq i \end{array} \\ c_{ij} & \text{otherwise} \end{cases} \tag{7}$$

(7)の定義は，th以上の正の相関を持つ相手がない商品は，そのうち最大の相関を持つ相手とリンクを張るものとしているが，これは孤立ノードとなることを避けるための条件である．正の相関がない商品は孤立することになるが，本研究で扱ったデータでは存在しなかった．相関分析では経験的に相関係数が 0.2 未満な場合は正の相関なしと判断するため，本研究ではthを 0.2 とする．

3.3 商品間の類似度

提案モデル上で商品間の類似度を定めるために，ネットワーク上の商品間の最短経路を求める．経路の長さはAの対応する要素の逆数とする．最短経路の計算には Dijkstra 法を用いる．図 1 は，商品iからjまで辿るある 1 つの経路をP_{ij}={$(i,k),(k,l),(l,j)$}と表している．経路P_{ij}の長さ

は$\sum_{(v,w)\in P_{ij}}\frac{1}{a_{vw}}$とする．

商品間の類似度はネットワーク上の距離が遠いほど小さくなる． 3.1 節の 2)の対策として，商品間の類似度をネットワーク上の商品間の最短経路における商品間の相関係数の積とする．これにより，推薦の正確さを維持しつつ直接的には相関の低い商品を推薦するため，相関が高い商品を辿ったとき類似度が下がる割合が低くなる．商品iとjの類似度s_{ij}は以下になる．

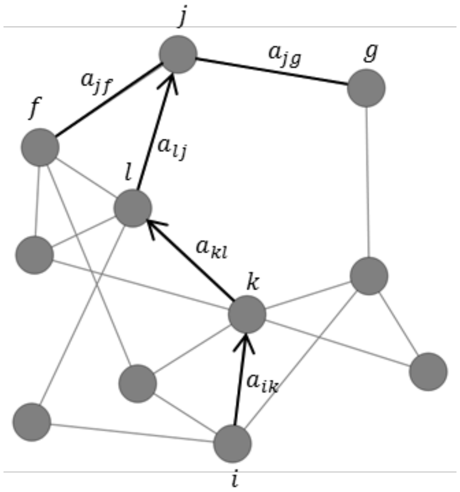

図 1 類似度とハブ度の算出方法
Figure 1 Calculation of similarity and hub

$$s_{ij}=\prod_{(v,w)\in P^{*}_{ij}} a_{vw} \quad (8)$$

ただし，P^{*}_{ij}を商品iとjの最短経路とする．s_{ij}は 0 から 1 までの値をとる．

3.4 商品のハブ度

3.1 節の 1)の対策として，ハブ性の数量的な尺度としてハブ度を定義する．提案手法では，相関の高い商品が多い商品をハブ性が強いと考え，商品jのハブ度h_jを，ネットワーク上で商品jに隣接している商品の相関係数の総和と定める．

$$h_j=\sum_{e\in N_j} a_{je} \quad (9)$$

ただし，図 1 より商品jに隣接している商品を$N_j=\{f,g,l\}$とする．h_jは 0 以上の値をとる．

3.5 推薦スコアの計算

提案モデルは類似度が高くハブ度の低い商品を推薦することを目指す．類似度とハブ度の線形和とした推薦スコアを求める．直感的には，ハブ度が低ければ直接的な相関が高くても推薦し，ハブ度が高ければ相関が高い商品を辿ってハブ度がより低い商品を探索することになる．商品iに対する商品jの推薦スコアr_{ij}は，類似度s_{ij}とハブ度h_jから

$$r_{ij}=s_{ij}+\lambda h_j \quad (10)$$

と定義する．ただし，λはデータから学習するパラメータであり，4.4 節で述べる目的関数の値が最大になるよう学習する．たとえば，商品iを購入した利用者に対して，推薦スコアの高い順に定められたN個の商品を推薦する．

3.6 評価指標

本節は，推薦の正確さを測る指標と意外性を測る指標について述べる．前者には推薦商品の新規性を考慮した適合率，再現率，F 値[1,2]などがあり，後者には文献[10]の unexpectedness がある．一般的な推薦モデルの推薦の正確さを測る場合，利用者の購買履歴データを，訓練データとテストデータに分ける．訓練データで利用者に対する推薦商品を決定し，テストデータで利用者が推薦商品を購入しているか検証する．

