C10　摂動化によってプライバシーを保護した情報推薦方式

発表者 0BDRM039 望月 安菜
指導教員 菊池 浩明 教授

A Privacy-Preserving Recommendation Method using Perturbation

Abstract: This paper proposes a new perturbation method for privacy preserving in recommendation. The proposed method improves accuracy of recommendation based on the well-known item-based collaborative filtering with simplified item-item similarity defined by difference of rating values.

1　はじめに

情報推薦の主流は，複数のユーザによって複数のアイテムが評価付けされているデータベースにおいて，他のユーザの値を基に評価付けされていないアイテムの評価値を予測する協調フィルタリング (Collaborative Filtering) である．しかし，協調フィルタリングは複雑な類似度の計算を行わなければならないため，評価値の差分を類似度とし，アイテムベース推薦方式の摂動化 Slope One によって推薦値を求める方式を提案する．

2　既存手法

2.1　摂動化と再構築

再構築問題 (Reconstruction Problem) とは，摂動化された $Y = X + R$ から，真の値 X の確率分布を見積もる問題である．R. Agrawal and R. Srikant[1] によって最初に発表された摂動化アルゴリズムである．秘匿したい情報に意図的にランダムノイズを乗せて，格納されたデータのプライバシーを保護する．

暗号化による方法と異なり，時間のかかる暗号化はなく，通信効率も計算効率も高い．大規模なデータベースにおいても適用可能である．

オリジナルデータ X を Randomized Response によって摂動化を行い，偽データ Y を作成する．評価値の集合を $V = 1, 2, \cdots, v$，維持確率を p とする．評価値 x の摂動化 y は

$$y = \begin{cases} x & \text{確率 } p \\ a \in V & \text{確率 } 1-p \end{cases} \tag{1}$$

2.2　Slope One

D.Leniel and A. Maclachlan [2] によって提案された Slope One はアイテムベースの情報推薦アルゴリズムである．シンプルなアルゴリズムと高い性能で商用にも採用されている．Slope One とは，アイテム間の相関に傾き 1 の一次式 $f(x) = x + b$ を用いているところからその名が付いている．特異値分解 (SVD) などの既存の推薦方式と比較して，アイテム間平均差分に基づいて推薦を行うので実装も容易で処理性能も高い．

3　提案方式

3.1　概要

オリジナルデータ X の評価値行列 R^X を維持確率 p で Randmized Response によって摂動化を行い，偽データ Y の評価値行列 R^Y を作成する．この偽データ R^Y と p についてベイズ推定を行い，真のデータ X の再構築を行う．再構築の過程で得られた条件付き確率 $P(X|Y)$ を用いて，推薦精度を向上させることを試みる．

3.2　提案方式 - 摂動化 Slope One

Slope One を行う際に使用する共生起行列 $\Phi_{i,j}$ と平均差分行列 $\Delta_{i,j}$ を求める．偽データの評価値行列 R^Y のまま Slope One を行うと大きな誤差が起こるため，再構築を行い R^X に近似した R^Z によって，Slope One で情報推薦を行う．

3.2.1　共生起行列

維持確率 p より，欠損値数の予測を行う．摂動化を行って生成した行列 R^Y，維持確率 p よりオリジナルデータに期待欠損値数を予測する．n 個の要素を持つオリジナルデータ X の欠損値数を表す確率変数を K_X，摂動化した偽データで観測した欠損値数を表す確率変数を K_Y とする．摂動化を行うと欠損値を維持する確率は p，欠損値から評価値へと変化する確率は $1-p$ である．同様に評価値を維持する確率は $1 - \frac{1-p}{v}$ であり，評価値から欠損値へと変化する確率は $\frac{1-p}{v}$ である．これより，真の欠損値数 K_X の時に，偽データに K_Y 個の欠損値が生じる条件付き確率は

$$P(K_Y|K_X) = \sum_{j=0}^{K_X} \binom{K_X}{j} (1-p)^j p^{K_X - j} \binom{N-K_X}{K_Y - j} \left(1 - \frac{1-p}{v}\right)^{K_Y - j} \left(\frac{1-p}{v}\right)^{N - K_X - K_Y + j} \tag{2}$$

