摂動化によってプライバシーを保護した情報推薦方式

†望月 安菜 †菊池 浩明

†東海大学

1. 情報推薦

Apple MacBook 2.4GHz Core 2 Duo/13.3"/2G/250G/8xSuperDrive/Gigabit/802.11n/BT/Mini DisplayPort MC516J/A

アップル

★★★★★ (7件のカスタマーレビュー) いいね (3)

7レビュー

星5つ:	(7)
星4つ:	(0)
星3つ:	(0)
星2つ:	(0)
星1つ:	(0)

おすすめ度

★★★★★ (7件のカスタマーレビュー)

あなたの意見や感想を教えてください

自分のレビューを作成する

これにも注目

最近チェックした商品 | 同じテーマの商品

Apple MacBook 2.4GHz Core 2 Duo/13.3...
アップル
~~¥ 94,800~~ ¥ 82,248

Apple MacBook Pro 2.4GHz Core 2 Duo...
アップル
~~¥ 114,800~~ ¥ 104,167

Apple MacBook Air 1.4GHz Core 2 Duo...
アップル
~~¥ 88,800~~ ¥ 80,500

Apple MacBook Pro 2.66GHz Core 2 Duo...
アップル
~~¥ 138,800~~ ¥ 109,000

› 表示履歴を管理する

2. 目的

❖情報推薦の問題点

- 評価値のプライバシー
- 暗号化：計算コストがかかる

❖研究目的：摂動化を利用した新しい情報推薦方式の提案

- ランダム化のみによる高速処理

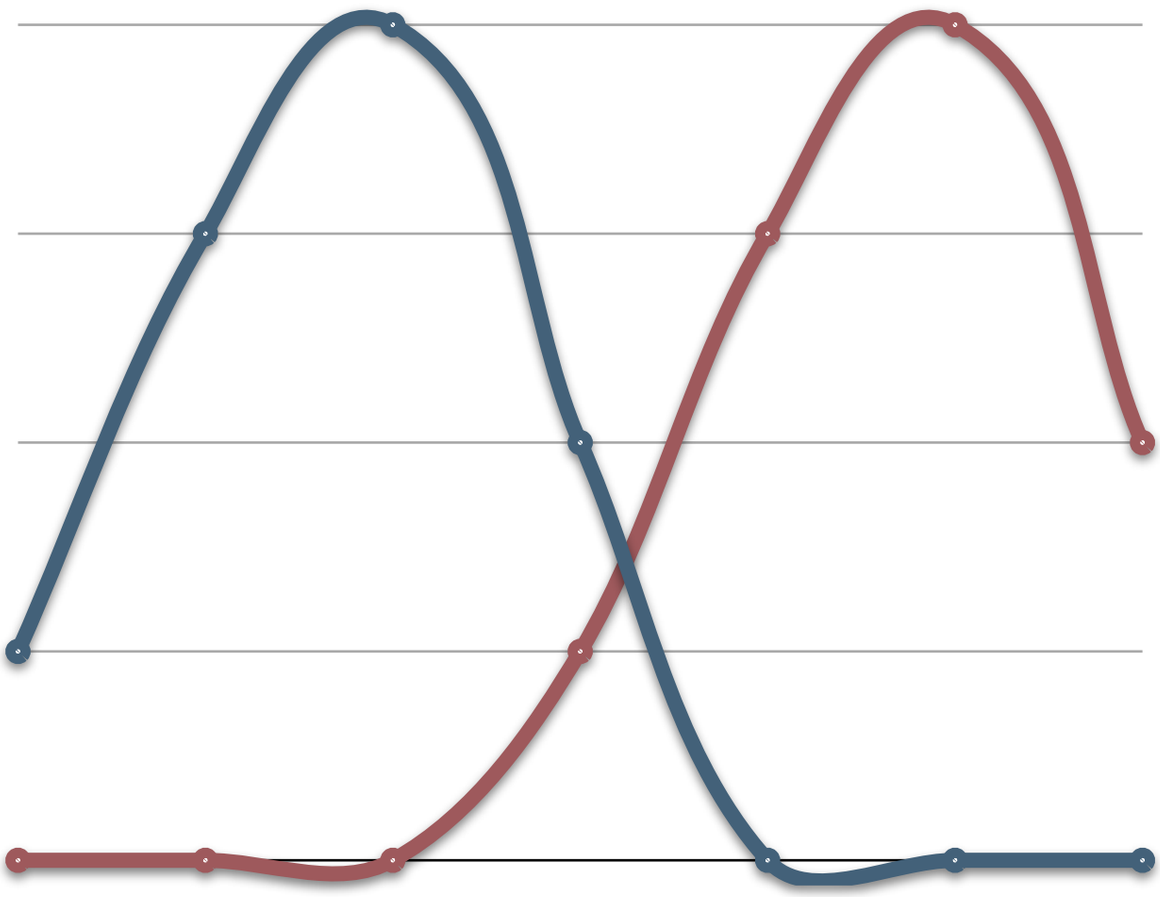
3. 従来研究

	H. Polat , W. Du 2003[3]	提案手法
ランダム化 プライバシー 協調フィルタリング	加法摂動化 PCAを用いた 攻撃あり [2] 容易	回答ランダム化 （目標）

3.1 加法摂動化（Polat and Du, 2003）

オリジナル X	乱数 R	摂動化 Y
3	0	3
1	1	2
2	2	4
0	2	3

$$Y = X + R$$



3.1.1 加法摂動化による推薦（Polat and Du, 2003）

	i_1	i_2	i_3	i_4	i_5
u_1	2	2	3	1	
u_2	1	3	2		3
u_3	2		3	3	2
u_4	3	2	*	2	2

R →

	i_1	i_2	i_3	i_4	i_5
	2	3	1	1	0
	1	1	2	0	1
	1	0	3	3	0

→ CF

＊＝２ (推薦値)