文献[1,2]に基づき，新規性を考慮した適合率，再現率，F 値を以下のように定める．まず，訓練データにおける利用者uの購入した商品の集合を$\Omega(u,TR)$とする．次に，利用者uに対する推薦商品の集合を$\Omega(u,R)$とする．最後に，テストデータで利用者uが購入した商品の集合を$\Omega(u,TST)$とする．適合率は，推薦商品に対するテストデータで購入した商品の積集合から，訓練データで購入した商品を除いた割合である．

$$\text{適合率}=\frac{1}{|U|}\sum_{u\in U}\frac{|((\Omega(u,R)\cap\Omega(u,TST))\text{¥}\Omega(u,TR)|}{|\,\Omega(u,R)\text{¥}\Omega(u,TR)\,|} \quad (11)$$

推薦商品を全て訓練データで購入している利用者について(11)式は計算できないが，そのような利用者は本研究で扱ったデータでは存在しなかった．

再現率は，テストデータで購入した商品に対する推薦商品の積集合から，訓練データで購入した商品を除いた割合である．

$$\text{再現率}=\frac{1}{|U|}\sum_{u\in U}\frac{|(\Omega(u,R)\cap\Omega(u,TST))\text{¥}\Omega(u,TR)|}{|\,\Omega(u,TST)\text{¥}\Omega(u,TR)\,|} \quad (12)$$

テストデータで購入した商品を全て訓練データで購入している利用者について(12)式は計算できないが，そのような利用者は本研究で扱ったデータでは存在しなかった．

F 値は，適合率と再現率の調和平均である．

$$\text{F 値}=\frac{2\,\text{適合率}\cdot\text{再現率}}{\text{適合率}+\text{再現率}} \quad (13)$$

文献[10]では，評価を行う対象モデルの推薦商品と，原始的な基準モデルの推薦商品との違いからunexpectedness を定義している．対象モデルと基準モデルの推薦商品の集合をそれぞれ$\Phi(u,R)$と$\acute{\Phi}(u,R)$とする．unexpectedness は基準モデルで推薦できなかった商品を対象モデルが推薦した割合であり，

$$\text{unexpectedness}=\frac{1}{|U|}\sum_{u\in U}\frac{|(\Phi(u,R)\cap\Omega(u,TST))\text{¥}(\acute{\Phi}(u,R)\cup\Omega(u,TR))|}{|\Phi(u,R)\text{¥}\Omega(u,TR)|} \quad (14)$$

となる．推薦商品を全て訓練データで購入している利用者

ⓒ 2016 Information Processing Society of Japan

について(14)式は計算できないが，そのような利用者は本研究で扱ったデータでは存在しなかった．本実験は NMF を基準モデル，提案モデルと IBCF を対象モデルとして両者の unexpectedness を比較する．

4. 数値実験

4.1 分析対象

本論文では，平成 26 年度データ解析コンペティション（http://jasmac-j.jimdo.com/データ解析コンペティション/平成 26 年度）において使用された，総合ブランド企業の購買履歴データ（データ 2）を対象として分析を行う．購買履歴データは年間 10 万円以上の購買を行っている優良顧客と，それ以外の顧客のものから構成される．本論文では，優良利用者の購買履歴データを対象とした．

購買履歴データは，2013 年 5 月 15 日から 2014 年 6 月 30 日まで約 1 年の期間について, 取引した商品の時刻別に 1 単位とした 1,503,727 件，全 11,030 利用者，52,703 商品から構成される．

本実験では，提案モデルの学習と評価のため購買履歴データを取引時刻順に 2 分割し，それぞれを訓練データとテストデータとする．さらに，訓練データを学習データと検証データに分割した．分割した購買履歴の概要を表 1 に示す．提案モデルおよびベースラインモデルの評価では訓練データの期間に購入した商品から推薦商品を決定し，テストデータとの一致の度合いを評価した．このため，学習データ，検証データ，テストデータの 3 つのデータ全てで購買がある利用者と商品のみを分析対象とした．

