

擬準同型性を満たす類似度による 分散協調フィルタリングプロトコル

東海大学大学院 工学研究科 情報理工学専攻

修士論文発表

9KDRM001 青木 良樹

指導教員 菊池 浩明

はじめに



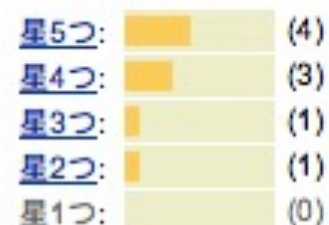
[スカーゲン]SKAGEN 腕時計 basic titanium mens T233XLTMN ケース幅: 37mm メンズ [正規輸入品]

SKAGEN(スカーゲン)

★★★★☆ (9件のカスタマーレビュー)

カスタマーレビュー

9レビュー



おすすめ度

★★★★☆ (9件のカスタマーレビュー)

Rating

これにも注目

最近チェックした商品

同じテーマの商品



Mac Fan 2011年 02月号
雑誌
¥ 730



Mac People 2011年 02月
号
雑誌



Mac People 2011年 01月
号
雑誌
¥ 890



MacBook Air パーフェク
トガイド
マックピープル編集部
大型本
¥ 1,029



MAC POWER 2010Vol.2
大型本
¥ 1,449



Mac Fan 2010年 11月号
雑誌
¥ 730

おすすめ

最近チェックした商品

同じテーマの商品



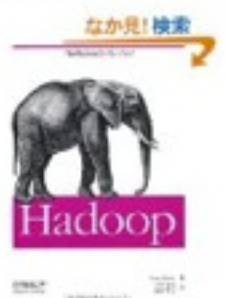
集合知プログラミング
Toby Segaran, 當山 仁
健...
大型本
¥ 3,570



集合知イン・アクション
Satnam Alag, 堀内 孝
彦, ...
大型本
¥ 3,990



ビジュアライジング・
データ...
Ben Fry, 増井 俊之 (監
訳)...
大型本
¥ 3,780



Hadoop
Tom White, 玉川 竜司,
...
大型本
¥ 4,830



入門 自然言語処理
Steven Bird, ...
大型本
¥ 3,990



パターン認識と機械学習
上 - ベイズ...
C. M. ビショップ, 元田
浩, ...
単行本
¥ 6,825

研究目的

NTT docomo

A	i_1	i_2	Sim
u_1	3	2	0.2
u_2	1	4	0.7
u_3	3		0.1
u_4	4	3	-

東海大学

B	i_3	i_4	i_5	Sim
u_1	5	1		0.1
u_2	3		4	0.3
u_3	4	5	2	0.6
u_4	*	2	2	-

要請1

互いの情報を秘匿

要請2

協力して精度向上

プライバシーを保護する2つの分割

	アイテムID				
ユーザID	i_1	i_2	i_3	i_4	i_5
u_1	3	2	5	1	
u_2	1	4	3		4
u_3	3		4	5	2
u_4	4	3		2	2

水平分割

	i_1	i_2	i_3	i_4	i_5
u_1	3	2	5	1	
u_2	1	4	3		4
u_3	3		4	5	2
u_4	4	3		2	2

垂直分割

	i_1	i_2	i_3	i_4	i_5
u_1	3	2	5	1	
u_2	1	4	3		4
u_3	3		4	5	2
u_4	4	3	*	2	2

	協調フィルタリング	他DM手法
水平分割	Canny 02 (特異値分解) 木澤 09 (秘匿積集合) Lathia 07(一致度)	佐久間ら 07(k-means) Clifton 04a(相関ルール)
垂直分割	本発表	Vaidya 03(k-means) 菊池, 香川 10(NaiveBayes)

ナイーブな方式

$$\hat{r}_{u,0}^{AB} = \frac{\sum_{v \in U - \{u\}} s_{u,v} r_{v,i}}{\sum_{v \in U - \{u\}} s_{u,v}}$$

A	i_1	i_2	B	i_3	i_4	i_5
u_1	3	2	u_1	5	1	
u_2	1	4	u_2	3		4
u_3	3		u_3	4	5	2
u_4	4	3	u_4	*	2	2

AとBが協力して計算

$$= \frac{\begin{pmatrix} r_3 a_1 a_2 + \\ r_1 a_2 a_3 + \\ r_2 a_1 a_3 \end{pmatrix} + \begin{pmatrix} r_3 b_1 a_2 + r_2 b_1 a_3 + \\ r_3 b_2 a_1 + r_1 b_2 a_3 + \\ r_2 b_3 a_1 + r_1 b_3 a_2 \end{pmatrix} + \begin{pmatrix} r_3 b_1 b_2 + \\ r_1 b_2 b_3 + \\ r_2 b_1 b_3 \end{pmatrix}}{\begin{pmatrix} a_1 a_2 + \\ a_2 a_3 + \\ a_1 a_3 \end{pmatrix} + \begin{pmatrix} b_1 a_2 + b_1 a_3 + \\ b_2 a_1 + b_2 a_3 + \\ b_3 a_1 + b_3 a_2 \end{pmatrix} + \begin{pmatrix} b_1 b_2 + \\ b_2 b_3 + \\ b_1 b_3 \end{pmatrix}}$$

Aのみ
Bのみ

問題点 送信する暗号文の数 $O(n^2)$

提案方式：基本方式

ナイーブ方式

$$\hat{r}_{u,o}^{AB} = \frac{\begin{pmatrix} r_3 a_1 a_2 + \\ r_1 a_2 a_3 + \\ r_2 a_1 a_3 \end{pmatrix} + \begin{pmatrix} r_3 b_1 a_2 + r_2 b_1 a_3 + \\ r_3 b_2 a_1 + r_1 b_2 a_3 + \\ r_2 b_3 a_1 + r_1 b_3 a_2 \end{pmatrix} + \begin{pmatrix} r_3 b_1 b_2 + \\ r_1 b_2 b_3 + \\ r_2 b_1 b_3 \end{pmatrix}}{\begin{pmatrix} a_1 a_2 + \\ a_2 a_3 + \\ a_1 a_3 \end{pmatrix} + \begin{pmatrix} b_1 a_2 + b_1 a_3 + \\ b_2 a_1 + b_2 a_3 + \\ b_3 a_1 + b_3 a_2 \end{pmatrix} + \begin{pmatrix} b_1 b_2 + \\ b_2 b_3 + \\ b_1 b_3 \end{pmatrix}}$$

基本方式

$$\hat{r}_{u,o}^{A*B} = \frac{\begin{pmatrix} r_1 a_2 a_3 + \\ r_2 a_1 a_3 + \\ r_3 a_1 a_2 \end{pmatrix}}{\begin{pmatrix} a_2 a_3 + \\ a_1 a_3 + \\ a_1 a_2 \end{pmatrix}} w_A + \frac{\begin{pmatrix} r_3 b_1 b_2 + \\ r_1 b_2 b_3 + \\ r_2 b_1 b_3 \end{pmatrix}}{\begin{pmatrix} b_1 b_2 + \\ b_2 b_3 + \\ b_1 b_3 \end{pmatrix}} w_B$$

擬準同型類似度

A	i_1	i_2	B	i_3	i_4	i_5
u_1	3	2	u_1	5	1	
u_2	1	4	u_2	3		4
u_3	3		u_3	4	5	2
u_4	4	3	u_4	*	2	2

