

商品の推薦理由がわかる情報推薦システム

岡本 一志

電気通信大学 大学院情報理工学研究科 情報学専攻 准教授

はじめに

商品や店舗, サービスなどの情報がインターネットを介して発信

- ・ 大量の商品の中から価値のある商品を見つける手助けが重要
- ・ 購買履歴や商品への評価といった行動ログを活用した意思決定支援

情報推薦システムのアプローチ

- ・ 内容ベースフィルタリング
 - ・ 商品の属性やユーザの属性などから推薦リストを作成
- ・ 協調フィルタリング (本研究の対象)
 - ・ 口コミによる推薦をモデル化
 - ・ 過去にユーザが商品に付与したスコアを用いて推薦リストを作成
 - ・ 内容ベースフィルタリングに比べ, 多様性を持った推薦リスト

研究背景

購買履歴や貸出履歴，商品へのレーティングなどのデータの活用

- ・ ユーザ/顧客による暗黙的・明示的な嗜好を表現する基礎的なデータ
- ・ ユーザ/顧客を行に，商品（アイテム）を列とする行列として扱うことが可能
- ・ ユーザ/顧客の数や商品数の増加に伴い次元数やスパース率も増加

センシングにより観測した位置や行動などもユーザ・アイテム行列として扱うことが可能
(推薦だけでなく，データ分析へも応用可)

ユーザ・アイテム行列の例
(4段階評価，空欄は未評価)

	商品A	商品B	商品C
ユーザ1	★★★★	★★★	★
ユーザ2	★	★	
ユーザ3	★★		★
ユーザ4	★★★	★★★★	
ユーザ5		★★★★	★★

ユーザが全てのアイテムを評価することは現実的ではなく，基本的にはスパースな行列

研究背景

協調フィルタリング

- ・ 情報推薦システムの実現手法のひとつ
- ・ 他者の評価を活用

当研究室でのアプローチ

- ・ Word2Vecのアイテムベクトル学習への応用
- ・ スパースなベクトルを対象とした類似度関数の開発
- ・ **推薦の透明性を有したモデルベース協調フィルタリング技術の開発**

ユーザ・アイテム行列の例
(4段階評価, 空欄は未評価)

	商品A	商品B	商品C
ユーザ1	★★★★	★★★	★
ユーザ2	★	★	
ユーザ3	★★		★
ユーザ4	★★★	★★★★	
ユーザ5		★★★★	★★

観測スコアから
未観測のスコアを予測

推薦の説明

「この商品はオススメです. なぜなら・・・」

「あなたと類似したユーザがこのアイテムを気に入っています」

推薦の説明 [N. Tintarev 07]



推薦結果が出てきたロジックを説明
(推薦システムがどう判断したか)

透明性は推薦アイテムの受け入れられやすさや信頼性, 満足度に寄与

[J. L. Herlocker 00, R. Sinha 02]

推薦の説明に関する既存研究

	タスク	手法	対象とする推薦の説明の役割							
			透明性	精密性	信頼性	有効性	説得力	効率性	満足度	
[R. Sinha 02]	Amazonなどの外部サービスを利用		○							
[M. Nilashi 16]			○		○					
[P. Symeonidis 08]	リスト予測	ハイブリッド型			○					
[F. Gedikli 14]		メモリベース法	○			○	○	○	○	
[J. L. Herlocker 00]	スコア予測	メモリベース法	○			○				○
[M. Bilgic 05]		ハイブリッド型				○				○
[D. Bridge 14]		メモリベース法					○			
[B. Abdollahi 17]		モデルベース法			○					
[M. Kaminskis 17]		メモリベース法					○			
[M. Fu 18]		メモリベース法	○							
[X. Wang 18]		ハイブリッド型	○							

協調フィルタリングにおける 推薦の透明性に関する研究

既存研究：メモリベース法が主流

当研究室での取り組み

- ・ 科研費（若手研究, 代表）
- ・ 回帰分析に基づき推薦理由を説明する情報推薦システムの開発
- ・ 2018～2020年度
- ・ モデルベース法でのアプローチ
 - ・ ユーザ・アイテム行列中の未観測スコアの予測と併せて、スコア算出理由を提示する技術の開発

