

個人情報フィルタを使った エッジ型情報収集支援における 個人クエリ間の違いの定量化

金沢工業大学工学部情報工学科

芦川 拓実(3年)

風尾 勇佑(3年)

佐治 寿一(3年)

小林 暉(3年)

金道 敏樹(指導教官)

目次

- 構想
- 結果予測
- 実験・結果
- まとめ

ネット利用について私たちの考えたこと

- 人の知恵やノウハウを、プライバシーを保ちながら利用したい！
- 興味が似た人の個人クエリを利用することは便利かも！
- 二つの個人クエリを直接比較すれば、類似性は分かる。
- 直接比較だと、個人のプライバシーを保てない。
- 個人クエリの中身を見ないで、類似性を評価したい。
- それぞれのコンテンツの評価値を使うと、類似性は分かるのではないか。

我々が実現したいこと

スコア利用前

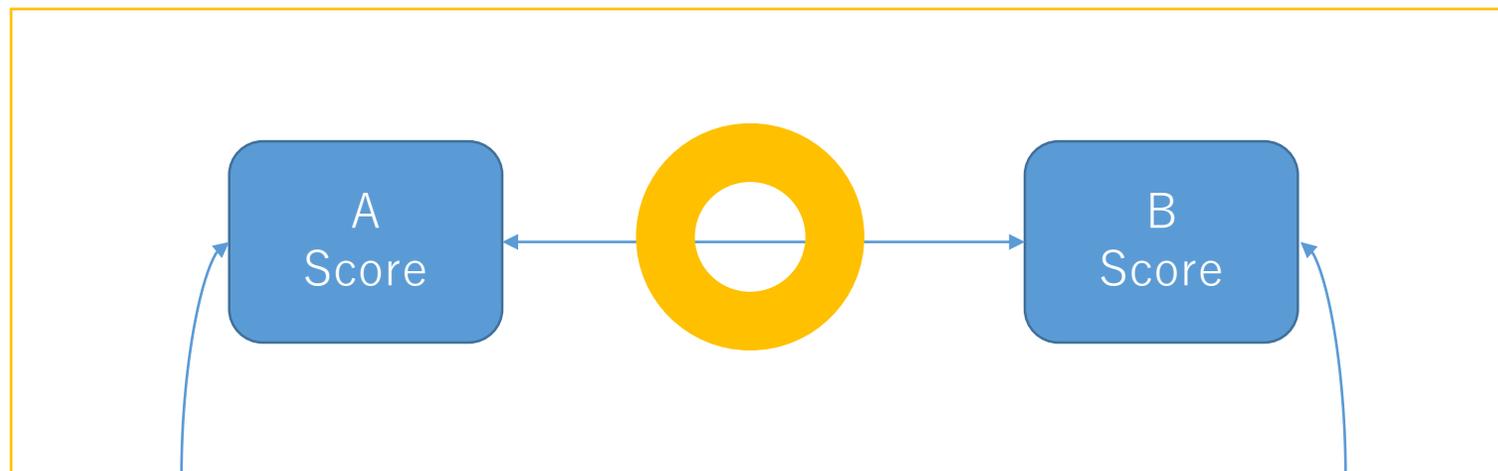
○ ×		ID:60「明日に向かって撃て！」
○ ×		ID:187「夕陽の群像」
○ ×		ID:184「GiveandGo-ギブアンドゴー」
○ ×		ID:294「オリーブの林を歩いて」
○ ×		ID:30「ゴジラ キング・オブ・モンスターズ」
○ ×		ID:34「劇場版仮面ライダーエグゼイドトゥルーエンディング」
○ ×		ID:109「マダーオブキャット」
○ ×		ID:191「ノルトの楽園」
○ ×		ID:85「コララインとボタンの魔女」
○ ×		ID:24「ディーブインセクト」
○ ×		ID:283「悪魔で最高のサリー」
○ ×		ID:173「ひろしま」
○ ×		ID:297「ファントム・スレッド」
○ ×		ID:293「ういらぶ」