全域的

AB	i_1	i_2	i_3	i_4	i_5
u_1	3	2	5	1	
u_2	1	4	3		4
u_3	3		4	5	2
u_4	4	3	*	2	2

局所的

\tilde{S}^A	Sim
u_1	0.2
u_2	0.7
u_3	0.1
u_4	-

\tilde{S}^B	Sim
u_1	0.1
u_2	0.3
u_3	0.6
u_4	-

合成

\tilde{S}^{A*B}	Sim
u_1	0.1
u_2	0.5
u_3	0.4
u_4	-

\tilde{S}^{AB}	Sim
u_1	0.1
u_2	0.5
u_3	0.4
u_4	-

$$|\tilde{S}^{A*B} - \tilde{S}^{AB}| < \epsilon$$

ϵ 擬準同型類似度

- **方式1：基本方式（本日の発表）**
- 方式2：事前計算方式
- 方式3： k - 近傍方式

基本方式の流れ

A	i_1	i_2	i_3	\tilde{s}
u_1	3	2	[5]	0.4
u_2	1	4	[3]	0.1
u_3	3		[4]	0.5
u_4	4	3	*	-
	Sum			1

B	i_3	i_4	i_5	\tilde{s}
u_1	5	1		0.6
u_2	3		4	0.3
u_3	4	5	2	0.1
u_4	*	2	2	-
	Sum			1

基本方式の流れ

A	i_1	i_2	i_3	\tilde{s}
u_1	3	2	[5]	0.4
u_2	1	4	[3]	0.1
u_3	3		[4]	0.5
u_4	4	3	*	-
	Sum			1

Step 1

$(4, E[5], E[3], E[4])$



B	i_3	i_4	i_5	\tilde{s}
u_1	5	1		0.6
u_2	3		4	0.3
u_3	4	5	2	0.1
u_4	*	2	2	-
	Sum			1

基本方式の流れ

Step2

$$E[5]^{0.4} \cdot E[3]^{0.1} \cdot E[4]^{0.5} = E[4.3]$$

A	i_1	i_2	i_3	\tilde{s}
u_1	3	2	[5]	0.4
u_2	1	4	[3]	0.1
u_3	3		[4]	0.5
u_4	4	3	*	-
	Sum			1

Step1

$(4, E[5], E[3], E[4])$

B	i_3	i_4	i_5	\tilde{s}
u_1	5	1		0.6
u_2	3		4	0.3
u_3	4	5	2	0.1
u_4	*	2	2	-
	Sum			1

基本方式の流れ

Step2

$$E[5]^{0.4} \cdot E[3]^{0.1} \cdot E[4]^{0.5} = E[4.3]$$

A	i_1	i_2	i_3	\tilde{s}
u_1	3	2	[5]	0.4
u_2	1	4	[3]	0.1
u_3	3		[4]	0.5
u_4	4	3	*	-
	Sum			1

Step1

$(4, E[5], E[3], E[4])$



Step3

$(E[4.3])$



B	i_3	i_4	i_5	\tilde{s}
u_1	5	1		0.6
u_2	3		4	0.3
u_3	4	5	2	0.1
u_4	*	2	2	-
	Sum			1

基本方式の流れ

Step2

$$E[5]^{0.4} \cdot E[3]^{0.1} \cdot E[4]^{0.5} = E[4.3]$$

A	i_1	i_2	i_3	\tilde{s}
u_1	3	2	[5]	0.4
u_2	1	4	[3]	0.1
u_3	3		[4]	0.5
u_4	4	3	*	-
Sum				1

Step1

$(4, E[5], E[3], E[4])$

B	i_3	i_4	i_5	\tilde{s}
u_1	5	1		0.6
u_2	3		4	0.3
u_3	4	5	2	0.1
u_4	*	2	2	-
Sum				1

Step3

$(E[4.3])$

Step4

$$\hat{r}_{4,3}^{A*B} = D[E[4.3]] \frac{2}{5} + 4.4 \frac{3}{5} = 4.36$$

- 変形ユークリッド距離（中間発表）
- コサイン類以度
- ピアソン相関係数
- **Somers' d 類以度（本日の発表）**

Somers' d 類以度

次の式を計算し，比較することによって算出

$$f_{a,i} = r_{a,i} - \bar{r}_{a,i}$$

1. Concordant : 一致

$$f_{a,i} > 0 \text{ and } f_{b,i} > 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} < 0$$

2. Discordant : 不一致

$$f_{a,i} > 0 \text{ and } f_{b,i} < 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} > 0$$

3. Tied : 同順

$$f_{a,i} = 0 \text{ or } f_{b,i} = 0$$

$r_{a,i}$ は欠損値

$r_{b,i}$ は欠損値

	i_1	i_2	i_3	i_4	i_5	ave
u_1	4	2	5	1	5	3.4
u_2	1	4	3		4	3

$$d_{a,b} = \frac{C - D}{N - T}$$

Somers' d 類以度

次の式を計算し，比較することによって算出

$$f_{a,i} = r_{a,i} - \bar{r}_{a,i}$$

1. Concordant : 一致

$$f_{a,i} > 0 \text{ and } f_{b,i} > 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} < 0$$

2. Discordant : 不一致

$$f_{a,i} > 0 \text{ and } f_{b,i} < 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} > 0$$

3. Tied : 同順

$$f_{a,i} = 0 \text{ or } f_{b,i} = 0$$

$r_{a,i}$ は欠損値

$r_{b,i}$ は欠損値

	i_1	i_2	i_3	i_4	i_5	<i>ave</i>
u_1	4	2	5	1	5	3.4
u_2	1	4	3		4	3
$f_{1,*}$	0.6	-1.4	1.6	-2.4	1.6	
$f_{2,*}$	-2	1	0		1	

$$d_{a,b} = \frac{C - D}{N - T}$$

Somers' d 類以度

次の式を計算し，比較することによって算出

$$f_{a,i} = r_{a,i} - \bar{r}_{a,i}$$

1. **Concordant : 一致**

$$f_{a,i} > 0 \text{ and } f_{b,i} > 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} < 0$$

2. **Discordant : 不一致**

$$f_{a,i} > 0 \text{ and } f_{b,i} < 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} > 0$$

3. **Tied : 同順**

$$f_{a,i} = 0 \text{ or } f_{b,i} = 0$$

$r_{a,i}$ は欠損値

$r_{b,i}$ は欠損値

	i_1	i_2	i_3	i_4	i_5	<i>ave</i>
u_1	4	2	5	1	5	3.4
u_2	1	4	3		4	3
$f_{1,*}$	0.6	-1.4	1.6	-2.4	1.6	
$f_{2,*}$	-2	1	0		1	
	d	d	t	t	c	

$$d_{a,b} = \frac{C - D}{N - T}$$

Somers' d 類以度

次の式を計算し，比較することによって算出

$$f_{a,i} = r_{a,i} - \bar{r}_{a,i}$$

1. **Concordant : 一致**

$$f_{a,i} > 0 \text{ and } f_{b,i} > 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} < 0$$