	メモリベース法	モデルベース法
処理	ユーザ・アイテム行列を用いて予測	ユーザ・アイテム行列から得られた規則を用いて予測
長所	ユーザやアイテムに変更があってもモデル再構築が不要	規則に従った推薦処理のため高速
短所	推薦の都度、全てのデータ参照があり低速	ユーザやアイテムに変更があるとモデル再構築が必要
代表手法	k近傍法 <ul style="list-style-type: none"> ・ 利用者間型 ・ アイテム間型 	<ul style="list-style-type: none"> ・ 特異値分解 ・ 非負値行列因子分解 ・ Matrix Factorization

推薦の透明性を有した モデルベース協調フィルタリング

ユーザベクトル
(one-hot形式) $\mathbf{u} \in \{0, 1\}^m$

アイテムベクトル
(one-hot形式) $\mathbf{v} \in \{0, 1\}^n$

提案する回帰モデル

$$\hat{y} = c + \sum_{i=1}^m \alpha_i u_i + \sum_{j=1}^n \beta_j v_j + \sum_{i=1}^m \sum_{j=1}^n \gamma_{i,j} u_i v_j$$



入力が2つのone-hotベクトルのため

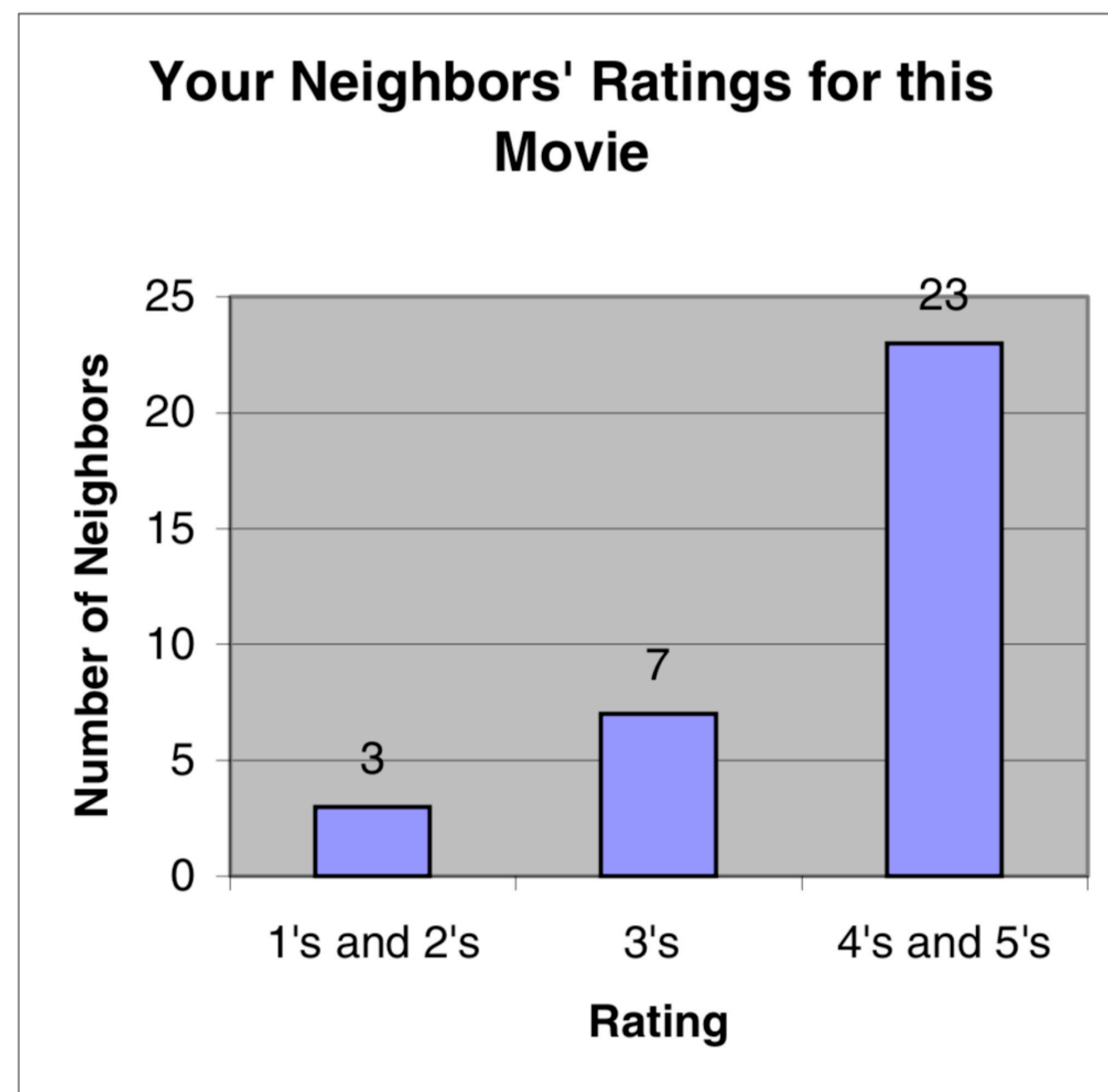
$$\hat{y} = c + \alpha_i + \beta_j + \gamma_{i,j}$$

