

レコメンデーション

情報科学概論 1
高間 康史 (1回目講義)

レコメンデーションとは

- Recommend … 推薦する
- 相手が望むもの・好むもの・必要としているもの・・・を薦める
- 情報推薦

レコメンデーションの例

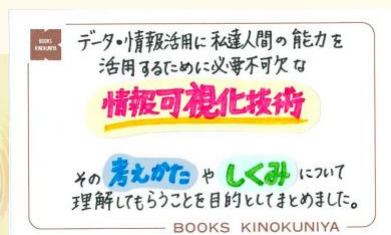
- オンラインショッピング
 - ベストセラー
 - 「この商品を買った人はこんな商品も買っています」



3

レコメンデーションの例

- 書店（実店舗）
 - 平積み・タワー積み
 - ポップ



4

アニメにも・・・

- あのアニメにも・・・
 - どこでも○○, ○○こんにゃく, etc.
 - 問題解決に役立つ道具の推薦



<https://doraega.com/2019/>

5

2種類のレコメンデーション

- ベストセラー, 平積み, ポップ
 - 誰に対しても同じものを推薦
 - 「この商品を買った人はこんな商品も買っています」
 - 相手に応じて異なるものを推薦
 - パーソナライゼーション (personalization)
- ➡現在主流

6

インターネット上の レコメンデーション

- オンラインショッピング
- インターネット広告
 - コンテンツ連動型広告：閲覧中のWebページの内容にマッチした広告を表示
 - 行動ターゲティング広告：ネット上での行動を解析・分析して対象ユーザを選別

7

インターネット上の レコメンデーション

- オンラインニュースサイト・アプリ
 - ニュースフィルタリング
 - キュレーションサービス（紙面を個人向けに作成）



<https://www.smartnews.com/ja/>



<https://gunosy.co.jp/>

8

人材推薦

- 採用支援：企業が求める技能・経歴を持つ人材の推薦
- お見合い
- ボランティアマッチングにも？

9

レコメンデーションの 仕組み

10

用語説明

- アイテム・・・推薦対象となるモノ
 - － 商品，書籍，ニュース，etc.
- 対象ユーザ・・・推薦を受けるヒト
- 嗜好・選好・・・好み，興味，関心：何が好きか？
- 特徴（属性）・・・アイテムやヒトの性質を表すもの

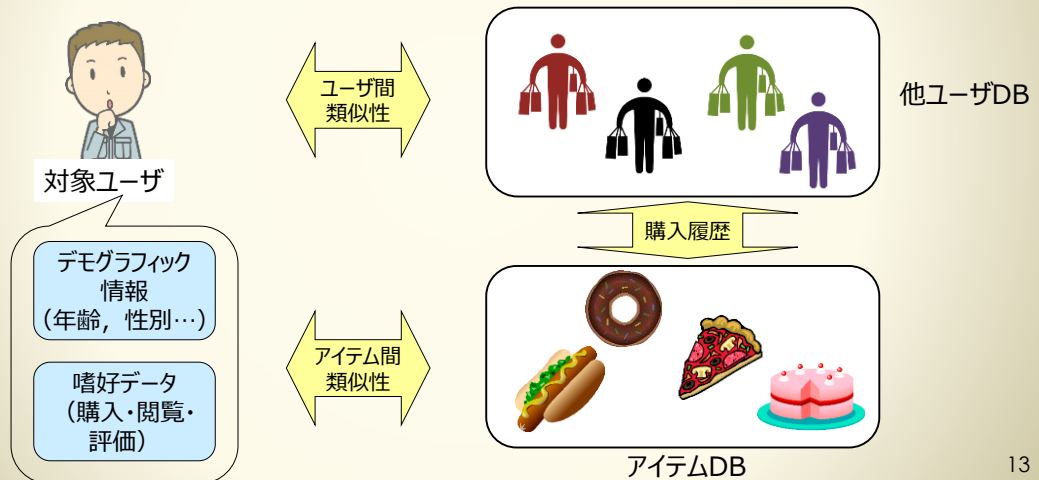
11

仮説：背景にある考え方

- 特徴が類似したユーザの嗜好は類似
 - － デモグラフィック情報の類似：
「30代男性に人気」，「関西の味付け」
- 行動が類似したユーザの嗜好は類似
 - － 過去に同じアイテムを購入・気に入った
- ユーザの嗜好は（ある程度）一貫
 - － 同じジャンルの映画，同じ歌手の曲を好む