2. **Discordant : 不一致**

$$f_{a,i} > 0 \text{ and } f_{b,i} < 0$$

$$f_{a,i} < 0 \text{ and } f_{b,i} > 0$$

3. **Tied : 同順**

$$f_{a,i} = 0 \text{ or } f_{b,i} = 0$$

$r_{a,i}$ は欠損値

$r_{b,i}$ は欠損値

	i_1	i_2	i_3	i_4	i_5	<i>ave</i>
u_1	4	2	5	1	5	3.4
u_2	1	4	3		4	3
$f_{1,*}$	0.6	-1.4	1.6	-2.4	1.6	
$f_{2,*}$	-2	1	0		1	
	d	d	t	t	c	

$$d_{a,b} = \frac{C - D}{N - T} \quad d_{1,2} = \frac{1 - 2}{5 - 2} = -0.3$$

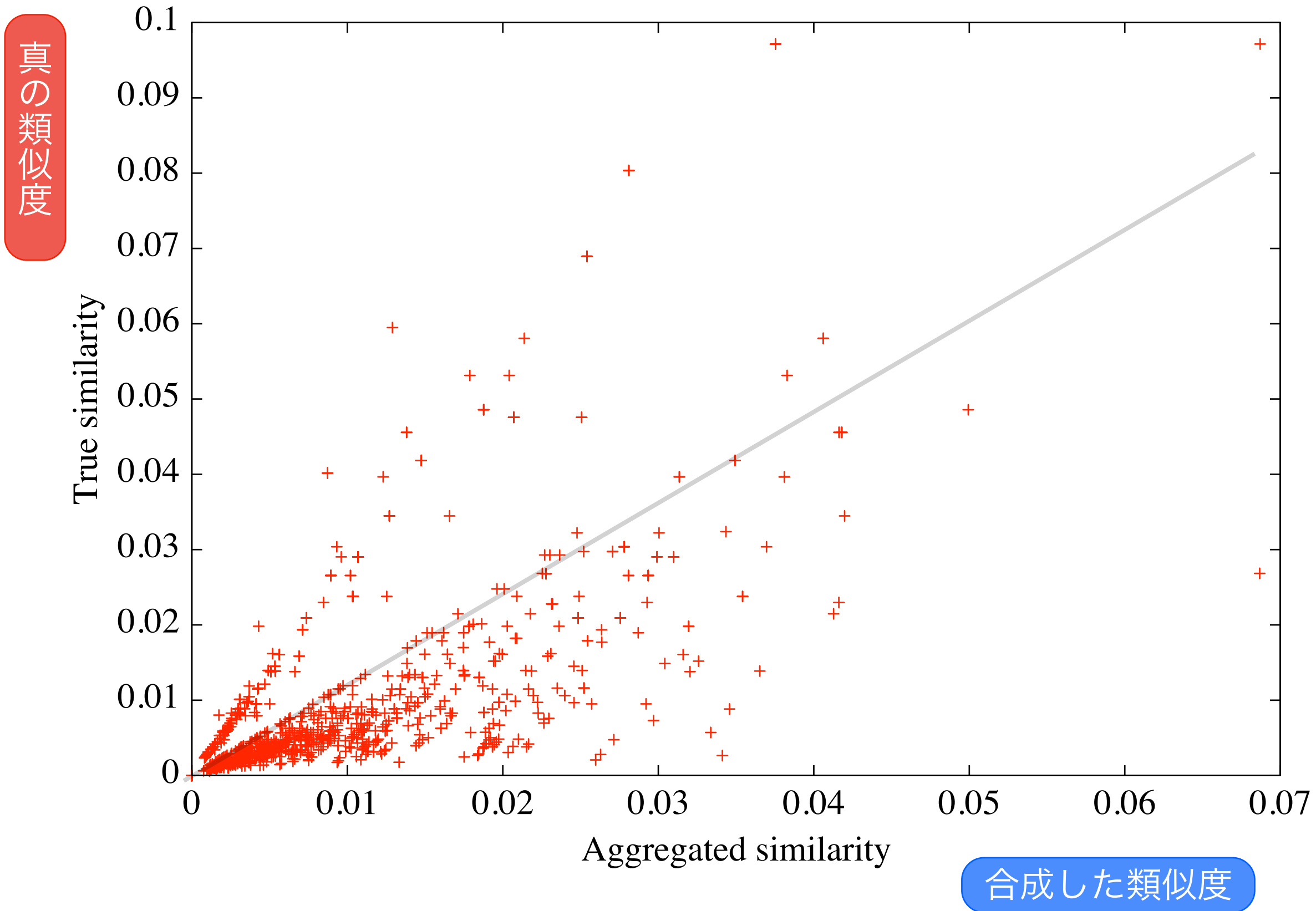
■ 評価項目

- i. 速度評価（中間発表で報告済み）
- ii. **二つの類似度の相関関係**
- iii. 精度評価(MAE)（中間発表で報告済み）
- iv. 類似度の違いによる精度
- v. **DBの違いによる精度（本日の発表）**

■ 実験環境

- Intel Core 2 Duo 2.26GHz, 4GB, Java ver. 1.6
- 使用データ：MovieLens data set
 - 評価数**10万件**, $n=943$, $m=1628$