類似度

	i_1	i_2	i_3	i_4	i_5
i_1	-	0.7	0.9	0.8	0.7
i_2		-	0.7	0.3	0.7
i_3			-	0.8	0.7
i_4				-	0.6
i_5					-

3.2 本提案

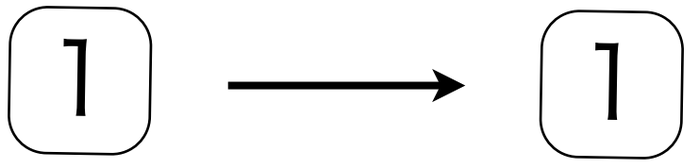
❖概要

1. 加法摂動化の代わりに回答ランダム化（安全性）
2. ベイズ推定に基づいて再構築した確率分布による協調フィルタリング（精度の向上）

3.1.2 Randomized Response

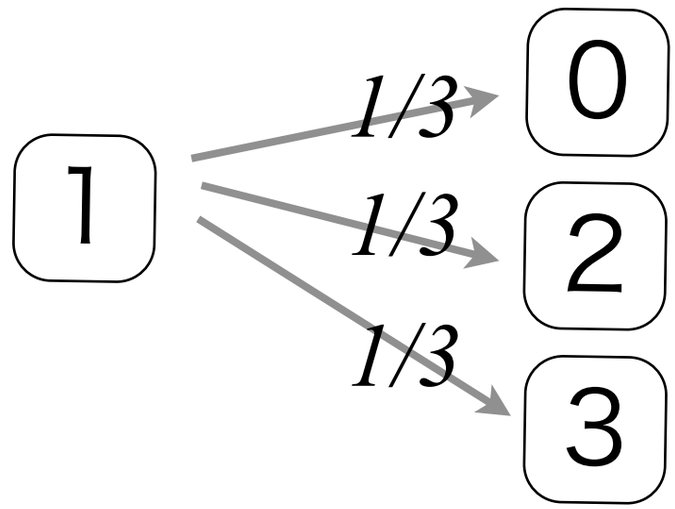
❖ p の確率で維持

$p = 0.4$



❖ $1\text{-}p$ の確率で撹乱

$1\text{-}p = 0.6$



3.1.2 Randomized Response

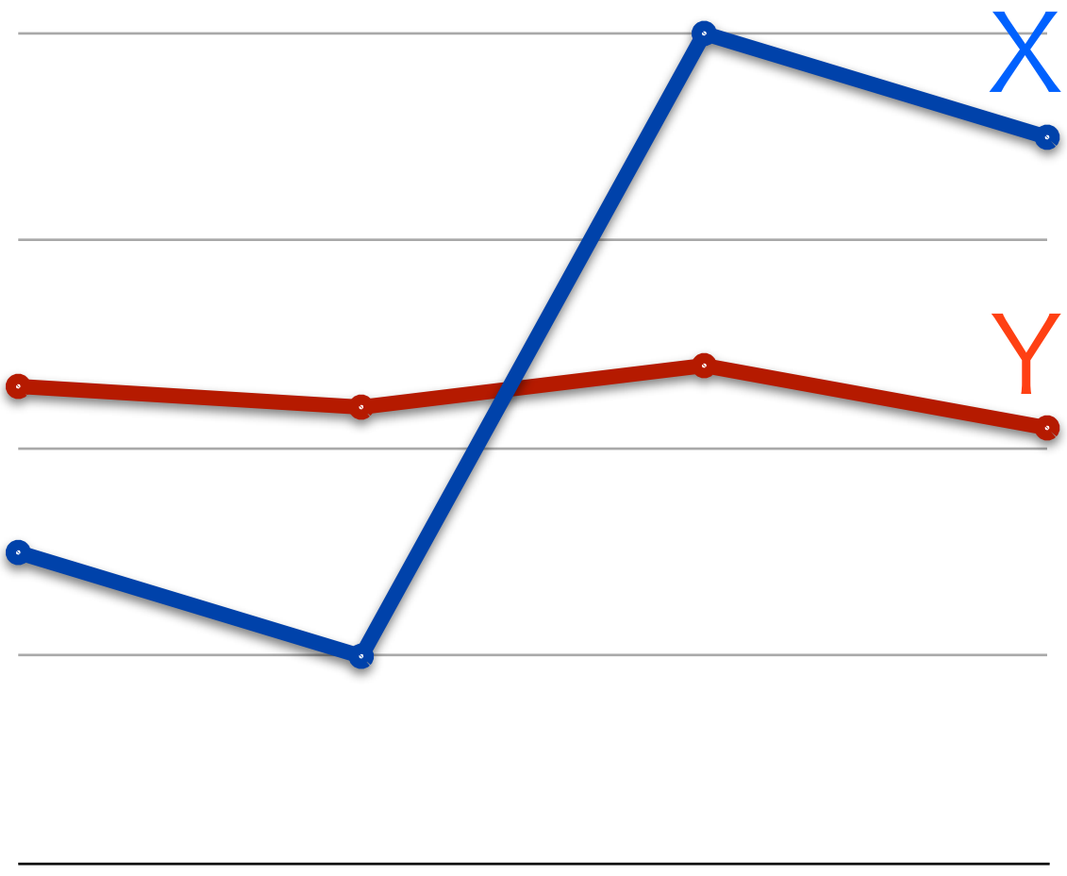
オリジナルデータの確率分布：X

x	0	1	2	3
P(X)	0.1	0.3	0.1	0.5

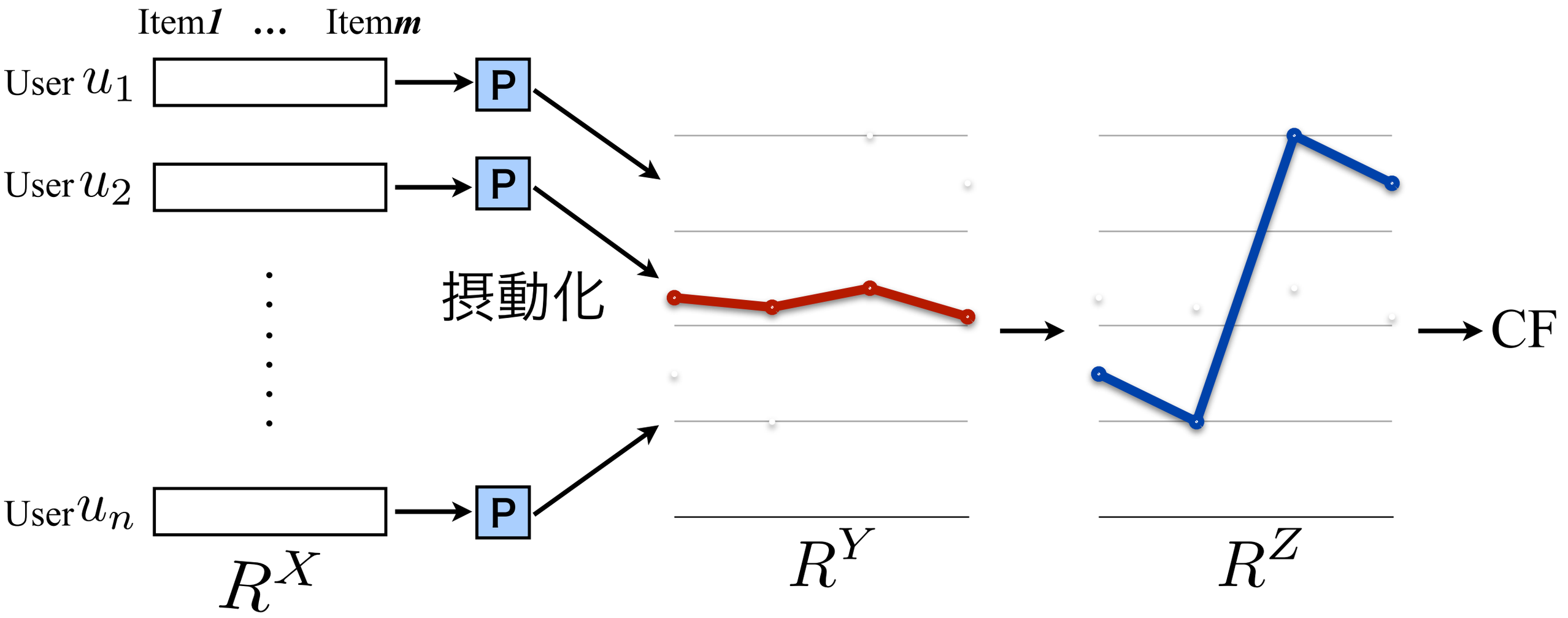
$p = 0.4$

偽データの確率分布：Y

P(Y)	0.22	0.26	0.22	0.3



4. ベイズ再構築の原理



評価値行列

X	i_1	i_2	i_3	i_4	i_5
u_1	2	2	3	1	0
u_2	1	3	2	0	3
u_3	2	0	3	3	2
u_4	3	2	＊	2	2

オリジナルデータ

摂動化

Y	i_1	i_2	i_3	i_4	i_5
u_1	2	3	1	1	0
u_2	1	1	2	0	1
u_3	1	0	3	3	0
u_4	3	2	＊	2	3

偽データ

3.2 ベイズ推定

❖ ベイズ推定

a	0	1	2	3
X	0.1	0.3	0.1	0.5
P(Y)	0.22	0.26	0.22	0.3
P^1(X)	0.21	0.26	0.21	0.33
⋮				
P^{150}(X)	0.101	0.299	0.101	0.499

$$P^i(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$= \frac{P(B|A)P^{i-1}(A)}{\sum_{a \in A} P(B|A=a)P^{i-1}(A=a)}$$