12

レコメンデーションで利用する情報



13

内容ベースフィルタリング

- 代表的なレコメンデーション手法の一つ
- 考え方
 - ユーザが好むアイテムは、共通する特徴を持つ
- 特徴：2種類
- カタログ的な情報
 - 映画：ジャンル，主演，監督，etc.
 - 書籍：ジャンル，著者，出版社，etc.
- 内容的な情報
 - 書籍・ニュース：使用されている単語

- ▶ 特徴が類似したユーザの嗜好は類似
 - ▶ デモグラフィック情報の類似：「30代男性に人気」，「関西の味付け」
- ▶ 行動が類似したユーザの嗜好は類似
 - ▶ 過去に同じアイテムを購入・気に入った
- ▶ ユーザの嗜好は（ある程度）一貫
 - ▶ 同じジャンルの映画，同じ歌手の曲を好む

14

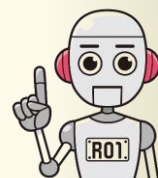
推薦アイテムの決定方法

1. 対象ユーザの嗜好を計算 ➡ ユーザプロフィール
– 特徴のベクトル (or集合)
2. データベース中のアイテムの特徴を計算
➡ アイテムモデル
3. ユーザプロフィールとアイテムモデルの
類似度を計算
– 類似度の高いアイテムを推薦

15

例題

店名	食事	座席	個室	飲み放題	喫煙
A	居酒屋	テーブル	あり	なし	可
B	居酒屋	座敷	あり	あり	可
C	イタリアン	テーブル	なし	なし	可
D	寿司	座敷	あり	なし	可



店名	食事	座席	個室	飲み放題	喫煙
X	居酒屋	テーブル	あり	なし	可
Y	寿司	座敷	なし	なし	不可

どちらを推薦すべき？

16

協調フィルタリング (Collaborative Filtering, CF)

- 内容ベースフィルタリング
 - 好むアイテムの特徴を利用
- 協調フィルタリング
 - アイテムの特徴は使わない
- 類似度はユーザの行動から計算
 - 同じユーザが好んだアイテムは類似
 - 同じアイテムを好んだユーザは類似

- ▶ 特徴が類似したユーザの嗜好は類似
 - ▶ デモグラフィック情報の類似：「30代男性に人気」、「関西の味付け」
- ▶ 行動が類似したユーザの嗜好は類似
 - ▶ 過去に同じアイテムを購入・気に入った
- ▶ ユーザの嗜好は（ある程度）一貫
 - ▶ 同じジャンルの映画、同じ歌手の曲を好む

17

評価値行列

- アルゴリズムで利用する情報
- 各ユーザのアイテムに対する評価：
 - 例. 1（嫌い）～5（好き）
- x: 欠損値（未評価）
 - スパース：疎な行列（多数のx）
 - 密度：x以外の要素の割合
 - 例. $15/30 = 0.5$
- 問題
 - 欠損値の予測 ➡ 予測評価値
 - 予測評価値の高いアイテムを推薦