で与えられる．

事前確率 $P(K_Y|K_X)$ より，再構築アルゴリズムによりベイズ推定を行うことで事後確率 $P(K_X|K_Y)$ を求める．

$$P(K_X|K_Y) = \frac{P(K_Y|K_X)P(K_X)}{\sum_X P(K_Y|K_X)P(K_X)} \quad (3)$$

3.2.2 平均差分行列

摂動化評価値行列 R^Y と維持確率 p から定まる条件付き確率 $P(Y|X)$ より，摂動化 Slope One のための Δ_R を求める．偽データの Δ_Y より，再構築の再ベイズ推定で得られた条件付き確率 $P(X|Y)$ を用いて，ある列における差分の総和 Δ_Y から，新の差分総和 Δ_X を予測する．

$$P(\Delta_Y|\Delta_X) = \sum_{\Delta=\Delta_x} \sum_{\delta\in\Delta} P(\Delta_Y|\delta) \quad (4)$$

3.2.3 摂動化 Slope One

3.3 結果

オリジナルデータ，摂動化を行った偽データ，提案手法である再構築データのそれぞれで Slope One を行い予測した値の MAE(Mean Absolute Error) と，協調フィルタリングによって予測した値の MAE との比較を Table 1 に示す．$MAE^Z = \sum_{u,i} |r^X_{u,i} - r^Z_{u,i}|$ と定める．摂動化したデータにナイーブに協調フィルタリングを適用したものは，$MAE^Y = \sum_{u,i} |r^X_{u,i} - r^Y_{u,i}|$ である．偽データより提案手法である再構築データの方が誤差が少ない．

3.4 考察

Fig.1 の散布図は，Slope One の推薦値を真のデータした時の，摂動化データのみによる推薦値と提案手法による推薦を図示している．Slope One を行った際，再構築データはオリジナルデータに近づいている．これは Table 1 からも分かるように，偽データより再構築データも方が誤差が小さいことが分かる．また，協調フィルタリングについてもゴさを測り Slope One との比較を行った．全てのデータにおいて Slope One の方が誤差が少ない．これら 2 点より，Slope One は，協調フィルタリングより精度が高く，摂動化との相性の良さが分かる．しかし，本結果で用いた評価値行列は，オリジナルの推薦値の誤差が大きいことから人工的で歪んでいた可能性がある．

4 おわりに

プライバシーを保護したまま，摂動化したデータから再構築を行う新しい情報推薦方式を提案した．提案方式は，摂動化により一定確率で評価値をランダマイズされたデータができるためプライバシーが保護される．再構築によって，オリジナルデータへ近似させることで，情報推薦の精度が向上することを示した．情報推薦の主流は，協調フィルタリングであったが，Slope One を使って評価値を予測することで，より誤差の少ない推薦を行うことが期待できる．

Table 1: Mean Absolute Error

	CF	Slope One
Original	0.97	0.73
Disguised	1.03	0.89
Proposed	1.01	0.86

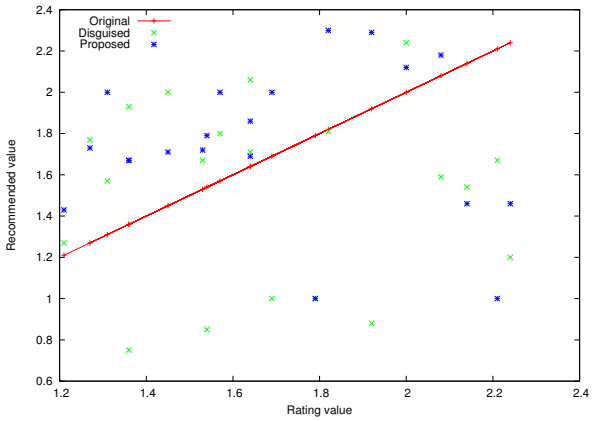

Fig. 1: Slope One による推薦値の分布

参考文献

[1] R. Agrawal and R. Srikant, “Privacy-Preserving Data Mining”, ACM SIGMOD 2000, pp. 439-450, 2000.

[2] D. Leniel and A. Maclachlan, “Slope One Predictors for Online Rating-Based Collaborative Filtering”, Society for Industrial Mathematics, pp. 1-5, 2005.

業績リスト

1. 望月, 菊池, “摂動化によってプライバシーを保護した情報推薦方式”, マルチメディア, 分散, 協調とモバイル (DICOMO2011) シンポジウム, 3G2, pp. 536-541, 2011, 優秀論文賞受賞.
2. 望月, 菊池, “Slope One を用いた摂動化プライバシー保護情報推薦方式”, コンピュータセキュリティシンポジウム 2011(CSS2011), 2D2-3, pp. 379-389, 2011.
3. 望月, 菊池, “アイテム依存の摂動化によるプライバシー保護情報推薦”, 暗号と情報セキュリティシンポジウム (SCIS2012), 2D1-6, pp. 1-5, 2012.