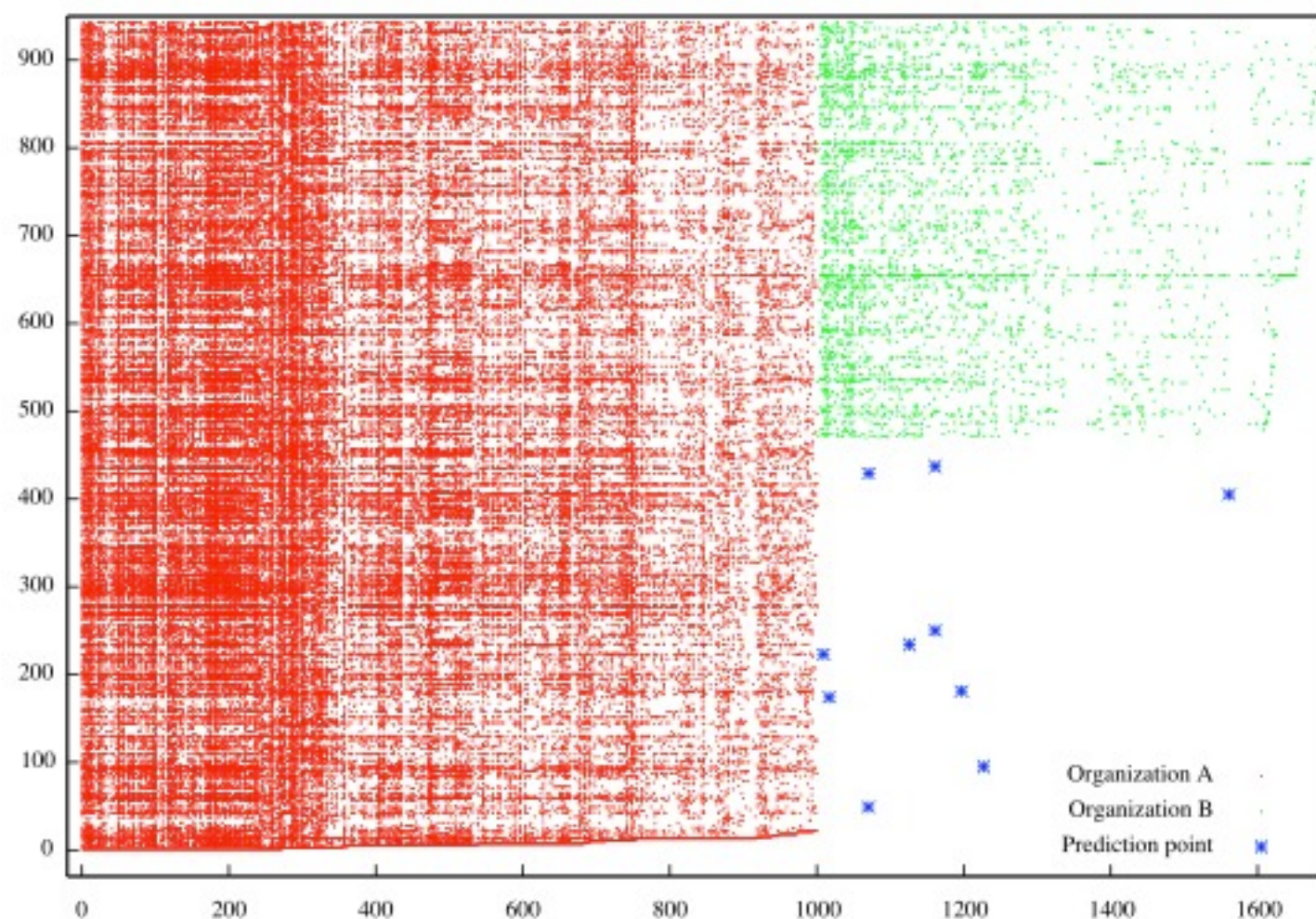
二つの類似度の散布図



密度の異なるDBでの推薦

組織A

組織B



Item ID

- ピアソン相関係数とSomers' d 類以度は局所DBのみの時より、全DBの値に近づけた

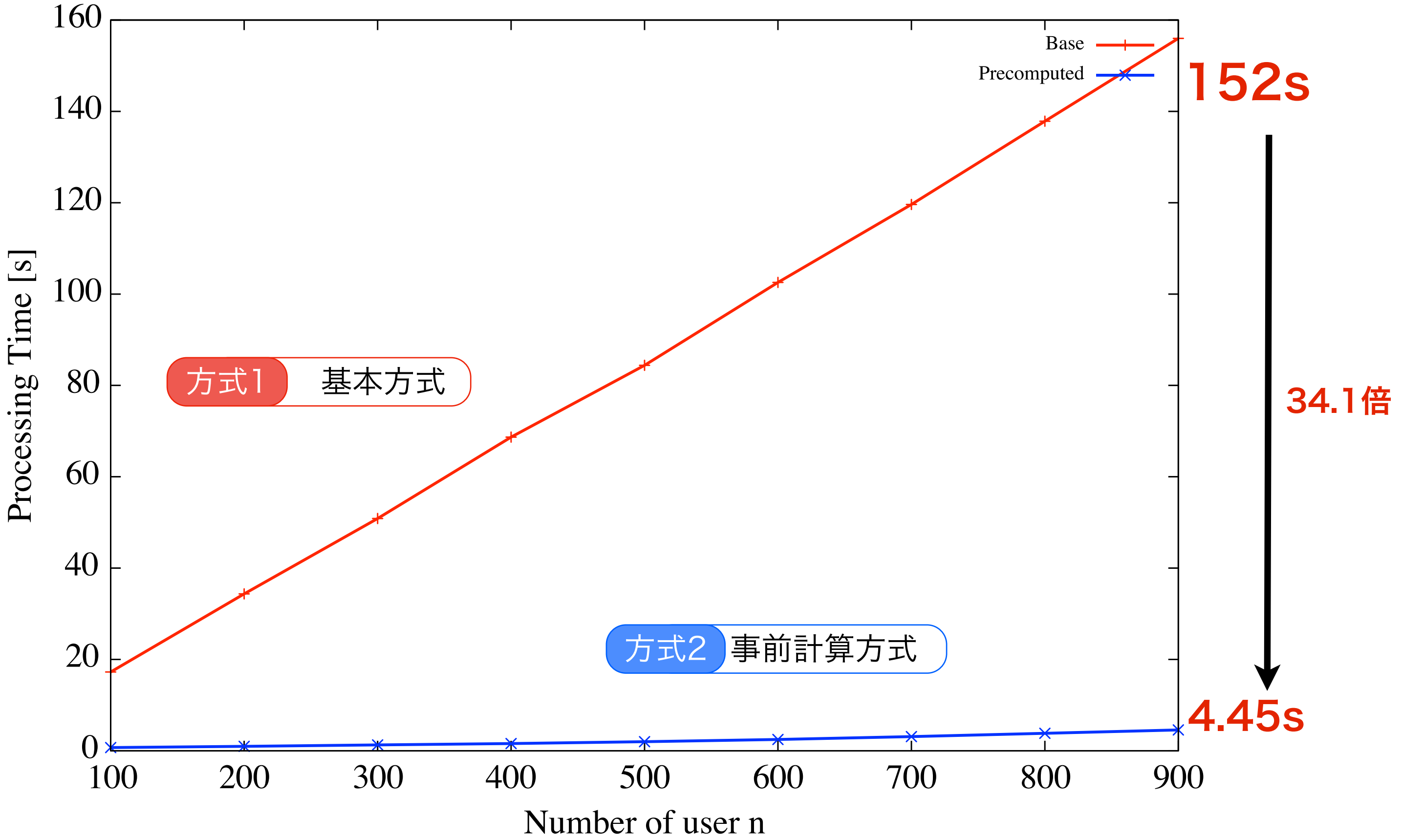
	Mod. Euc.	Cosine	Pearson	Somer's d
全 DB	1.325	1.298	0.922	1.024
分散 DB	1.517	1.490	1.086	1.226
局所 DB	1.292	1.292	1.559	1.292

- 垂直分割,二団体間の秘匿分散協調フィルタリングを提案
- ナイーブな方式, **基本方式**, **事前計算方式**, **k -近傍方式**
- **擬準同型類似度**を提案し, スパース率が極端に異なるDBに適用した