ユーザ:
 U

4	x	2	1	x	x
4	x	x	x	4	x
x	3	x	2	2	x
x	x	2	5	x	4
x	x	1	4	3	4

アイテム: I

18

ユーザベースの協調フィルタリング

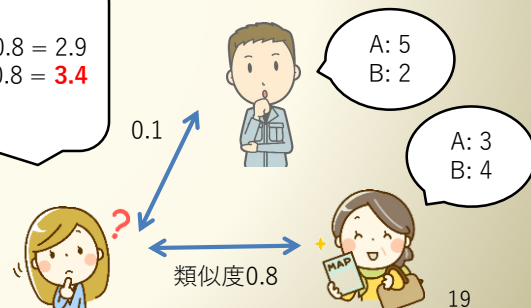
- 近傍ユーザの計算
 - ユーザ間類似度 = アイテムに対する評価の類似度に基づき計算
 - 相関係数やコサイン類似度を利用
- 予測評価値の計算
 - 推薦対象ユーザが未評価のアイテムについて評価値を予測
 - 加重平均

$$A \propto 5 \times 0.1 + 3 \times 0.8 = 2.9$$

$$B \propto 2 \times 0.1 + 4 \times 0.8 = \mathbf{3.4}$$

類似度

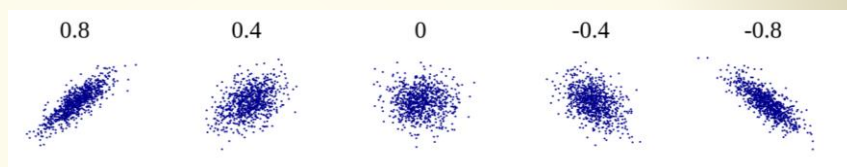
4	x	2	1	x	x
4	x	x	x	4	x
x	3	x	2	2	x
x	x	2	5	x	4
x	x	1	4	3	4



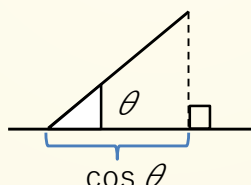
19

(参考) 相関係数とコサイン

- 相関係数：二つの変数の関係



- コサイン (余弦)



同じ方向
= 各要素の値が類似

20

協調フィルタリングの バリエーション

- 近傍ベース：類似ユーザの情報に基づく方法
- ユーザーベース
- アイテムベース
 - アイテム間類似度を事前に計算
 - 対象ユーザが過去に評価したアイテムに類似したアイテムを推薦

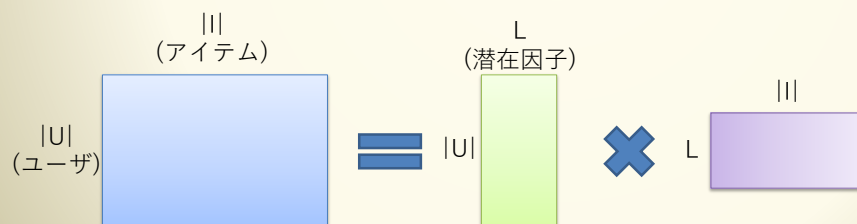
類似度

4	x	2	1	x	x
4	x	x	x	4	x
x	3	x	2	2	x
x	x	2	5	x	4
x	x	1	4	3	4

21

協調フィルタリングの バリエーション

- 行列分解ベース
 - 評価値行列 ($|U| \times |I|$) を
ユーザ行列 ($|U| \times L$), アイテム行列 ($|I| \times L$) に分解
 - 潜在因子: $L \ll |U|, |I|$. ユーザ, アイテムの本質的特徴?



22

レコメンデーションの課題

23

協調フィルタリングの魅力と課題

- 評価値行列：システム運用を続けると継続的に成長
 - 情報が増えるほど推薦精度が向上
- アイテムのカatalog・内容情報が不要
 - ➡ アイテムを問わず利用可能
- 課題：情報が少ない場合
 - 同じアイテムを評価したことがないユーザ
 - ➡ 類似度計算不可
 - 評価されたことがないアイテム ➡ 評価値予測不可

24

情報不足の問題

- コールドスタート (cold-start) 問題：新規ユーザー・アイテムに対しての評価情報が不足 → 推薦が困難
 - 例) 近傍ユーザーが不明
- スパースティ (sparsity) 問題
 - 一般的な密度：0.1 ~ 20%