分析対象に含まれる利用者数・商品数は表 2 の「全部門データ」の欄に示した通りである．この中で高額の商品は購買頻度が低く種類が少ないため, 購買頻度の最も高い「菓子」部門の商品を抽出したデータを用意した．これを一部門データと呼び，全部門のデータと区別する．一部門データは提案モデルのネットワークを簡潔に可視化し，従来モデルとの違いを例示するのに用いる．表 2 はさらに一部門データの概要および両データから生成した商品ネットワークのリンク数を示す．

4.2 ベースラインモデル

提案モデルを客観的に評価するため，比較となるベースラインモデルを設定する．本実験では，商品ベース協調フィルタリング(IBCF)と非負値行列因子分解モデル(NMF)をベースラインモデルとする IBCF は類似度として相関係数を用いる．

4.3 推薦方法

典型的な推薦システムでは，利用者が購入した商品に対して推薦商品を提案するため，本実験においても，同様に提案モデルとベースラインモデルは，利用者が訓練データの期間で購入した商品毎に 2 つの商品を推薦するものとした．提案モデルは訓練データから商品ネットワークを作成し，訓練データを学習データと検証データを用いて 2-fold 交差検証を行い，最大の unexpectedness を与えるλを選択した．

表 1　分割データ
Table 1　Divided data

	訓練データ		テストデータ
	学習データ	検証データ	
取引件数	20,509	16,604	18,217

表 2　分析データ
Table 2　Analysis data

	部門データ	全部門データ
利用者数	1,238	2,701
商品数	231	1,409
リンク数	812	9,316

表 3　小規模データの評価指標の比較
Table 3　Comparison with an evaluation index in small data

	提案モデル	IBCF	NMF
適合率	6.4%	6.2%	5.3%
再現率	17%	16%	31%
F 値	9.4%	9.0%	9.1%
unexpectedness	0.45%	0.31%	

4.4 評価指標

本実験では推薦の正確さを測る指標に，推薦商品の新規性を考慮した適合率，再現率，F 値，意外性を測る指標に unexpectedness を用いる．unexpectedness は NMF を基準モデルに設定し，提案モデルと IBCF の値を比較する．また，提案モデルは意外性を評価することが目的のため，パラメータλは unexpectedness を目的関数として学習させる．

4.5 一部門データの実験結果

本節は一部門データの実験結果を示していく．λは-1 から 0 までの値を検証し，学習した結果-0.1 となった．λの値が負であることから，ハブ度が低い商品の方が推薦スコアは高くなっている．

4.5.1 評価指標の比較

評価指標について，提案モデルとベースラインモデルを比較したものを表 3 に示す．表 3 より，提案モデルは適合率，再現率，F 値，unexpectedness で IBCF を上回った．

4.5.2 ネットワークの特性分析

実験で得られたネットワークの全体図を図 2 に示す．ノード数は 231 個，リンクの数は 812 本，孤立ノードは存在しなかった．なお少数のノードから構成されるネットワー

クに属す商品は相関の高い商品がないため，一般的なコールドスタートにおける手続きにならいハブ度の最も低い商品を推薦するものとした．次に，ネットワークにおける各商品の隣接ノード数$|N_j|$のヒストグラムを図 3 に示す．図 3 より，大多数の商品は少ない隣接ノード数であるが，一部