スコア利用後

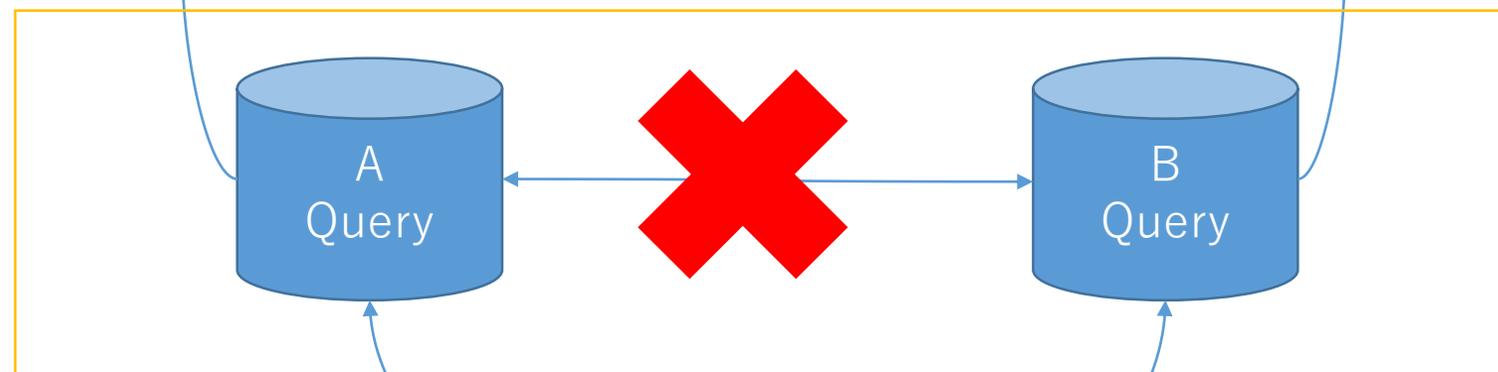
○ ×		ID:286「セカンドバージン」
○ ×		ID:294「オリーブの林を歩いて」
○ ×		ID:89「きみと、波にのれたら」
○ ×		ID:293「ういらぶ」
○ ×		ID:96「ONE PIECE FILM GOLD」
○ ×		ID:287「おとなの恋はまわり道」
○ ×		ID:290「夏ノ日、君ノ声」
○ ×		ID:217「心が叫びたがってるんだ。(2017)」
○ ×		ID:53「刑事戦隊トッキュウジャーTHEMOVIEギャラクシーラインSOS」
○ ×		ID:87「森の時間を食べたい(2018)」
○ ×		ID:296「想影」
○ ×		ID:125「マスカレードホテル」
○ ×		ID:75「ディリリとパリの時間旅行」
○ ×		ID:78「夏目友人帳うつせみに結ぶ」