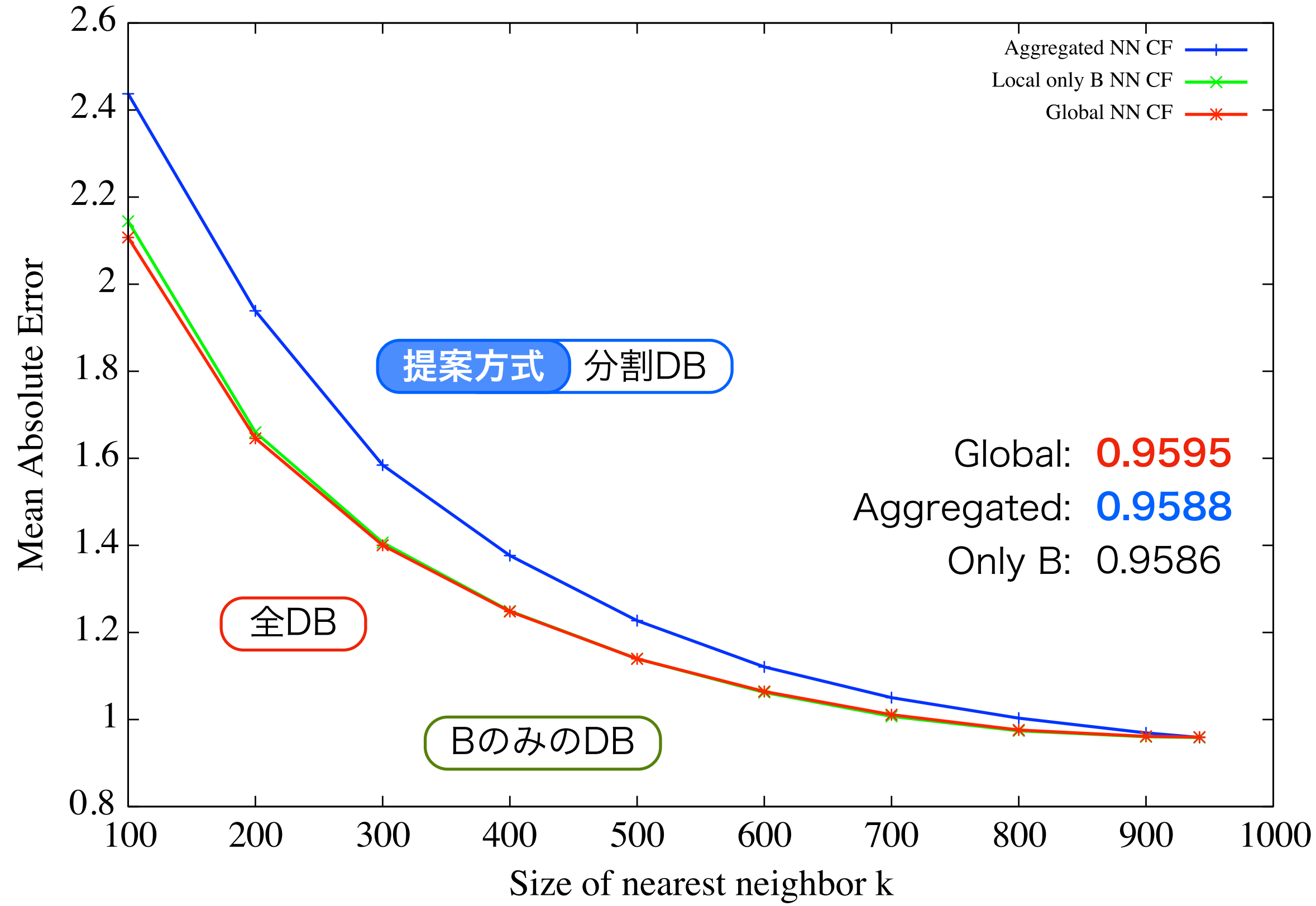
	計算コスト	通信コスト	精度
ナイーブな方式	×	×	○
方式1: 基本	○	○	△
方式2: 事前計算	◎	○	△
方式3: k -近傍	◎	◎	×

- 今後の課題として, 最適な組織に対する重みの検討や, 重みを使う必要のない完全準同型性を満たした類似度の検討がある.

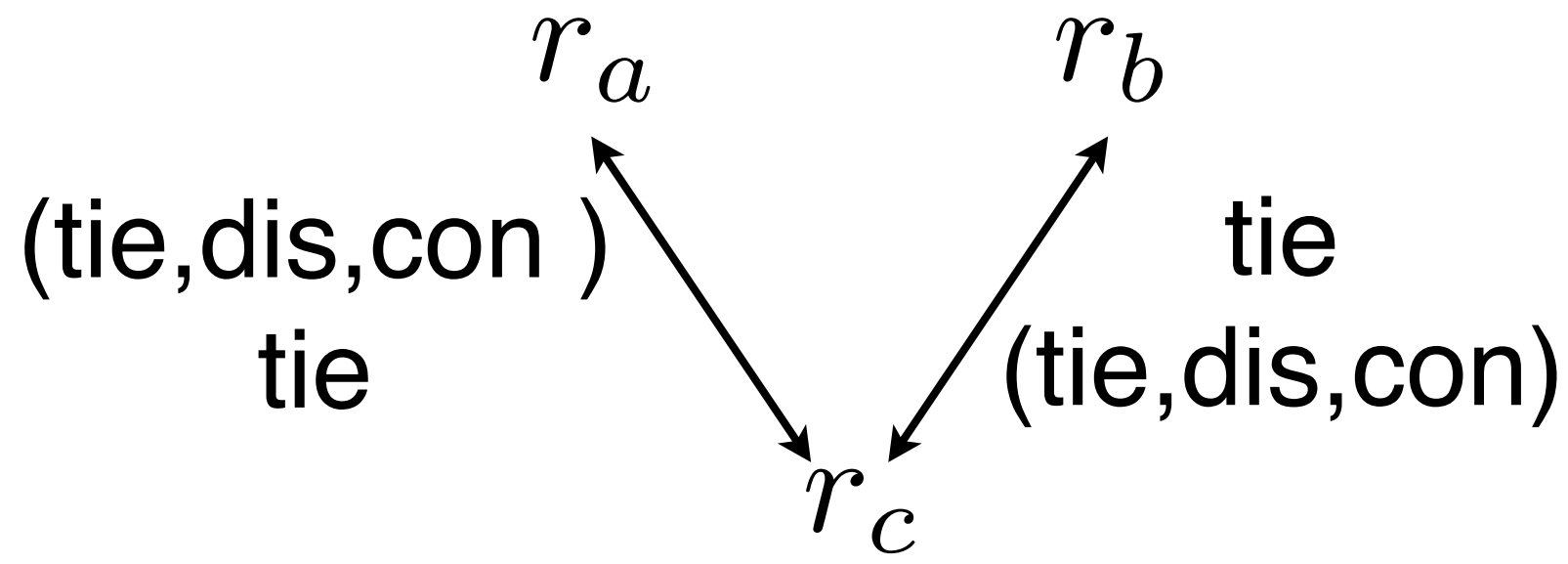
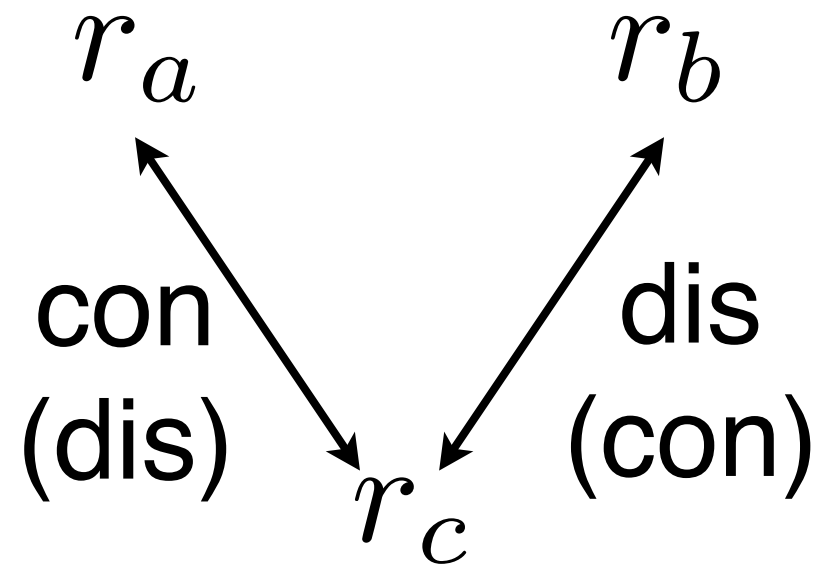
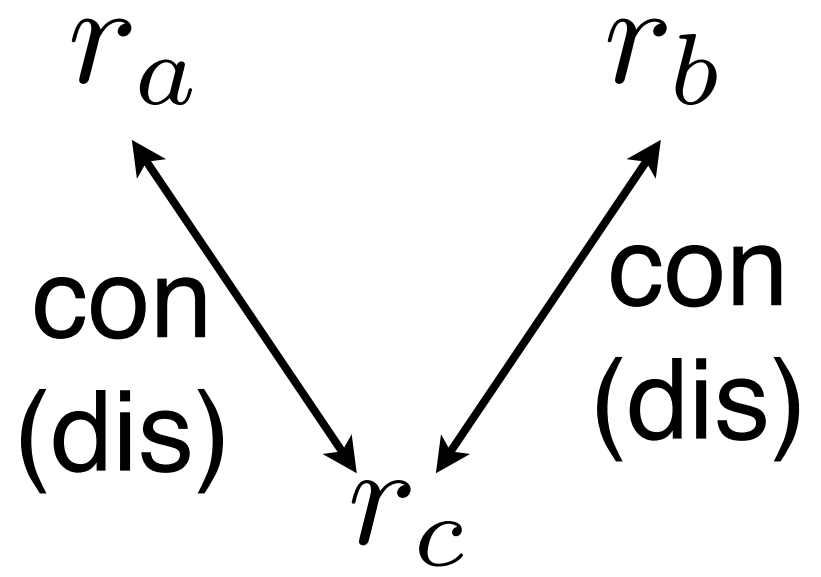
速度評価