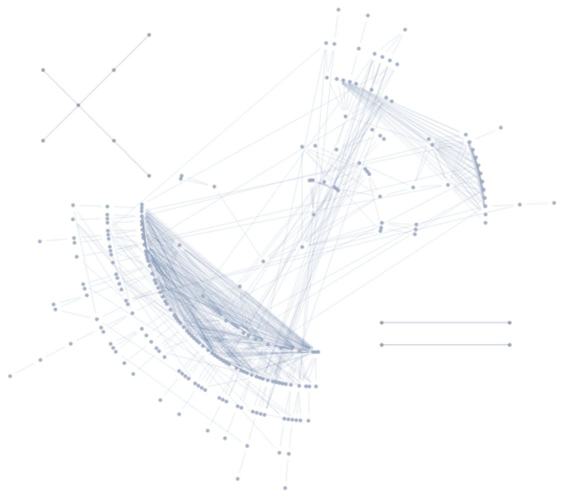

図 2　小規模データの商品ネットワーク
Figure 2　Item network of small data

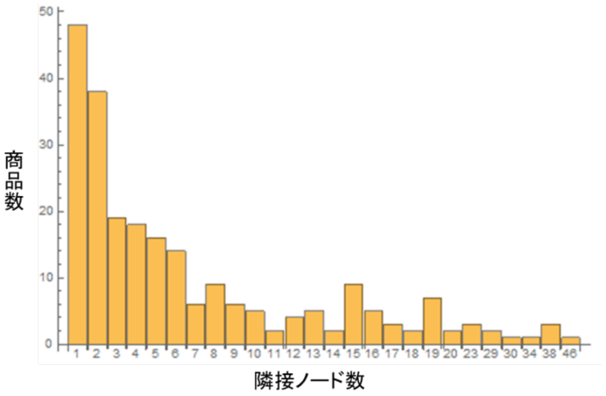

図 3　各商品の隣接ノード数のヒストグラム
Figure 3　Histogram of a number of adjacent nodes for every item

の商品は隣接ノード数が多くなっており，裾が広い分布となっている．選択した閾値thの値でハブ性の高い商品と周辺商品の違いがとらえられることが分かる．また，ネットワーク上の商品間の平均最短経路は 3.8 リンクであった．全商品間の類似度を計算するために，全商品間の最短経路を Dijkstra 法で求めたが，要した時間は 1.2 秒であった．計算機環境は CPU が Core i7@1.70Ghz でメモリが 8GB である．

4.5.3 推薦モデルの視覚化

4.3 節で述べた通り，商品毎に 2 つの商品を推薦するため一部門データでは 462 個の推薦商品がある．そのうち提案モデルと IBCF の推薦商品を比較すると，251 個が異なる商品であった．よって，提案モデルは IBCF とは大きく異なる商品を推薦していると言える．具体的に提案モデルがネットワーク上でどのような推薦を行っているか分析する．提案モデルと IBCF の推薦商品を比較した例を表 4 に示す．

表 4　推薦商品の比較
Table 4　Comparison with recommendation items

購入商品	推薦商品			
	提案モデル		IBCF	
シュークリーム	スイートポテト	スコーン	スイートポテト	カステラ
つまみ A	最中 A	最中 B	最中 A	つまみ B
バウム A	バウム B	大福	ブリュレ	バウム E
ブリュレ	バウム C	バウム D	バウムクーヘン	ロールケーキ

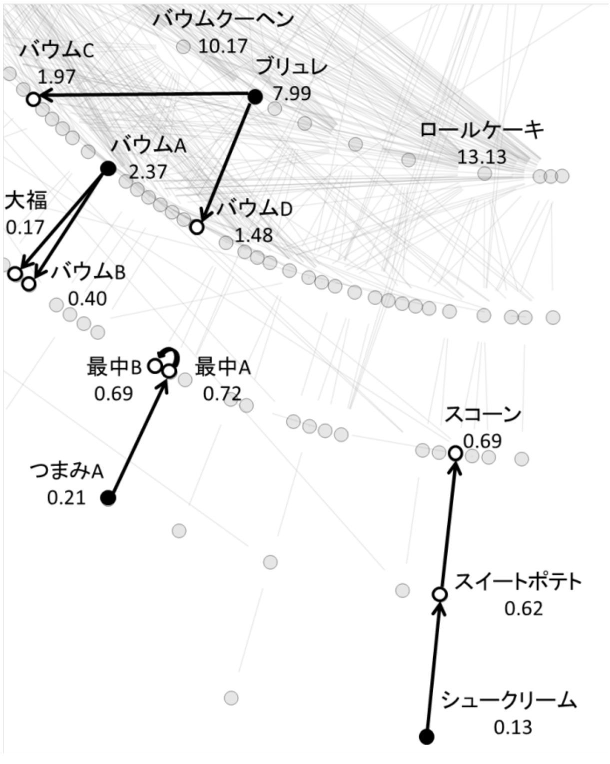

図 4　ネットワークの構造を用いた推薦
Figure 4　Recommendation to use a structure of networks

表 4 の商品について，提案モデルの推薦を表したものを図 4 に示す．商品名の下にはハブ度が表記してあり，末端ノードではハブ度が商品間の相関係数となる．購入商品●に対する推薦商品○に至るまでの経路を矢印で示している．