5. 協調フィルタリング

❖評価値の予測

	i_1	i_2	i_3	i_4	i_5
u_1	2	2	3	1	
u_2	1	3	2		3
u_3	2		3	3	2
u_4	3	2	*	2	2
$s_{j,3}$	0.3	0.9	-	0.9	0.9

$$r_{u,i} = \frac{\sum_j^m r_{u,j} s_{j,i}}{\sum_j |s_{j,i}|}$$

$$r_{4,3}$$

$$= \frac{s_{1,3} * r_{4,1} + s_{2,3} * r_{4,2} + s_{3,3} * r_{4,4} + s_{5,3} * r_{4,5}}{s_{1,3} + s_{2,3} + s_{4,3} + s_{5,3}}$$

$$= \frac{3 * 0.3 + 2 * 0.9 + 2 * 0.9 + 2 * 0.9}{0.3 + 0.9 + 0.9 + 0.9}$$

$$= 2.1$$

$$* = 2$$

5. アイテム間類似度 $S_{i,j}$

❖アイテム間類似度

	i_2	i_3
u_1	2	3
u_2	3	2
u_3		3
u_4	2	*
$s_{3,j}$	0.9	-

❖コサイン尺度

$$s_{i,j} = \frac{\sum_{k=1}^{n} r_{k,i} r_{k,j}}{\sqrt{r_{1,i}^2 + \cdots + r_{n,i}^2}\sqrt{r_{1,j}^2 + \cdots + r_{n,j}^2}}$$

$$s_{2,3} = \frac{2 * 3 + 3 * 2}{\sqrt{2^2 + 3^2}\sqrt{3^2 + 2^2}}$$

$$\simeq 0.92$$

6. 提案手法

X	i_1	i_2
u_1	2	2

↓

Y	i_1	i_2
u_1	2	3

y \ x	0	1	2	3
0	0.37	0.18	0.23	0.22
1	0.19	0.36	0.23	0.22
2	0.18	0.17	0.44	0.21
3	0.18	0.17	0.22	0.43

$$P(W|Y_1, Y_2) = \sum_{W=\alpha\beta} P(X = \alpha|Y_1) \cdot P(X = \beta|Y_2)$$

Ex.