推薦の透明性の実現方法： $c, \alpha_i, \beta_j, \gamma_{i,j}$ の提示

新技術の特徴・従来技術との比較

関連研究

類似ユーザの評価



[J. L. Herlocker 00]

類似ユーザの特定が必要
(計算コストが高い)

提案法

例

予測スコア (5点満点) : 4.1

- ・ユーザの影響 : 1.3
- ・アイテムの影響 : 0.9
- ・交互作用 : 0.7
- ・システムの影響 (バイアス) : 1.2

事前に学習したモデルに基づくため
推薦の説明の計算コスト削減が可能

図書推薦システムへの適用による評価

電気通信大学附属図書館の貸出履歴の利用

- ・ 倫理審査を経て提供を受けたデータ
- ・ 学生を学年と所属（学科・専攻）でユーザグループに分類

ユーザ グループ数	図書数	貸出履歴数		
		2015年度	2016年度	2017年度
36	19,186	17,935	16,859	15,917

入力：ユーザグループベクトルと図書ベクトル

- ・ ユーザグループベクトル：36次元のone-hotベクトル
- ・ 図書ベクトル：19,186次元のone-hotベクトル

出力：ある年度におけるユーザグループ*i*が図書*j*を借りる確率（スコア）

推薦精度の評価

比較対象（ベースライン）

- Factorization Machines (FM) [S. Rendle 12]

$$\hat{y} = c + \sum_{i=1}^m \alpha_i u_i + \sum_{i=1}^n \beta_i v_i + \sum_{i=1}^m \sum_{j=1}^n \langle \mathbf{w}_i, \mathbf{w}_j \rangle u_i v_j$$

- XGBoost [T. Chen 16]

データセットの分割

データセット	貸出履歴		
	2015年度	2016年度	2017年度
開発用	学習用データ	テストデータ	—
性能評価用	学習用データ		テストデータ

性能評価用データセットにおける推薦精度

	MAP@5	MAP@10	MAP@15	MAP@20
FM	0.286	0.250	0.226	0.208
XGBoost	0.279	0.232	0.204	0.186
提案法	0.321	0.262	0.228	0.208

評価指標：Mean Average Precision (MAP) @k

k件の推薦図書 R 適合図書 A $\text{precision}@k = \frac{|R \cap A|}{|R|}$

$$\text{average_precision}@k = \frac{1}{k} \sum_{i=1}^k \text{precision}@i$$

$$\text{mean_average_precision}@k = \frac{1}{g} \sum_{i=1}^g \text{average_precision}@i$$

推薦精度は
向上する傾向

性能評価用データセットにおける計算時間

	学習処理時間 [秒] (データ数：31,914件)	1スコアあたりの 予測処理時間 [秒]
FM	120	1.05×10^{-6}
XGBoost	17.0	1.33×10^{-5}
提案法	0.934	3.30×10^{-8}

計算マシン

- Intel Xeon Gold 6132 CPU @ 2.60GHz (14コア) ×2基
- メモリ196GB
- OS : Ubuntu 18.04.3, 実装 : Python + scikit-learn