たとえば，ネットワーク上で末端ノードである”シュークリーム”は他の商品と正の相関はないが，その中で最も高い

ⓒ 2016 Information Processing Society of Japan

相関の”スイートポテト”とリンクがある．また，”スイートポテト”は”スコーン”と中程度の正の相関がある．提案モデルは”スイートポテト”を辿って”シュークリーム”から”スコーン”を推薦している．つまり，提案モデルは他の商品と正の相関が低い商品や少ない商品の場合でも，ネットワーク上の相関の高い商品を辿ることで，利用者の好みを考慮した推薦を行うことが可能であると考える．また，直接の相関が高くない商品を推薦しているので，意外性は高まると考える．

次に，他の商品と正の相関が多い”ブリュレ”について分析する．本来であれば表 4 の IBCF の推薦のように，”ブリュレ”は相関が高い上位 2 つの”バウムクーヘン”と”ロールケーキ”を推薦するのが一般的である．しかし，この両者はハブ度が大きいポピュラーな商品のため意外性の低い推薦と考えられる．提案モデルは推薦の意外性を高めるために，ハブ度の低い”バウム C” と”バウム D”を推薦している．ネットワーク上では周辺に位置する商品をより多く推薦することを意味する．

4.6 全部門データの実験結果

本節は全部門データの実験結果を示していく．λは-1 から 0 までの値を検証し-0.02 となった．全部門データでも一部門データと同様な実験結果がみられた．重複をさけるため必要な箇所だけ説明する．

評価指標について各推薦モデルを比較したものを表 5 に示す．一部門データ同様，評価指標は提案モデルが IBCF を上回った．次に，ネットワークの全体図を図 5 に示す．ノード数は 1,407 個，リンクの数は 9,316 本，孤立ノードは存在しなかった．少数のノードから構成されるネットワークに属す商品については，一部門データと同様にハブ度の最も低い商品を推薦するものとした．また，ネットワーク上の商品間の平均最短経路は 4.4 リンクであった．一部門データと全部門データの違いとして，商品間の類似度を計算するコストがある．4.5.2 節で示した同じ計算機環境で 567 秒を要した．

全部門データでは 2,814 個の推薦商品がある．そのうち，提案モデルと IBCF の推薦商品を比較すると 1,177 個が異なる商品であった．一部門データ同様，提案モデルは IBCF とは大きく異なる商品を推薦していると言える．また，提案モデルの推薦の特徴も一部門データと同じく，ネットワーク上の相関の高い商品を辿った推薦やネットワークの周辺に位置する商品をより多く推薦していた．

5. おわりに

本論文は，商品間のネットワーク構造を用いて意外性を評価するモデルを提案した．提案モデルは類似度が高くハブ度の低い商品の推薦を行い，推薦の正確さと意外性の向上を目的としている．本実験では，提案モデルは適合率，再現率，F 値，unexpectedness で IBCF を上回った．

提案モデルと IBCF の推薦商品を比較すると，両者は大きく異なる商品を推薦していた．それにも関わらず，提案モデルは IBCF と同等以上の評価指標であることは意外性を高める上での意義があると考えられる．

3.1 節で述べた意外性が失われる要因 1)と 2)について，1)はネットワーク上で周辺に位置する商品をより多く推薦することで，2)は相関の高い商品間を辿って推薦することで対策した．評価指標について提案モデルは IBCF を上回ったので，これらの対策は効果があったと考える．

表 5 大規模データの評価指標の比較
Table 5 Comparison with an evaluation index in large data

	提案モデル	IBCF	NMF
適合率	4.3%	4.1%	3.5%
再現率	11%	10%	19%
F 値	6.2%	5.9%	5.9%
unexpectedness	0.13%	0.12%	