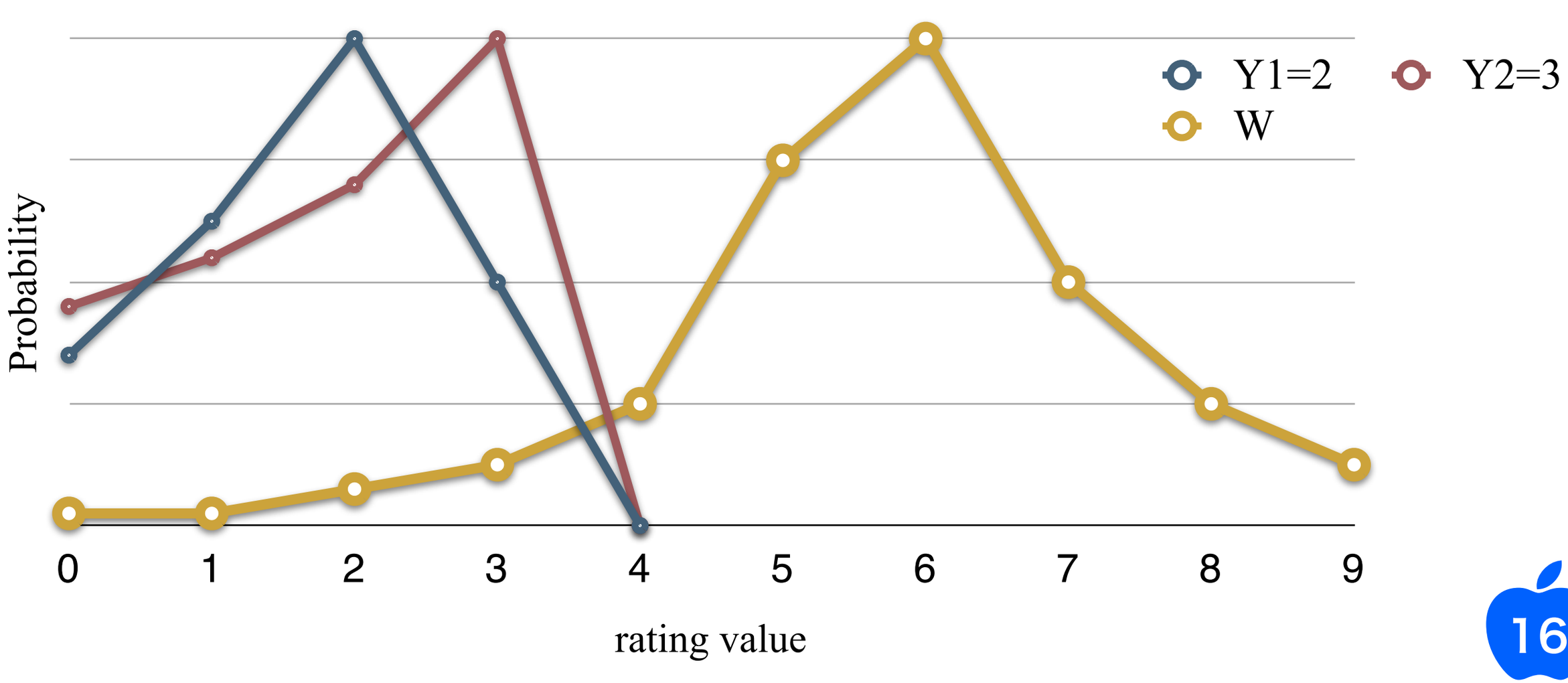
$$P(W = 6|Y_1 = 2, Y_2 = 3) = 0.20$$

5. 協調フィルタリング

❖コサイン尺度

$$s_{i,j} = \frac{\sum_{k=1}^{n} r_{k,i} r_{k,j}}{\sqrt{r_{1,i}^2 + \cdots + r_{n,i}^2}\sqrt{r_{1,j}^2 + \cdots + r_{n,j}^2}}$$