精度評価 (MAE)

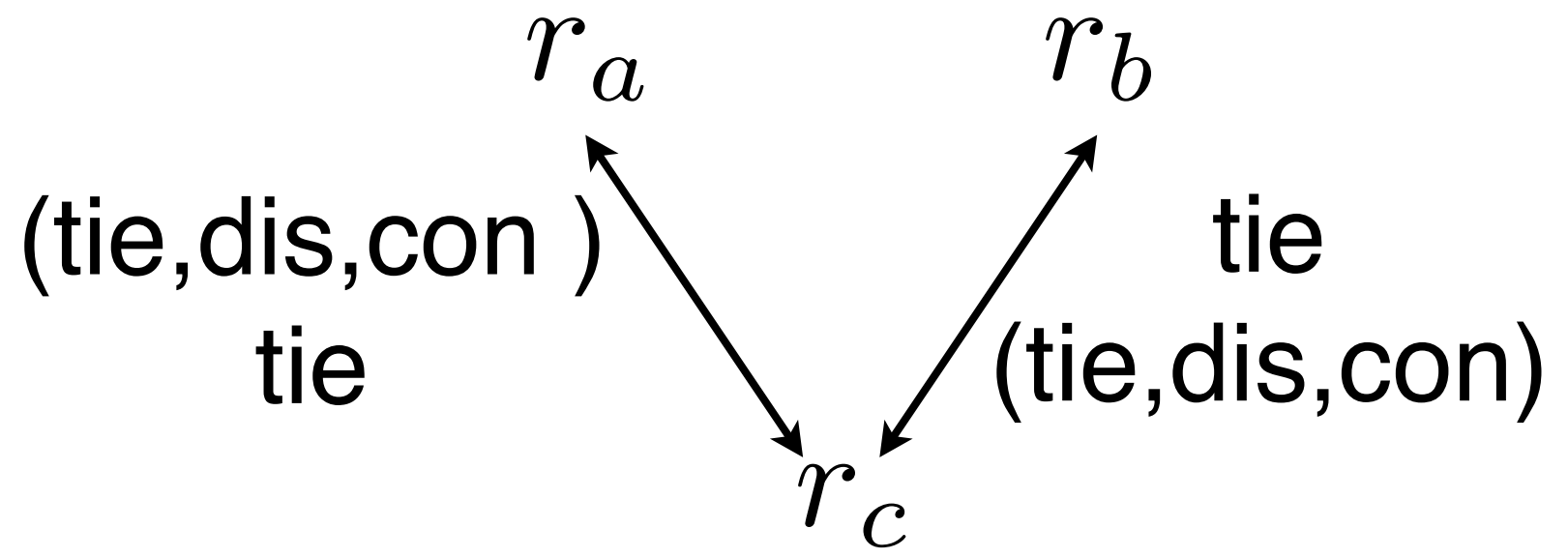
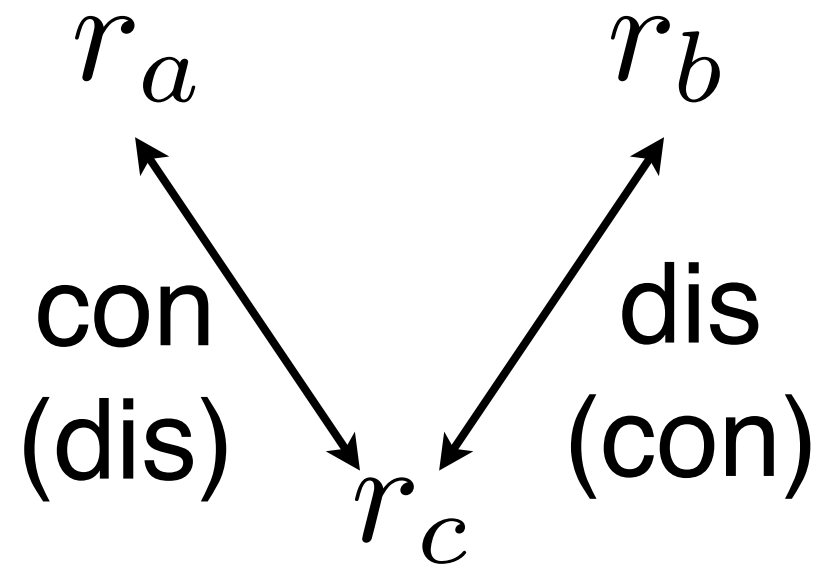
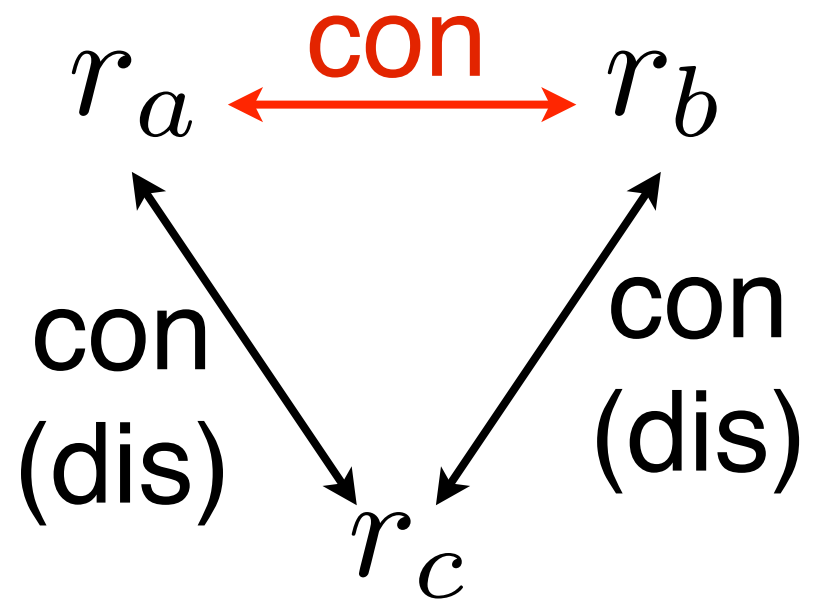