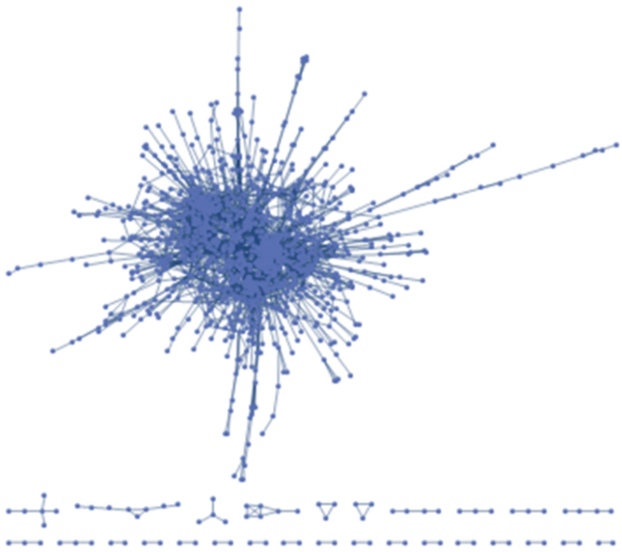

図 5 大規模データの商品ネットワーク
Figure 5 Item network of large data

今後の課題として，提案モデルとベースラインモデルとの意外性の差における,利用者満足度の検証が考えられる．また，提案モデルは推薦の仕組みを視覚化することができるので，商品の関係を分析する企業の担当者が，推薦の妥当性を評価・検証するツールとして利用できるように，提案モデルを改良していきたい．

謝辞 本研究のデータを与えてくださった企業様，ならびに経営科学系研究部会連合協議会の生田目崇様に深く感謝いたします．

ⓒ 2016 Information Processing Society of Japan

参考文献

[1] Jonathan L. Herloker, Joseph A. Konstan, Loren G. Terveen, John T. Ridel. Evaluating Collaborative Filtering Recommender Systems. ACM Transaction on Information Systems, 22(1): 5-53, 2004.

[2] 小川祐樹, 諏訪博彦, 太田敏澄, 山本仁志, 岡田勇. 共購買情報を用いた多様性のある推薦アルゴリズムの提案. 日本社会情報学会第 22 回全国大会研究発表論文集, 298-303, 2007.

[3] 村上知子, 森紘一郎, 折原良平. 推薦の意外性向上のための手法とその評価. 人工知能学会誌, 24(5): 428-436, 2009.

[4] Mouzhi Ge, Carla Delgado Battenfeld, Dietmer Jannach. Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity. Proceeding of the fourth ACM conference on Recommender Systems, 257-260, 2010.

[5] 杉本拓弥, 豊田哲也, 延原肇. 重み付き完全グラフに基づく異ジャンル間の嗜好傾向表現とそのレコメンデーションシステムへの応用. 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, 27(0): 237-240, 2011.

[6] Cai Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen. Improving Recommendation Lists Through Topic Diversifacation. Proceedings of the 14th international conference on World Wide Web, 22-32, 2005.

[7] 福島良典, 大澤幸生. ソーシャルメディアを利用したセレンディピティな情報推薦. 人工知能学会全国大会論文集, 26: 1-4, 2012.

[8] 奥健太, 服部文夫. セレンディピティ指向情報推薦のためのフュージョンベースアプローチのユーザ評価. 日本知能情報ファジィ学会誌, 25(1): 524-539, 2013.

[9] Mark S. Granovetter. The Strength of Weak Ties. American Journal of Sociology, 78(6): 1360-1380, 1973.

[10] 村上知子, 森紘一郎, 折原良平. 推薦結果の意外性を評価する指標の提案. 人工知能学会全国大会(第 21 回)論文集, 2C5-2, 2007.

[11] Daniel D. Lee, H. Sebastian Seung. Algorithms for Non-negative Matrix Factorization, In NIPS, 13: 556-562, 2000.

[12] 吉川大弘, 森貴章, 古橋武. Personalizability を考慮した推薦システムの提案. 情報処理学会論文誌 TOM, 6(1): 111-118, 2013.

[13] 桑田修平, 前田康成, 松嶋敏泰, 平澤茂一. 推薦システムのための状態遷移確率の構造を未知としたマルコフ課程. 情報処理学会誌 TOM, 6(1): 20-30, 2013.

ⓒ 2016 Information Processing Society of Japan