$$W = r_{k,i} r_{k,j}$$
$$Y_1 = r_{k,i}$$
$$Y_2 = r_{k,j}$$



7. 実験

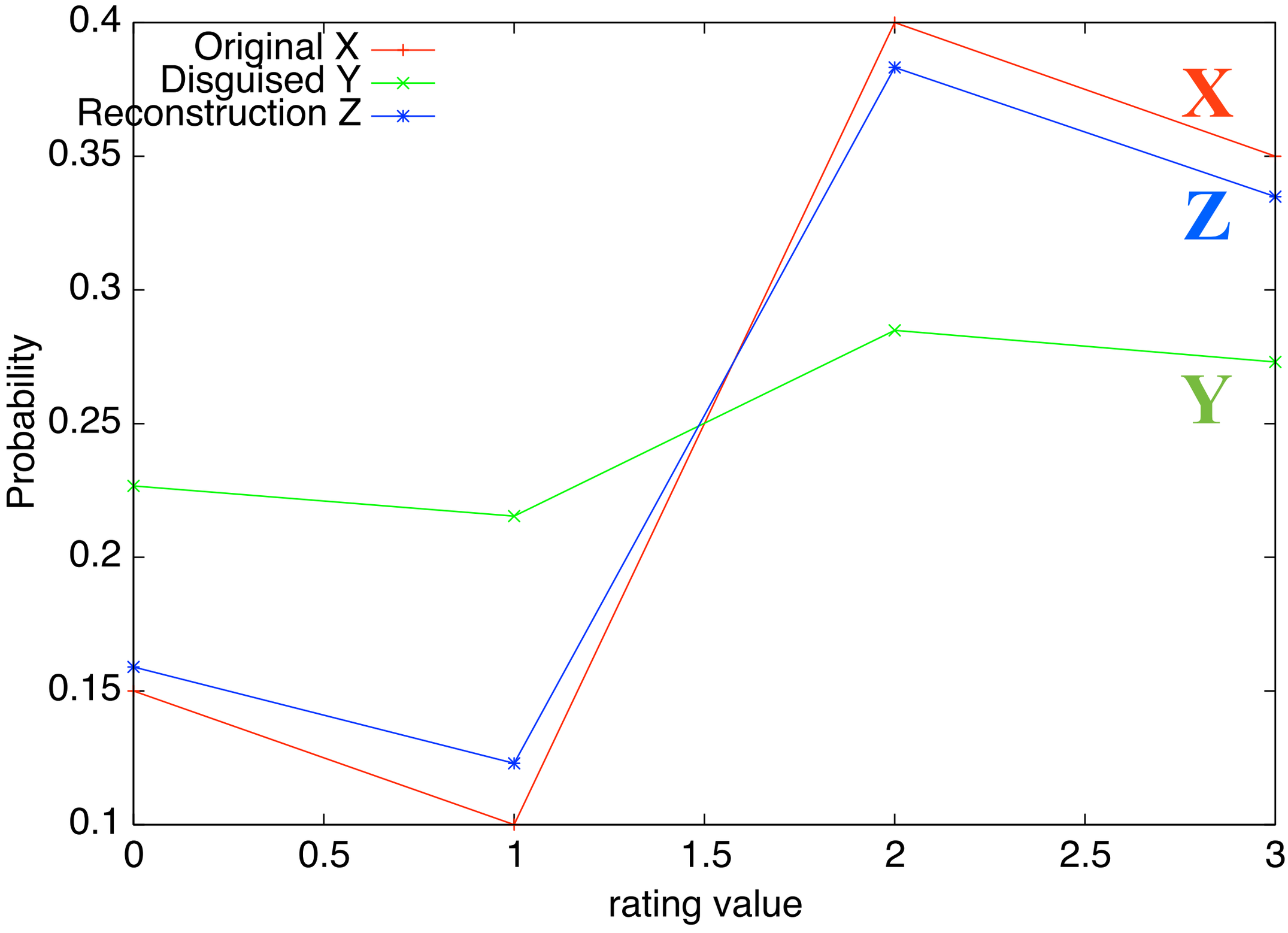
❖**実験１**：評価値の分布による再構築

❖**実験２**：評価値の積の分布

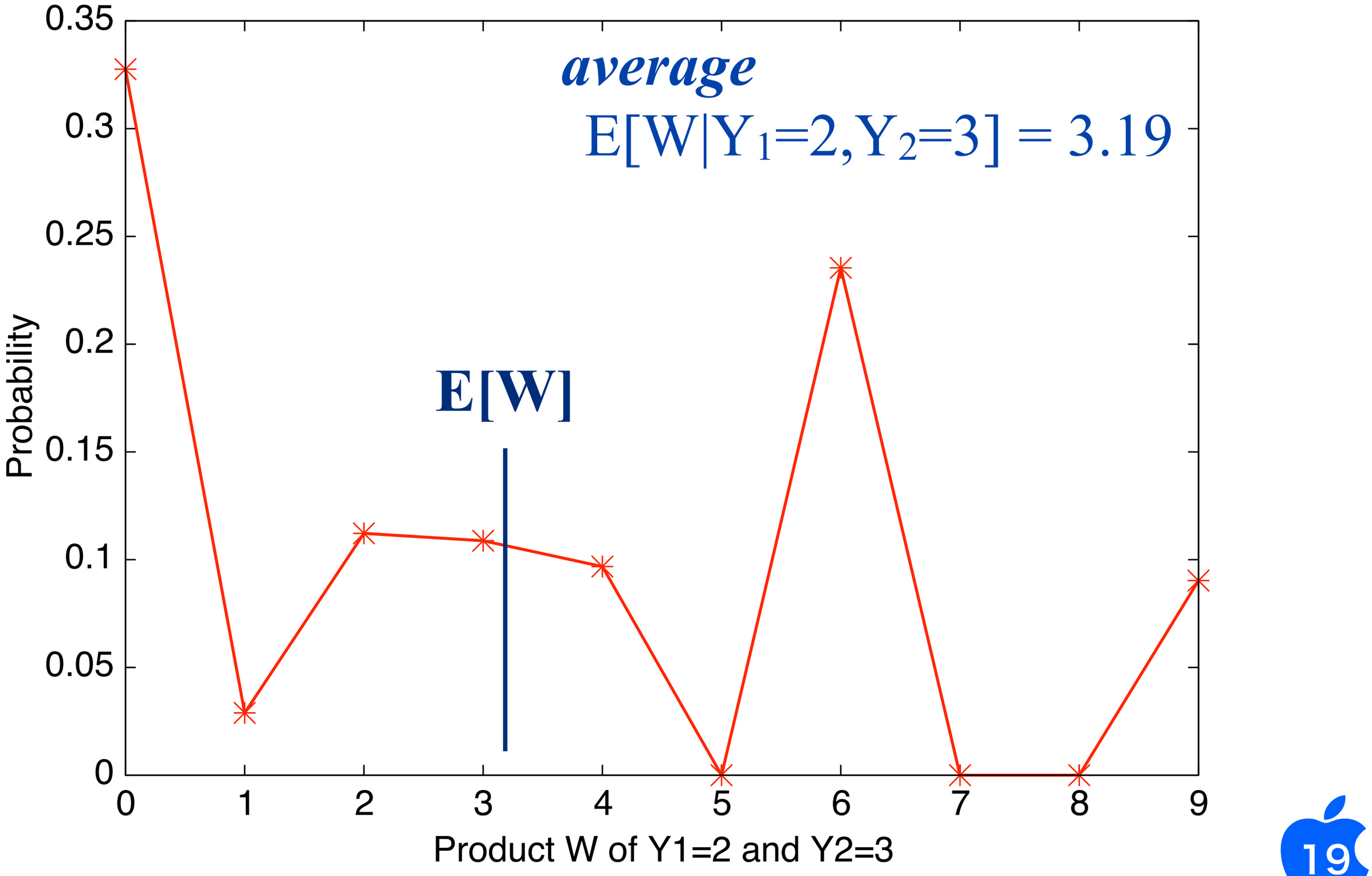
❖**実験３**：予測類似度

❖**実験４**：CFによる推薦値

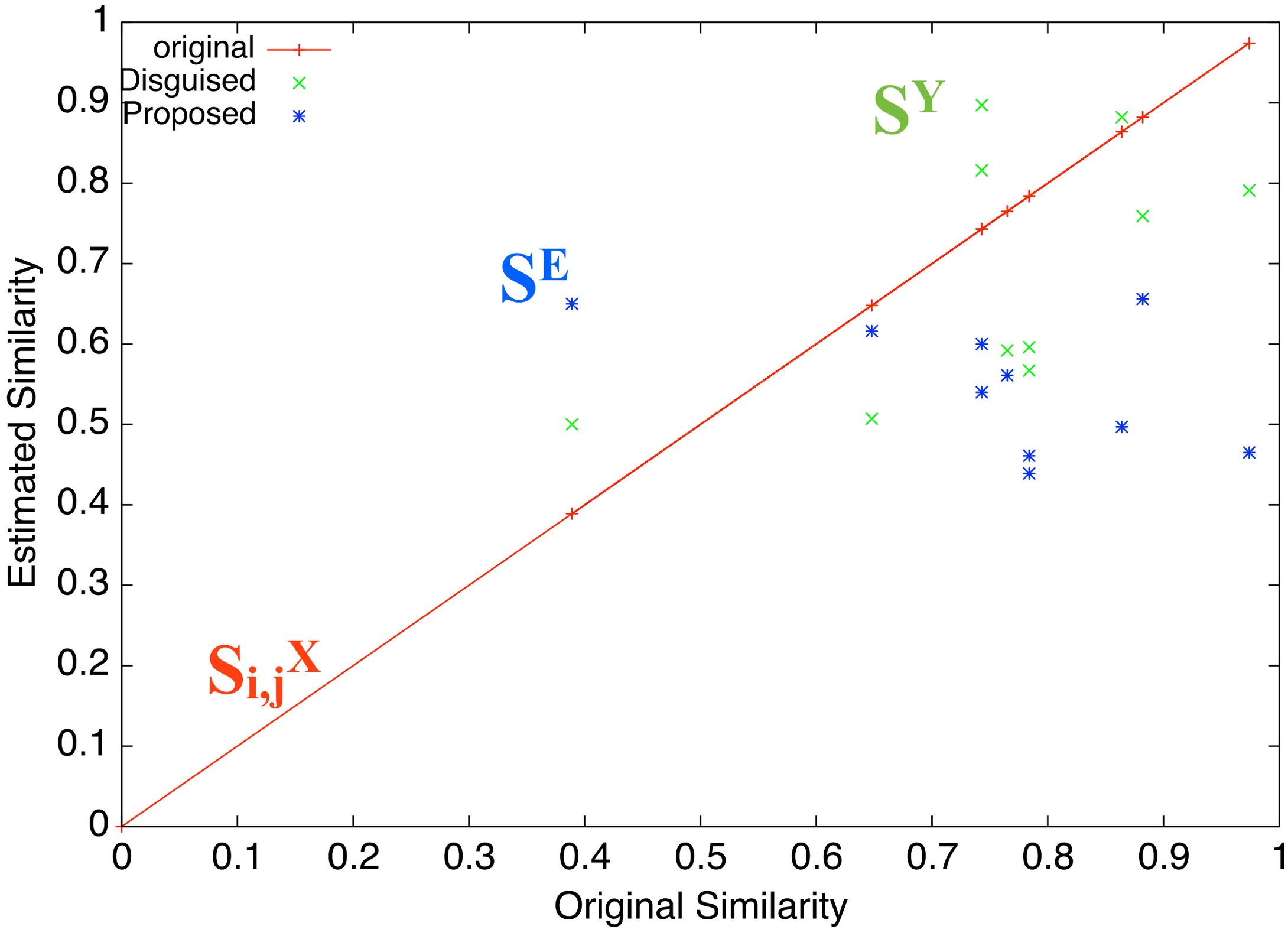
7.1 実験１：再構築された評価の分布