6.2. Transitivity of Concordance



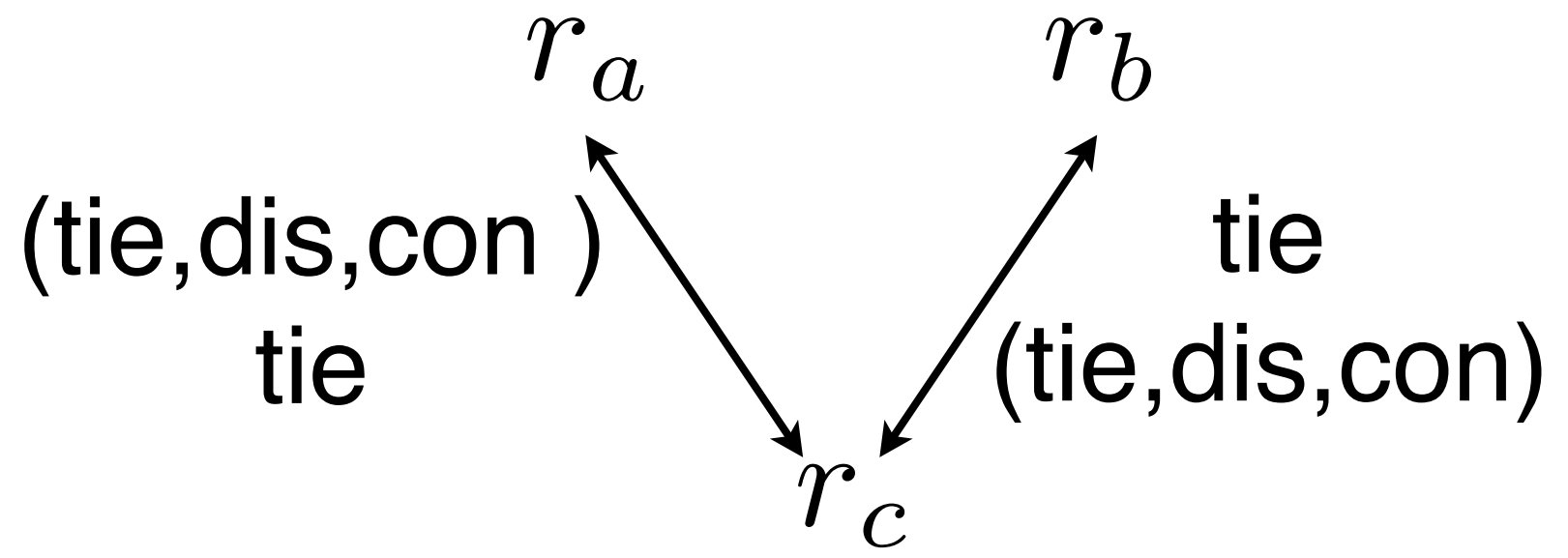
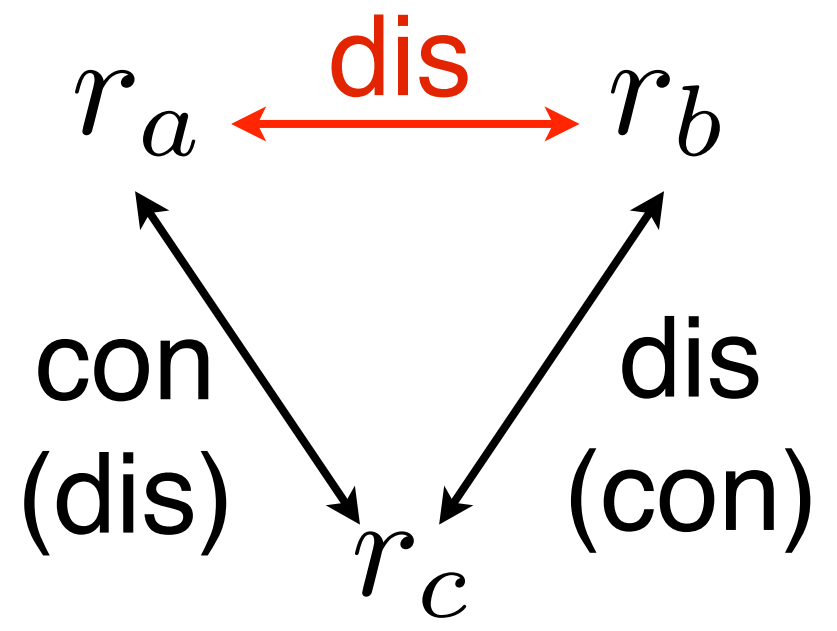
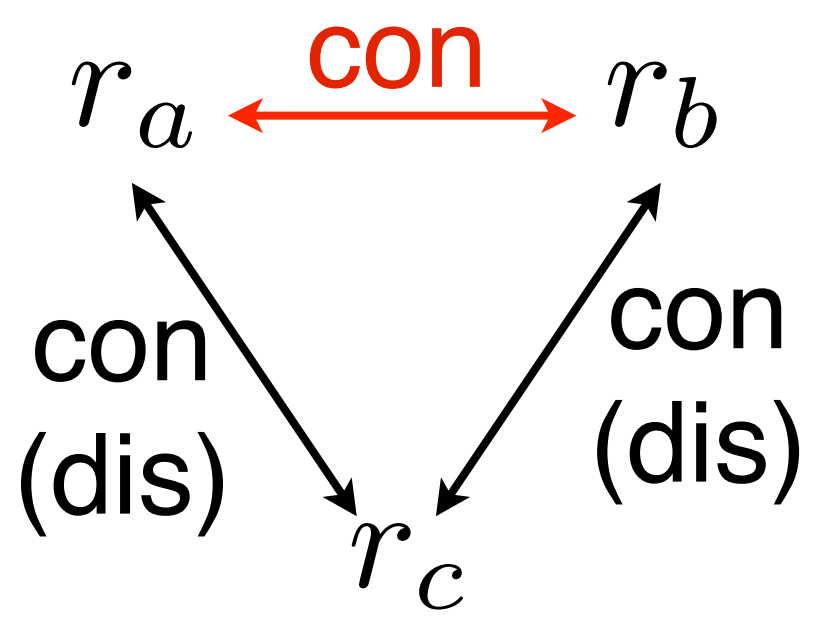
条件
 $r_c \neq \bar{r}_c$
 or
 $r_c \neq unrated$