ユーザが見落とし
ていたコンテンツ
を推薦したい

図で描くと・・・

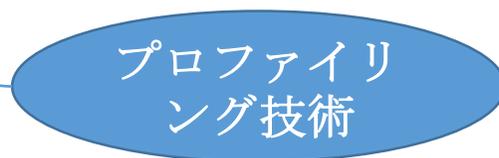


直接比較はユーザーの
プライバシーを
保護が難しい



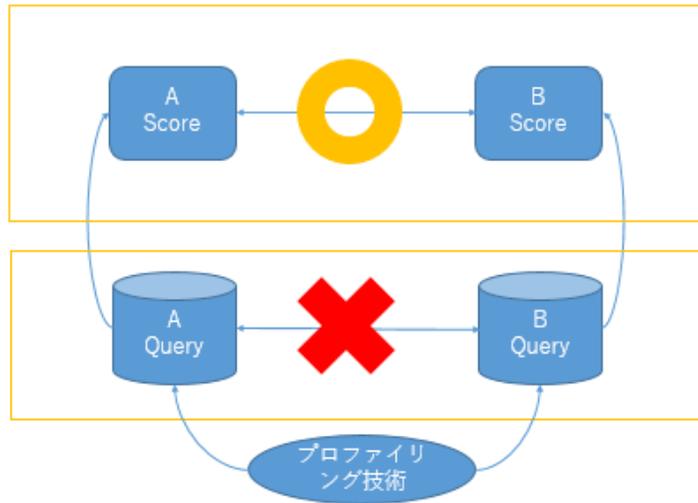
コンテンツのスコアの比較

プライバシーの
保護ができる



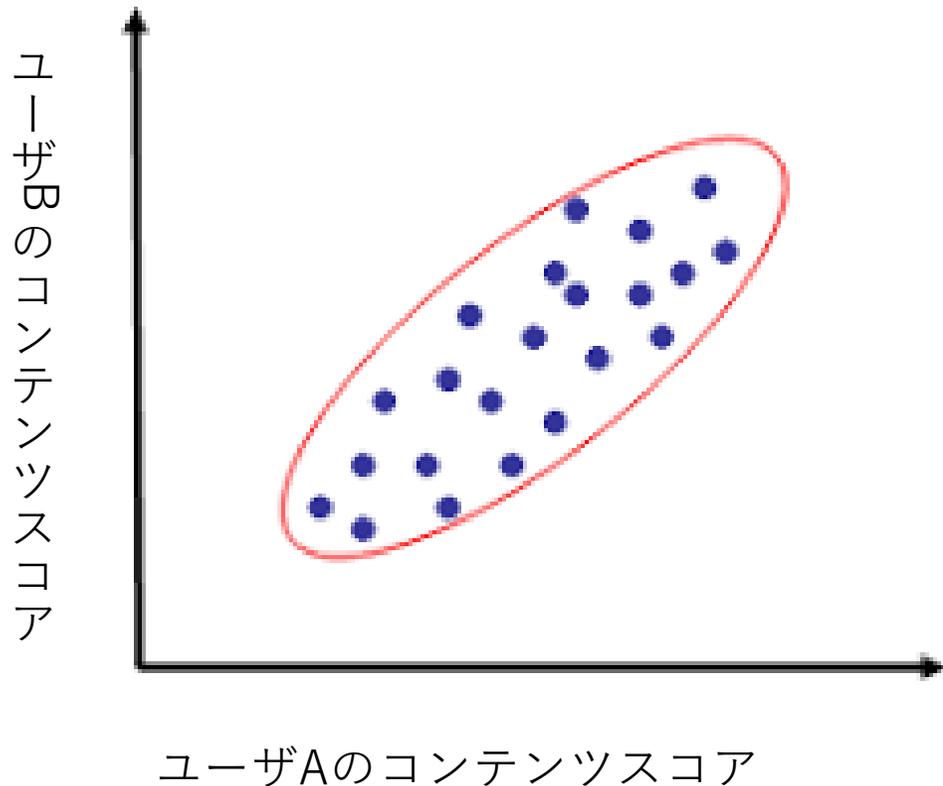
人の知識を利用するために解くべき課題

- 似ているということをどう評価する？

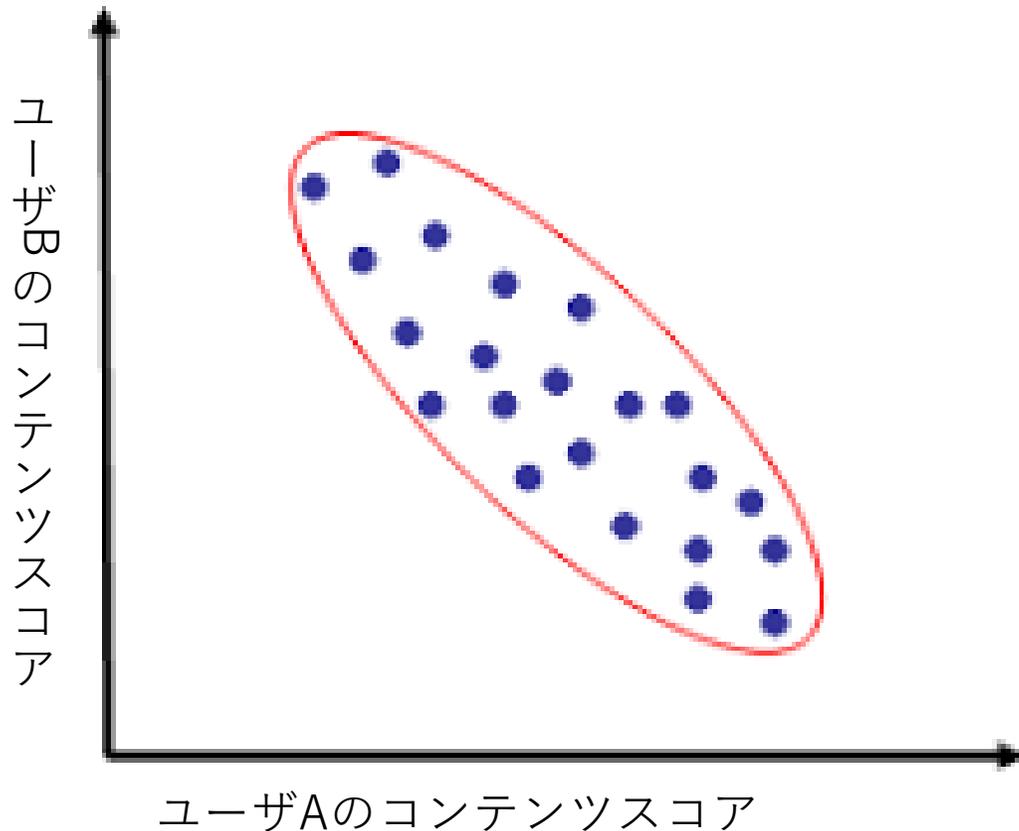


- コンテンツスコアの導入
 - →二人のコンテンツスコアの相関係数を調べる

結果予測① 個人クエリが似ている場合 コンテンツスコアに正の相関がある！



結果予測② 個人クエリが反対の場合 コンテンツスコアに負の相関がある！



- 個人クエリの類似性はコンテンツスコアの**相関係数**で評価できる。

今回使ったコンテンツスコアの計算

- プロファイリング技術 (INSOP)

	興味有 回数	興味無 回数
user name	P^i	P^u
k_1	Q_1^i	Q_1^u
k_2	Q_2^i	Q_2^u
\vdots	\vdots	\vdots
k_N	Q_N^i	Q_N^u

キーワード

今回使ったコンテンツスコアの計算

2値の入力からキーワードに

興味有=正の値

興味無=負の値

$$SKC(\alpha, k_n) = q^i(k_n) \log \frac{q^i(k_n)}{p^i} \quad (1)$$

$$-(1 - q^i(k_n)) \log \frac{1 - q^i(k_n)}{1 - p^i}$$

$$p^i = \frac{p^i}{p^i + p^u} \quad (2)$$

$$q^i(k_n) = \frac{Q^i(k_n)}{Q^i(k_n) + Q^u(k_n)} \quad (3)$$

今回使ったコンテンツスコアの計算

$$SKC(\alpha, k_n) = q^i(k_n) \log \frac{q^i(k_n)}{p^i} \quad (1)$$

コンテンツスコア

割り振られた値の
総和

$$-(1 - q^i(k_n)) \log \frac{1 - q^i(k_n)}{1 - p^i}$$

$$p^i = \frac{P^i}{P^i + P^u} \quad (2)$$

$$q^i(k_n) = \frac{Q^i(k_n)}{Q^i(k_n) + Q^u(k_n)} \quad (3)$$

今回使ったコンテンツスコアの計算

- プロファイリング技術 (INSOP)