ユーザー

4		2	4	
			4	
2			5	
	3			
	4		4	5

アイテム

類似度計算
不可能！

4		2	4	
	3	2	4	
2	3		5	4
	3		4	3
3	4		4	5

密な行列
68%

4		2		
			4	
			5	
	3			
	4			5

疎な行列
28%

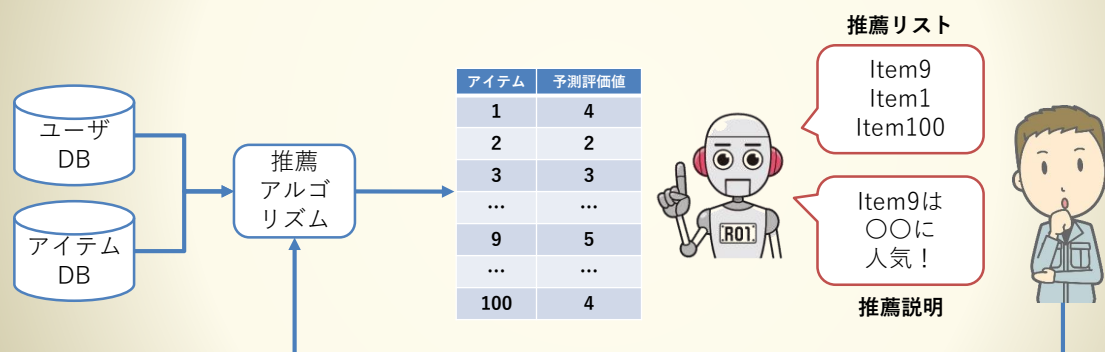
25

情報不足の問題

- 情報の独占が重要：例) ポイントカード
 - ← 利用者を識別しての情報収集が可能
 - 後発企業、中小企業に不利な展開
- 少ない情報で推薦可能な手法が必要不可欠
 - 情報推薦が適用可能な領域拡大にも貢献

26

情報推薦システムの全体像



27

評価の観点

- 推薦アルゴリズム：評価値の予測方法
- 評価指標
 - 予測精度：予測評価値と実際の評価値の差から計算
 - 平均絶対誤差 (Mean Absolute Error, MAE)
 - 平均二乗誤差 (Root Mean Square Error, RMSE)
 - 被覆率 (coverage)：全アイテムのうち、評価値の予測が可能なアイテムの割合
- 推薦システムとしての評価は？

28

どんなアイテムを推薦すべき？

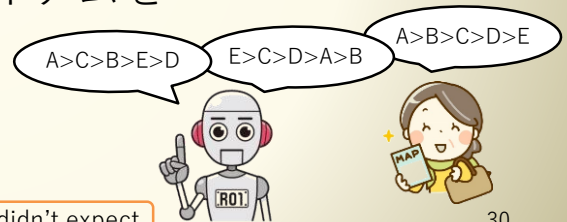
- 推薦システムとしての評価： Top-k推薦
 - （通常）予測評価値上位k件について推薦
- 評価の観点
 - ユーザが好むアイテムか？
 - ユーザにとって役立つか？
 - 受け入れてくれるか？



29

Top-k推薦の評価指標

- 適合率：推薦したアイテムの中に、どれだけ満足するものがあつたか？
- nDCG：ユーザが好む順序をどれだけ再現できたか？
- 多様性：様々なジャンルのアイテムを推薦できたか？
- 新規性：ユーザの知らないアイテムを推薦できたか？
- セレンディピティ：意外に感じるが気に入るアイテムを推薦できたか？



Serendipity = like + didn't expect

30

適合率, nDCG計算例

- 推薦順位上位3件を推薦した場合

- 評価4,5: 正解 (○) と仮定
- A○, B×, C○
⇒ 適合率 = 0.67 (2/3)