学習処理, 予測
処理ともに高速

想定される用途

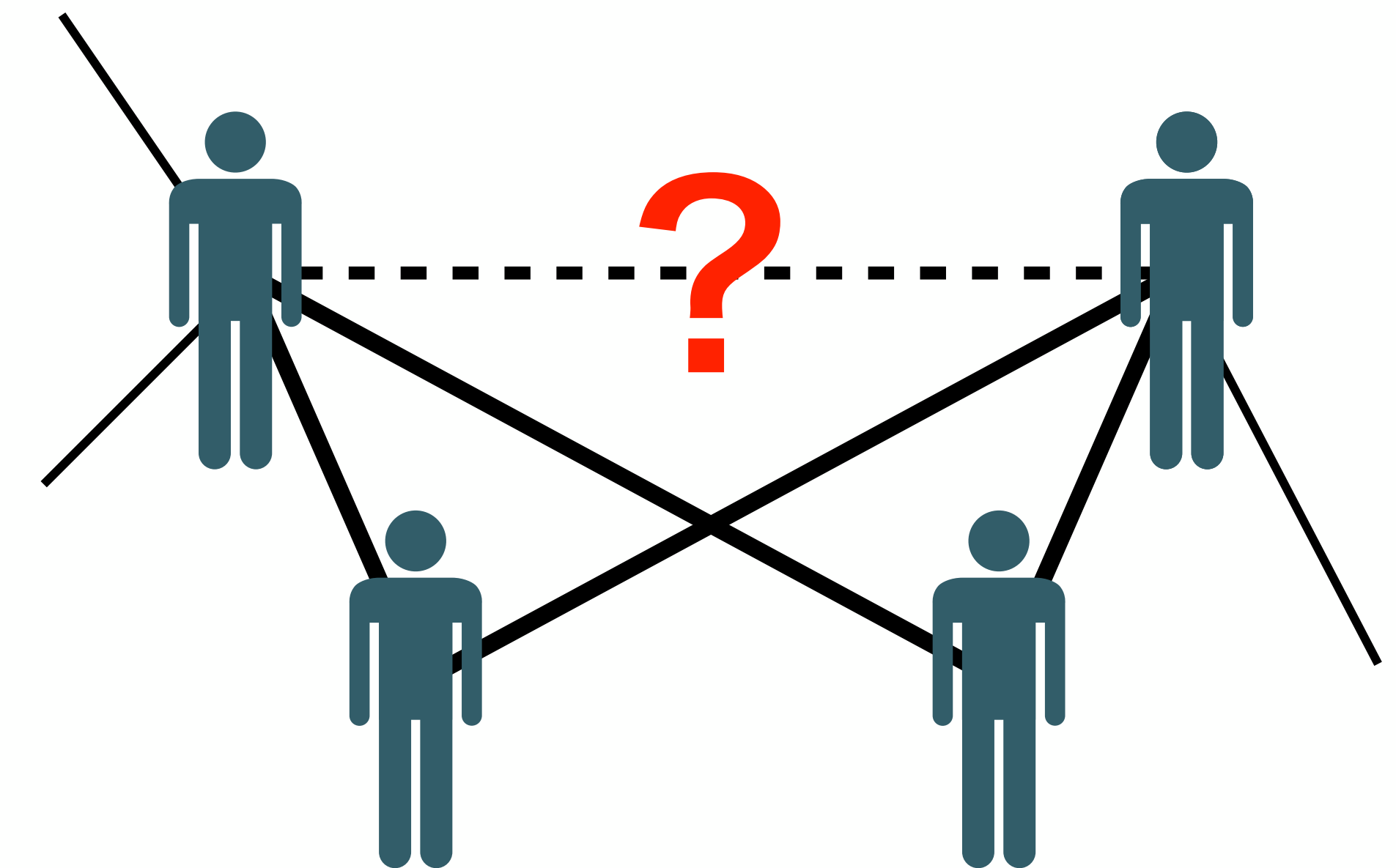
B to C

- ・ オンラインショッピング
- ・ 図書推薦システム（現在開発中）

B to B

- ・ ヒューマン・リソース・マッチング
- ・ 研究者推薦システム（現在開発中）
- ・ 行動ログを活用したマーケティング

予測に対する説明機能



実用化に向けた課題

スコアが付与されていない商品の推薦はできない

- ・ 他者の評価だけでなく，ユーザや商品の属性の活用も必要
- ・ 協調フィルタリング → ハイブリッド型フィルタリング

説明の言語化

- ・ 回帰係数の提示だけで理解してもらえるか？
- ・ 言語化による説明も重要

付与されたスコアは全て正しいという前提

- ・ スпамレビューなどが含まれていれば，期待されたとおりには動かない
- ・ 現在，スパムレビューの検出に関する研究を実施中

企業への期待

技術開発には実データが必要

- ・ 推薦精度に関する議論はベンチマークデータでの評価になりがち
- ・ 成功事例の記録に比べて、失敗事例の記録の取得は難しい
(研究費データベースだと、採択された課題のみ閲覧可能)
- ・ ユーザ/顧客の行動ログ (購買履歴や貸出履歴, 商品へのレーティングなど) のデータ活用を検討している企業との連携を希望