2値の入力からキーワードに

興味有 = 正の値

興味無 = 負の値

user name	P^i	P^u
k_1	Q_1^i	Q_1^u
k_2	Q_2^i	Q_2^u
\vdots	\vdots	\vdots
k_N	Q_N^i	Q_N^u

コンテンツスコア

割り振られた値の

総和

$$SKC(\alpha, k_n) = q^i(k_n) \log \frac{q^i(k_n)}{p^i} \quad (1)$$

$$-(1 - q^i(k_n)) \log \frac{1 - q^i(k_n)}{1 - p^i}$$

$$p^i = \frac{P^i}{P^i + P^u} \quad (2)$$

$$q^i(k_n) = \frac{Q^i(k_n)}{Q^i(k_n) + Q^u(k_n)} \quad (3)$$

類似性(相関係数)を使った他者クエリの利用方法

相関係数



$$\bullet C^+(\alpha, a) = C(\alpha, a) + S(\alpha, \beta)C(\beta, a)$$

- ユーザ α にユーザ β の興味を取り込む

S: 相関係数 **a**: 任意のコンテンツ

C: ユーザのコンテンツ **a** に対する興味スコア

類似性(相関係数)を使った他者クエリの利用方法

正の相関がある場合

- $C^+(\alpha, a) = C(\alpha, a) + S_{\alpha, \beta} C(\beta, a)$
- コンテンツスコアは上がる

類似性(相関係数)を使った他者クエリの利用方法

負の相関がある場合

- $C^+(\alpha, a) = C(\alpha, a) + S(\alpha, \beta)C(\beta, a)$

- コンテンツスコアは下がる

類似性(相関係数)を使った他者クエリの利用方法

相関が無い場合

- $C^+(\alpha, a) = C(\alpha, a) + S_{0.}(\alpha, \beta)C(\beta, a)$
- コンテンツスコアは変わらない

実験に使用したユーザについて

- **User1**

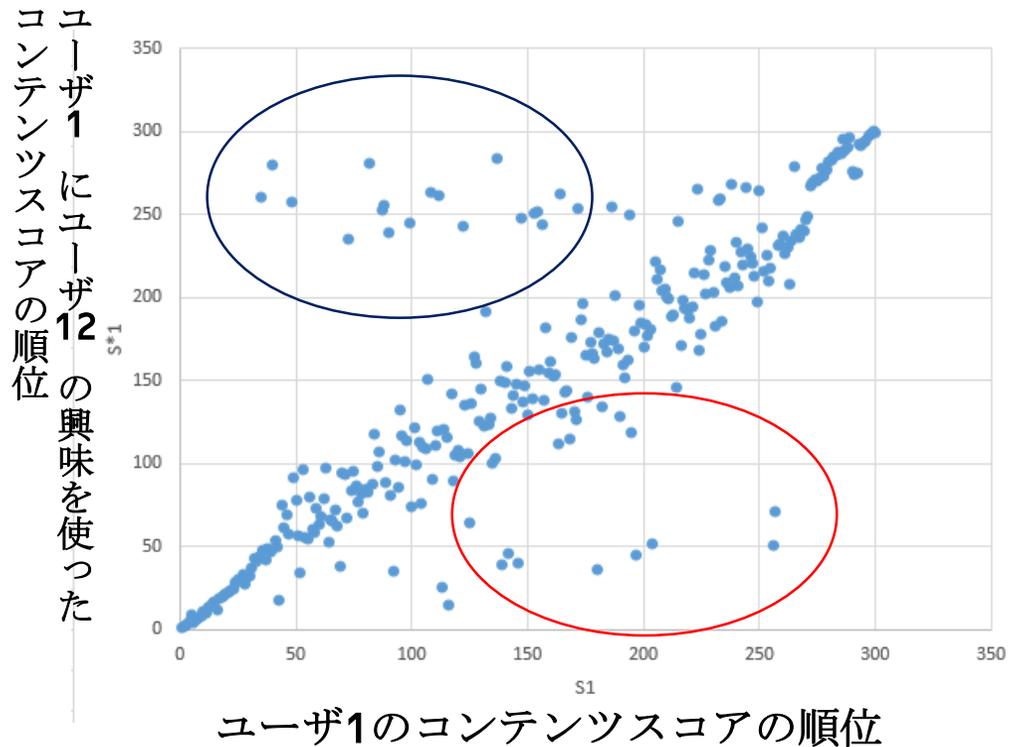
- 元々興味があるコンテンツ
- 恋愛、アニメ etc...