	A	B	C	D	E
ユーザ評価	4	2	5	2	4
推薦順位	1	2	3	4	5

- nDCG

- 理想の順位: C, A, E, B, D
 - $idealDCG = (2^5 - 1)/\log 2 + (2^4 - 1)/\log 3 + \dots + (2^2 - 1)/\log 6 = 50.4$
- 実際の推薦順位: A, B, C, D, E
 - $DCG = (2^4 - 1)/\log 2 + (2^2 - 1)/\log 3 + \dots + (2^4 - 1)/\log 6 = 39.5$
- $nDCG = 39.5/50.4 = 0.78$

$$DCG_k = \sum_{i=1}^k \frac{2^{r_i} - 1}{\log_2(i + 1)}$$

i番目アイテム
のユーザ評価

31

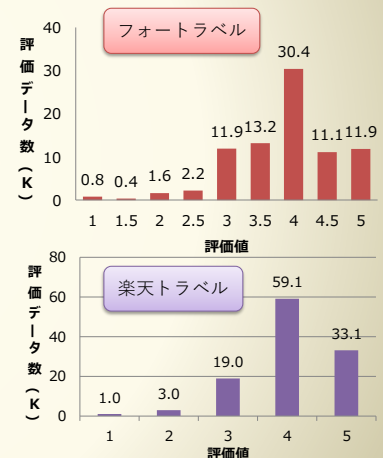
人気・評価の偏りの問題

- 人気アイテムの存在

- 無難な推薦
- 単純な手法で可能: 平均評価値,
人気ランキングなど
- そもそも推薦する意味は?

- 評価値の偏り

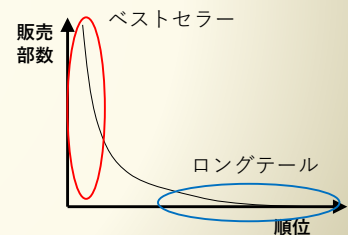
- 5段階の4あたりに集中: 予測が容易
- 低評価 (1,2), 高評価 (5) の予測は困難
- 低評価アイテムを正しく予測すること
の重要性
 - ロングテール (long-tail) アイテムの推薦
 - 受け入れ可能なアイテムの見極め



32

ロングテールとは

- ロングテール：売れないアイテムの集合
 - Amazon: 全売り上げの約1/3をロングテールから（※一時期）
⇔「80:20」の法則：80%の売り上げは20%の商品から
- ネットならではの現象？
 - 頭の面積<<しっぽの面積
 - 実店舗では困難（在庫管理コストなど）
- 見過ごされがちなロングテール商品を適切な顧客に推薦することで売り上げ増に貢献
- Win-Winの関係：顧客の満足度も向上