推薦システム (オススメする行為の自動化) の開発だけでなく, 機械学習システムのホワイトボックス化を検討している企業との連携も希望

おわりに

協調フィルタリングにおける推薦の透明性に関する研究

- ・ ユーザ・アイテム行列中の未観測スコアの予測と併せて、スコア算出理由を提示する技術の開発
- ・ 図書推薦システムへの適用による評価
- ・ 開発技術の特徴：高速かつ実装がシンプル
- ・ 今後の研究の方向性：説明インタフェースの開発

想定する応用例

- ・ オンラインショッピング / 記事推薦システム
- ・ ヒューマン・リソース・マッチング

参考文献

- [N. Tintarev 07] N. Tintarev: Explanations of Recommendations, Proc. of the 2007 ACM Conf. on Recommender Systems, 203-206, 2007
-
- [J. L. Herlocker 00] J. L. Herlocker, J. A. Konstan, J. Riedl: Explaining Collaborative Filtering Recommendations, Proc. of the 2000 ACM Conf. on Computer Supported Cooperative Work, 241-250, 2000
-
- [R. Sinha 02] R. Sinha, K. Swearingen: The Role of Transparency in Recommender System, Proc. of CHI '02 Extended Abstracts on Human Factors in Computing Systems, 830-831, 2002
-
- [M. Nilashi 16] M. Nilashi, D. Jannach, O. B. Ibrahim, M. D. Esfahani, H. Ahmadi: Recommendation Quality, Transparency, and Website Quality for Trust-building in Recommendation Agents, Electronic Commerce Research and Applications, 19(C), 70-84, 2016
-
- [P. Symeonidis 08] P. Symeonidis, A. Nanopoulos, Y. Manolopoulos: Providing Justifications in Recommender Systems, IEEE Trans on Systems, Man, and Cybernetics - Part A: Systems and Humans, 38(6), 1262-1272, 2008
-
- [F. Gedikli 14] F. Gedikli, D. Jannach, M. Ge: How Should I Explain? A Comparison of Different Explanation Types for Recommender Systems, Int. J. of Human-Computer Studies, 72(4), 367-382, 2014
-
- [M. Bilgic 05] M. Bilgic, R. Mooney: Explaining Recommendations: Satisfaction vs. Promotion, Proc. of Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research, 2005 Int. Conf. on Intelligent User Interfaces, 13-18, 2005
-
- [D. Bridge 14] D. Bridge, K. Dunleavy: If You Liked Herlocker et al.'s Explanations Paper, then You Might Like This Paper too, Proc. of Workshop on Interfaces and Human Decision Making for Recommender Systems, the 8th ACM Conf. on Recommender Systems, 2014
-
- [B. Abdollahi 17] B. Abdollahi, O. Nasraoui: Using Explainability for Constrained Matrix Factorization, Proc. of the 11th ACM Conf. on Recommender Systems, 79-83, 2017
-
- [M. Kaminskis 17] M. Kaminskis, F. Durao, Derek Bridge: Item-Based Explanations for User-Based Recommendations, Proc. of the 9th Int. Conf. on Information, Process, and Knowledge Management, 65-70, 2017
-
- [M. Fu 18] M. Fu, H. Qu, D. Moges, L. Lu: Attention based Collaborative Filtering, Neurocomputing, 31, 88-98, 2018
-
- [X. Wang 18] X. Wang, X. He, F. Feng, L. Nie, T. Chua: TEM: Tree-enhanced Embedding Model for Explainable Recommendation, Proc. of the 2018 World Wide Web Conference, 1543-1552, 2018
-
- [S. Rendle 12] Steffen Rendle: Factorization Machines with libFM, ACM Trans. on Intelligent Systems and Technology, 3(3), 57:1-57:22, 2012
-
- [T. Chen 16] T. Chen, C. Guestrin: XGBoost: A Scalable Tree Boosting System, Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 785-794, 2016

お問い合わせ先

国立大学法人電気通信大学 産学官連携センター
産学連携ワンストップサービス

TEL : 042-443-5871

FAX : 042-443-5725

E-mail : onestop@sangaku.uec.ac.jp