- **User12**

- **User1**とスコアの相関が高い人(相関係数=0.4)

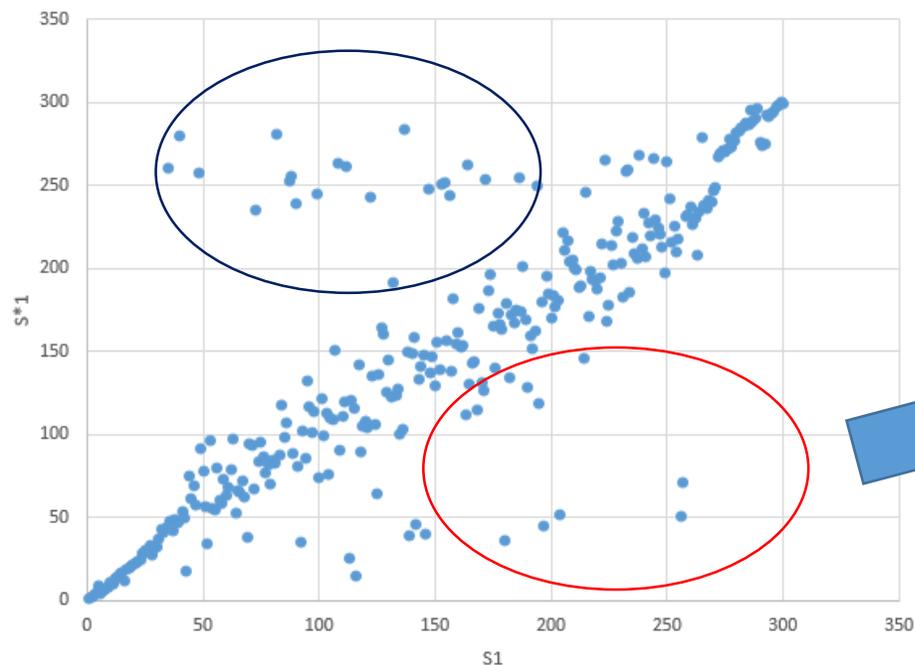
実験結果

- 他者クエリを利用する前と後での興味の変化
- 使用後上位に上がったコンテンツ (右下, 興味がある)
- 使用後下位に下がったコンテンツ (左上, 興味が無いもの)



実験結果(詳細について)

ユーザー1にユーザー12の興味を使ったコンテンツスコアの順位



ユーザー1のコンテンツスコアの順位

興味が似ている(相関係数が高い)
user1とuser12の比較

user1が興味があるコンテンツ

恋愛

- ユーザー1が見落としていた**恋愛映画**などのスコアが上昇した。
- 興味の似た他者クエリの利用でユーザーの視野拡大を手助けできそう。

まとめ

- 個人クエリを直接比較しなくても他人の興味を利用できる。
- 興味の似た他者クエリの利用でユーザーの視野拡大を手助けできることが示唆された。
- それによってユーザーが見落としていた情報の推薦が可能になる。
- それによってユーザーの視野拡大に繋がるのではないか。