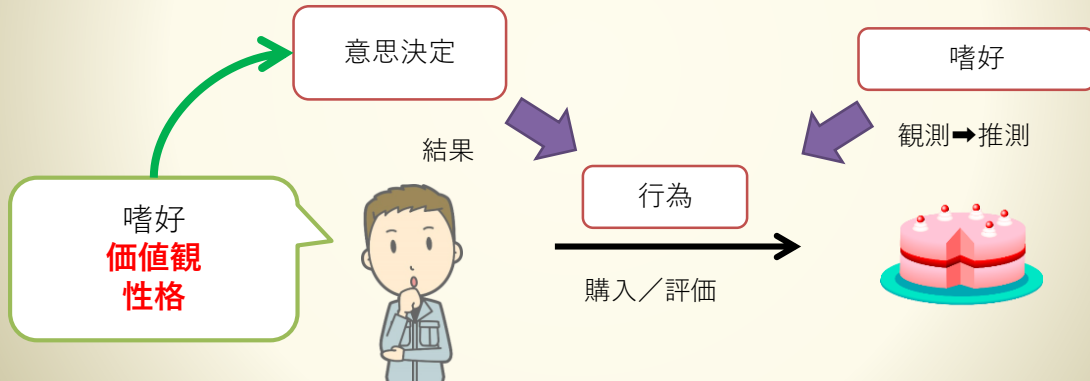


33

レコメンデーションと 価値観

34

ユーザ行動の背後にあるもの



35

性格（Personality）

- 個人の内面に由来する一貫した行動様式・プロセス
 - 先天的性質
- ビッグファイブ理論：パーソナリティの5因子
 - 経験への開放性：知的好奇心の強さ，想像力，新しいものへの興味
 - 勤勉性：自己統制力，目的達成への意志の強さ，計画性
 - 外向性：活動的である，上昇志向
 - 協調性：利他的な度合い，控えめさ，偽りのない態度
 - 情緒不安定性：不安や緊張の強さ，敏感さ

36

価値観 (Personal Values)

- 望ましい目標, 行動の選択基準
 - 後天的な性質
- Rockeachの価値体系
 - 最終価値 (18種類) : 豊かな生活, 平和な世界, 内面の調和, 社会的承認, 叡智, etc.
 - 手段価値 (18種類) : 野心・向上心, 寛容さ, 人の役に立つ, 愛情, 自制心, etc.
- Schwartzの価値体系
 - 自己超越: 博識, 善行
 - 保守: 調和, 伝統, 秩序
 - 変化に対する受容性: 自決, 刺激, 快楽
 - 自己高揚: 権勢, 達成, 快楽

37

価値観に基づくユーザモデルの課題

- 嗜好との距離
 - 野心的な利用者 → 何を推薦すべき?
 - 経験や勘に基づいた設定
- 汎用性: 多様なアイテムに対応できる

「方法論」

- 対象アイテムによらず共通して利用可能な「表現」 → 抽象的
- 対象アイテム依存の表現であっても, 同じ方法でモデルが作れることが望ましい → 方法論の共通化

「こだわり」
に着目

38

意思決定に見るこだわりの例



39

レビュー記事にみるこだわりの例

★★★★☆ あれ？, 2012/3/8
By [redacted] - レビューをすべて見る
レビュー対象商品: [redacted] [初回生産限定] [DVD] [DVD]
全然面白くない。(1票)
映像はお金かけてるだけあって見応えはあるけれど ストーリーは陳腐なもの。感動も何もない。
これは、ド派手な娯楽作品好きのファミリー向けですね。何も考えずに映像のすごさだけに驚ける人には良いでしょう。

★★★★★ 映像だけでも観る価値あり, 2012/4/24
By [redacted] - レビューをすべて見る
レビュー対象商品: [redacted] [初回生産限定] [DVD] [DVD]
この映画が収入記録を作ったと聞いた時には、「どうせ3Dの映像だけが特徴なんだろう」と思っていた。
しかし、自宅で3Dではない映像を観ても、その迫力には驚いた。とにかくよく出来ていて、飛行シーンの美しさは映画史に残るだろう。ストーリーにはそれほど感心したわけではないが、映像の力で充分に観客を魅了できる作品である。

個々の判断
(映像○, ストーリー×)
は同じなのに異なる評価
(★と★★★★★)



価値観・こだわりの
ちがいをモデル化

40

価値観に基づくユーザモデル

- 価値観 = 意思決定時にこだわりを持つ「属性」
 - 例. 「映画はストーリー重視」,
「(ホテルの) 水回りが汚いのが嫌」
- こだわりの強い属性が高評価 (低評価)
- 総合評価も高評価 (低評価)
- 評価一致率 (RMR) によるモデリング
 - 総合評価と「属性に対する評価」が一致する度合い
 - 高い属性 = 利用者の意思決定に強く影響 = こだわり
- ユーザモデル $U = \{RMR(a_1), \dots, RMR(a_n)\}$
 - a_i ... アイテムの持つ属性