6.2. Transitivity of Concordance



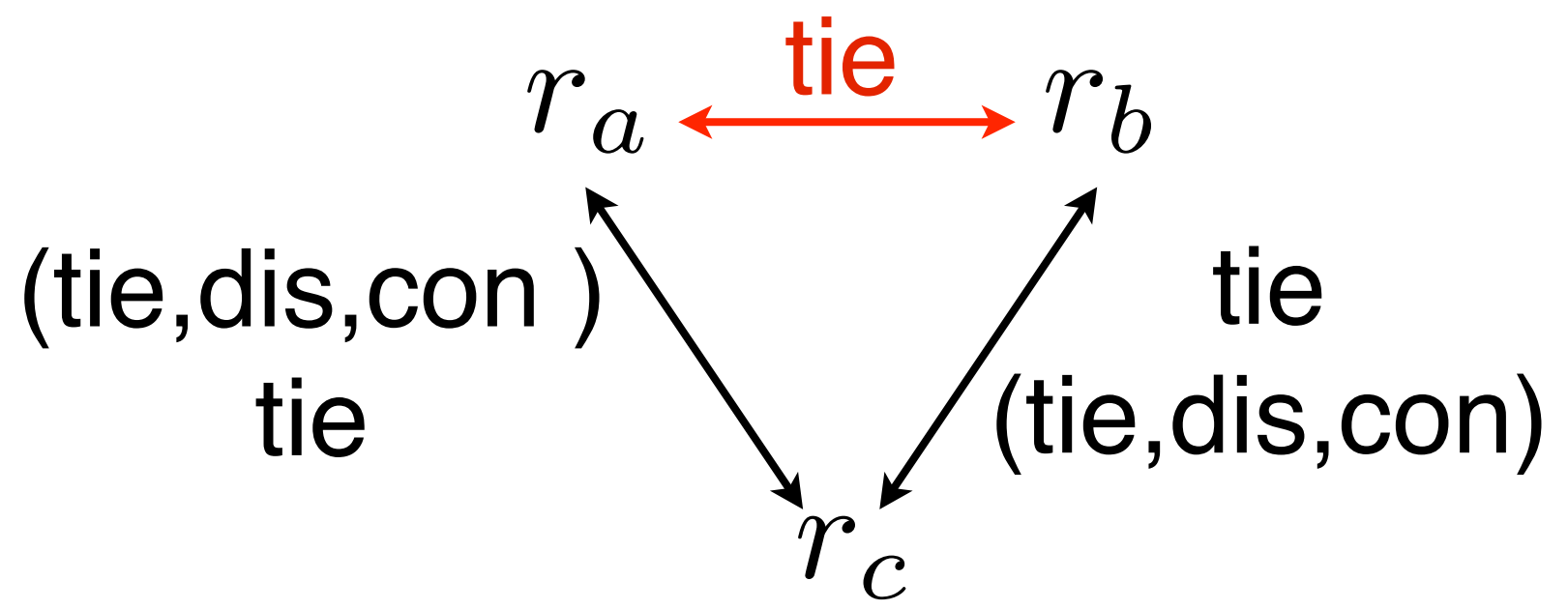
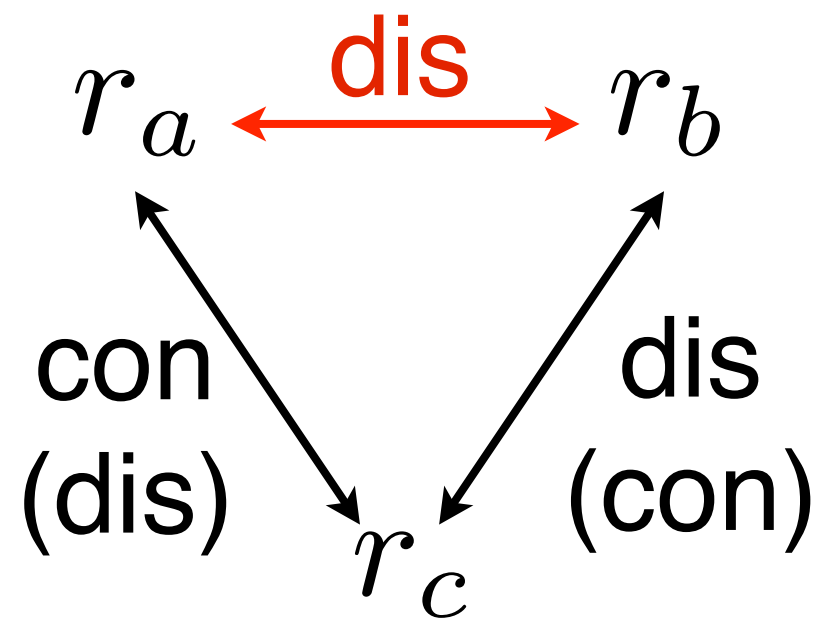
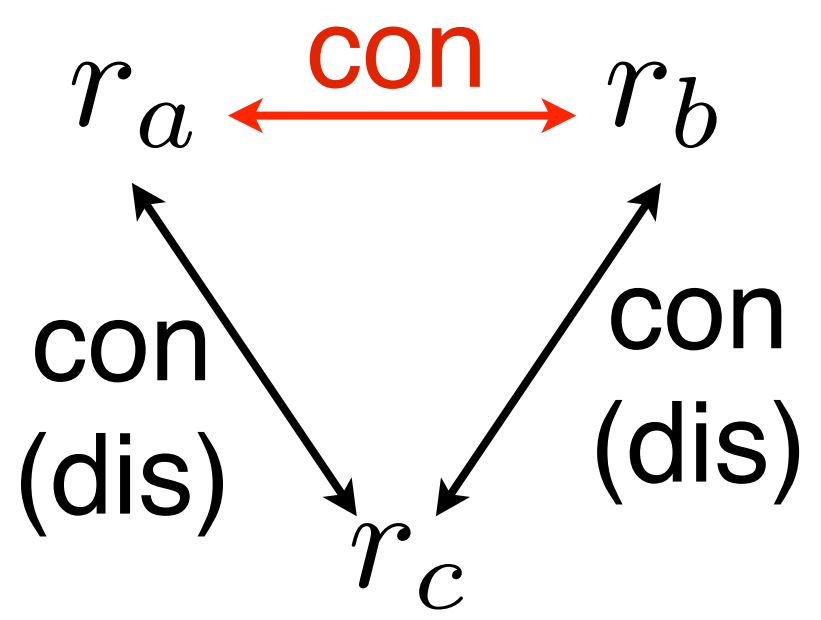
条件
 $r_c \neq \bar{r}_c$
 or
 $r_c \neq unrated$

6.2. Transitivity of Concordance



条件
 $r_c \neq \bar{r}_c$
 or
 $r_c \neq unrated$

6.2. Transitivity of Concordance



条件
 $r_c \neq \bar{r}_c$
 or
 $r_c \neq unrated$

6.3. Example

$$\begin{aligned}\bar{r}_a &= 3.5 \\ r_{a,i} &= 4 \\ f_{a,i} &= 0.5\end{aligned}$$

$$\begin{aligned}\bar{r}_b &= 2 \\ r_{b,i} &= 3 \\ f_{b,i} &= 1\end{aligned}$$

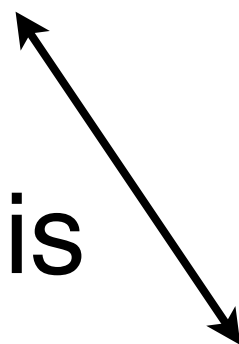
$$\begin{aligned}f_{c,i} &= -1.5 \\ \bar{r}_c &= 2.5 \\ r_{c,i} &= 1\end{aligned}$$

6.3. Example

$$\begin{aligned}\bar{r}_a &= 3.5 \\ r_{a,i} &= 4 \\ f_{a,i} &= 0.5\end{aligned}$$

$$\begin{aligned}\bar{r}_b &= 2 \\ r_{b,i} &= 3 \\ f_{b,i} &= 1\end{aligned}$$

dis



$$\begin{aligned}f_{c,i} &= -1.5 \\ \bar{r}_c &= 2.5 \\ r_{c,i} &= 1\end{aligned}$$

6.3. Example

$$\begin{aligned}\bar{r}_a &= 3.5 \\ r_{a,i} &= 4 \\ f_{a,i} &= 0.5\end{aligned}$$

$$\begin{aligned}\bar{r}_b &= 2 \\ r_{b,i} &= 3 \\ f_{b,i} &= 1\end{aligned}$$

$$\begin{aligned}f_{c,i} &= -1.5 \\ \bar{r}_c &= 2.5 \\ r_{c,i} &= 1\end{aligned}$$

6.3. Example