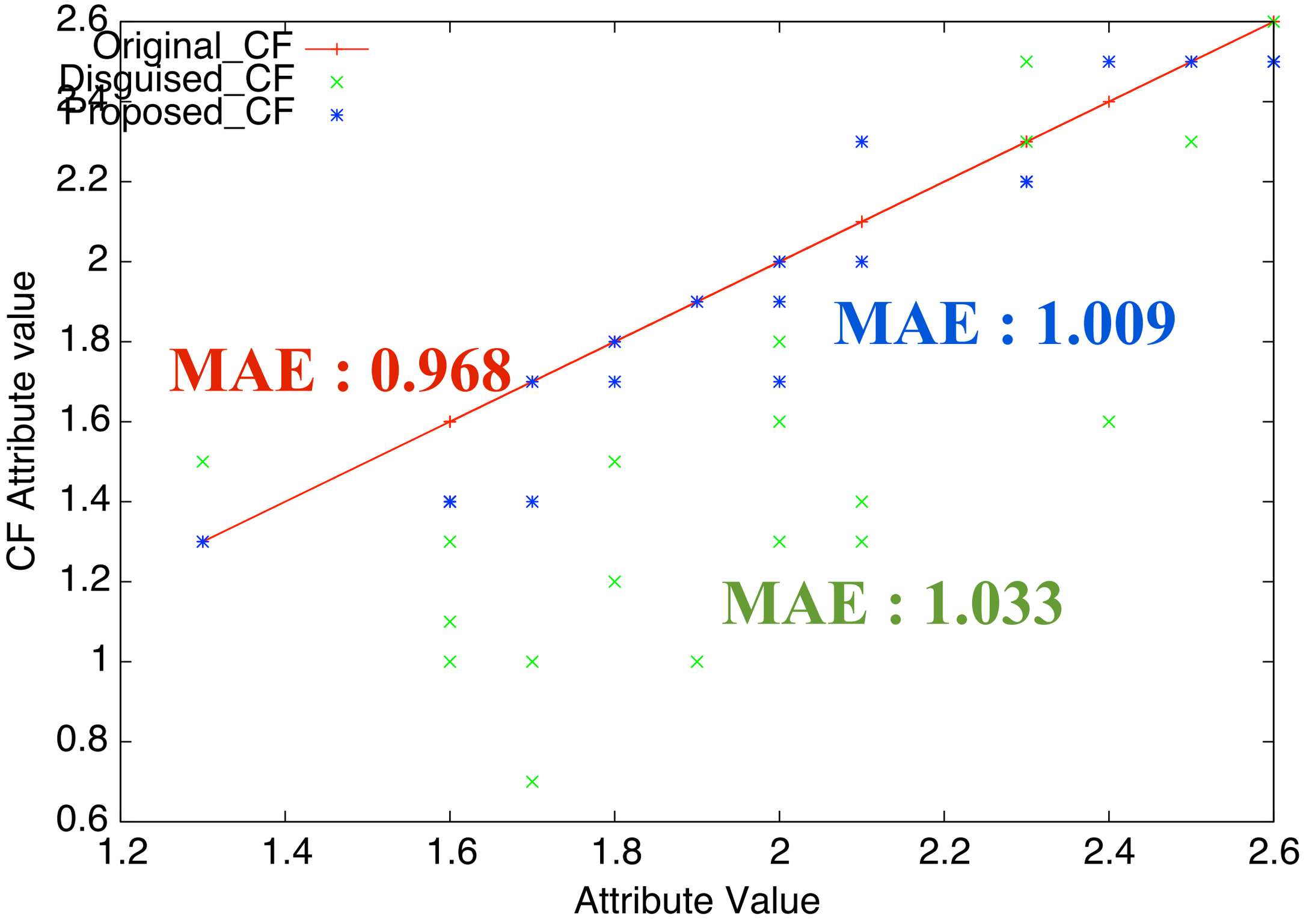
7.2 実験２：評価値の積の分布



7.3 実験３：類似度の分布



7.4 実験４：CFによる推薦値の分布



8. 結論

❖まとめ

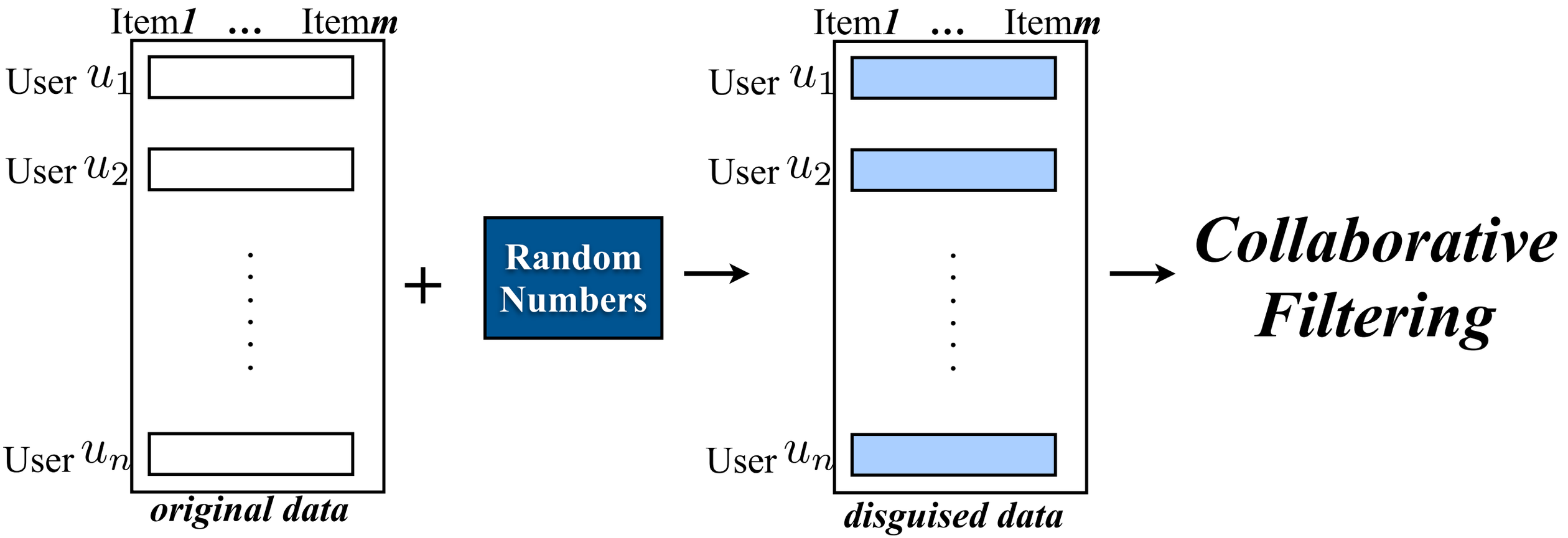
- 分布による再構築の精度は高い
- 偽データより，期待値を使ったCFの方が誤差が少ない

❖今後の課題

- 精度の向上
- 未評価の扱い

master

3.1.1 加法摂動化（Polat and Du, 2005）



❖欠点

1. 離散値に適用できない
2. 主成分分析により乱数除去が可能

6. 提案手法

- ２つの評価値の積と，その積の期待値より推薦を行う

6.1 期待値を用いたCF

❖**Step1** ２つの摂動化評価値が与えられた際，それらの積を取る確率変数の分布を求める

❖**Step2** 積の期待値を求める

❖**Step3** 摂動化行列の下で，コサイン尺度から期待値を求める

❖**Step4** 期待類似度より予測値を求める

- 評価値を知りたいユーザの他アイテムの評価した値は摂動化しない

- 評価していないものは，ユーザの評価の平均を入れる？

-