※RMR ... Rate Matching Rate

41

評価一致率 (RMR) の計算例

レビュー1

属性	評価極性
総合評価	好評
ストーリー	好評 ✓
俳優	好評 ✓
音楽	不評

✓ 総合評価と極性が一致

レビュー2

属性	評価極性
総合評価	不評 ✓
ストーリー	不評
俳優	好評
音楽	好評

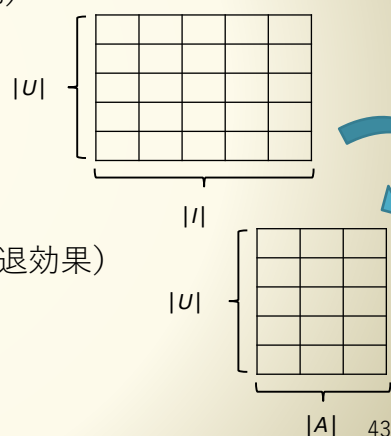
評価一致率

属性	ストーリー	俳優	音楽
一致回数	2	1	0
不一致回数	0	1	2
評価一致率	1.0	0.5	0.0

42

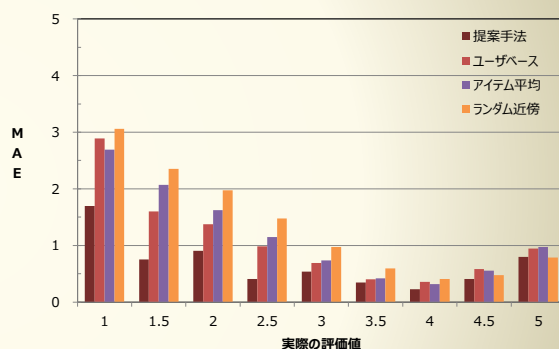
協調フィルタリングへの適用例

- 近傍ユーザの計算に適用
 - 通常：アイテムを要素とするベクトル ($|I|$ 次元)
 - 拡張：属性を要素とするベクトル ($|A|$ 次元)
- 期待される効果
 - 評価値の偏りに関する問題の軽減
 - (仮説) 極端な評価 (低評価or高評価) には、
こだわりが強く反映
 - 情報不足問題の軽減 → $|A| \ll |I|$ (次元縮退効果)
 - 近傍ユーザを見つけやすい
(同一アイテムを評価していなくてよい)



実験結果

- 対象データ：フォートラベル (ホテル)
 - ユーザ数：5,079,
 - ホテル数：7,295
 - 評価データ数：64,137 (スパース)
- MAEの結果
 - 平均的評価 (4近辺) はどの手法でも好結果 (低MAE)
 - 提案手法：特に低評価アイテムのMAEが低減

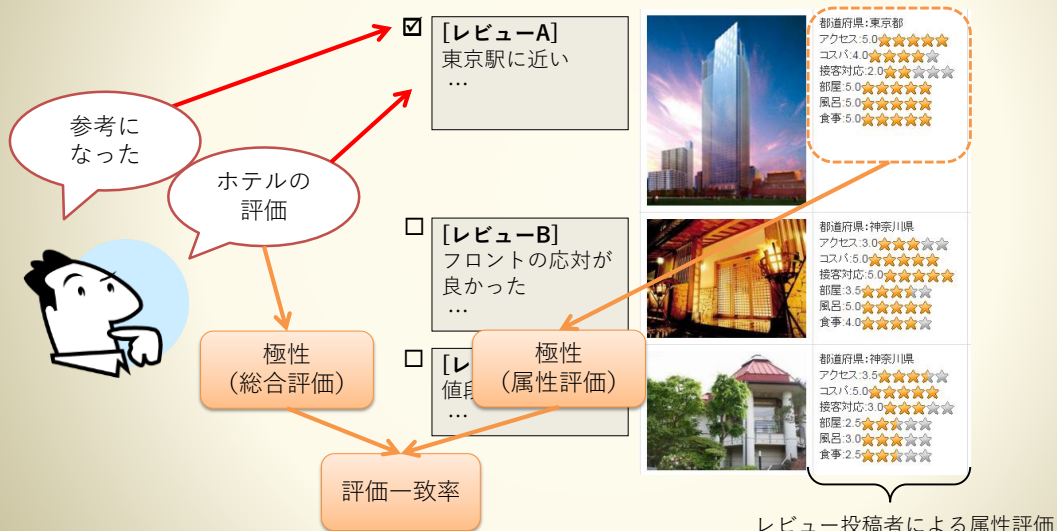


モデル構築方法の検討

- 投稿したレビュー記事からのモデリング
 - レビュー投稿者 << レビューサイト利用者
- 明示的なフィードバックの要求：属性別評価の入力
 - 余計な作業を強いる必要
- レビュー閲覧履歴からのモデリング
 - システムを利用しながら自然に獲得

45

レビュー閲覧履歴からのモデル構築



46

アイテムモデリング

- ユーザモデリング：特定ユーザのレビュー閲覧・投稿履歴から計算
 - ユーザのこだわりを推論
- アイテムモデリング：多数ユーザの評価結果（レビュー）から計算
 - そのアイテムが好まれる（嫌われる）ポイントを推論
- アイテムモデリングの利点：利用可能なレビューの多さ
 - より詳細な分析が可能：高（低）評価の時に好評（不評）の属性
- 想定用途
 - アイテムベース協調フィルタリング
 - 推薦理由の生成

騒音が気になる人は要注意

軽さにこだわる人に人気

47

レコメンデーションのこれから

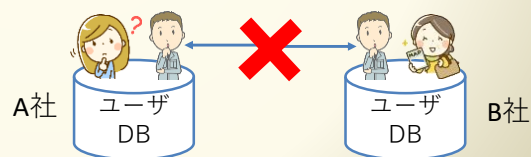
- 個人情報保護の問題
- 大きな意思決定の支援
- 脱「好きなモノの推薦」
- 社会問題解決への適用

48

個人情報保護の問題

- GDPR (EU一般データ保護規則)
 - 個人が自身のデータをコントロールする権利を取り戻す
- 個人データの「紐付け」が困難に
- 望まれる方法論
 - 大規模データ：匿名で活用
 - 対象ユーザの情報：少量からモデル構築

米Facebookは、Apple、Amazon、Netflix、Microsoftなどの米大手企業にFacebookユーザーの個人データを提供していたと、[米New York Times](#)が12月18日（現地時間）、入手したFacebookの内部文書と50人以上の元従業員やパートナー企業へのインタビューに基づいて報じた。(2018/12/20)



49

大きな意思決定の支援

- これまでの主な推薦対象アイテム
 - 映画，書籍，レストラン
 - ユーザにとって：低額，低リスク，経験機会豊富
- 不動産・就職・進学・結婚
 - 高額，高リスク，レアな体験
 - 推薦システムとユーザ間の知識のアンバランス
- 意思決定支援の必要性
 - 知識の提供
 - 推薦説明



50

脱「好きなモノの推薦」

- (例) 健康増進への応用

- レシピ・料理推薦

- 健康の観点：栄養バランスのよい食事
 - ユーザの好み：高カロリー・脂質の食事

- 散歩ルート推薦

- 健康の観点：目的カロリー消費が可能な経路
 - モチベーションの観点：飽きの発生

- 行動変容の支援

- 長期的な目標の達成：消費・摂取カロリー
 - 「好きなモノ」から「**受け入れ可能なモノ**」の推薦へ



51

個人から社会へ

- 個人の満足を維持しながら
コミュニティレベルの目標関数
を最大化

- マッチングの問題

- 観光（ホテル），お見合い
 - 全員に同じアイテムを推薦
できない：集中の回避

受け入れ・持続可能な省エネ活動
の推薦



こまめに消灯



ピークシフト



グリーン
カーテン



52