

エコメンテーション

東京工業大学博士課程 3年

佐々木 祥

自己紹介



> 佐々木 祥（ささき あきら）

» 集積システム専攻（情報通信系）

» 博士課程 3 年

> 無事就職決まりました

> 研究内容

» ソーシャルブックマーク

» リコメンデーション(情報推薦)

> 協調フィルタリング

> 登録情報

» brnw (twitter, delicious.com など)

第一回発表概要



- SBMカンファレンス一回目で発表
 - SBM研究全体の変遷紹介
 - Anti-Folksonomyアプローチに基づくSBMデータからのWeb推薦
 - タグなんて飾り，名称を見ないで推薦しよう！
 - 詳細
 - 「SBM研究会」でググる→プレゼン資料
 - [twitter/brnw](https://twitter.com/brnw)のプロフィールリンク先からDL(ppt,pdf)

発表概要



▶ 前半：佐々木

- ▶▶ ソーシャルまとめサイト
- ▶▶ エコメンテーションのイントロダクション

▶ 後半：上村

- ▶▶ エコメンテーション実現のための
プロファイル選択木生成手法の提案・実装

ミニプレゼンの何か ～SBMが普及しないたったひとつの理由





ブックマーク共有？何それおいしいの？

ミニプレゼンの何か ～SBMが普及しないたったひとつの理由

終

SBMが普及しないたったひとつの理由



- ▶ ブックマークを「共有する」メリットがわからない
 - ≫ 見知らぬ人にブックマーク見せて何が楽しい？
 - ≫ そんなにブックマークするほど見てないし
 - ≫ α ブックマーカーに叩かれて
変な噂されたら恥ずかしいし . . .

- ▶ 端的に言えば、魅力が感じられない

ブックマシエアに対するメリットの付与



ブックマシエアは必要とされていない？

- ▶ 否， URL列を提供するサイトは少なくない
 - ▶▶ 動画・壁紙などのまとめブログ・wiki

多くの人が利用するまとめサイト

SBMを売り込む（周知する）チャンス！！！！

だけど選択肢として， SBMが選ばれていない

何故SBMが選ばれていない？



- ▶ メリットがない・めどい
 - » アフィリエイト, 閲覧回数などの人気評価
 - » 登録が煩雑 (vs リンクのコピペ, wiki協調)
 - ▶ まとめてブックマリスト化できない
- ▶ ブックマリストを見せる仕組みになってない
 - » 「人気コンテンツ」 = ブクマ単体にフォーカス
 - » 関連するコンテンツ群を系統的に提示できない

点（ブクマ）から線（ブックマリスト）へ
SBMの見せ方をシフトできないか？

点から線へ 絶対タグから関係タグへ

ソーシャルまとめサイト

ブクマ間を繋ぐ相対的な「関係タグ」

コンテンツを始点とした関連コンテンツの登録 / 検索

多ユーザの意見を集約したまとめサイト作成



一次ソースと
関連コンテンツの
関連付け

コンテンツX

酒井容疑者
逮捕

常習性あり！
芸能界追放！

対立

コンテンツY

のりピー逮捕

かわいそう…
復帰したら
応援汁！

関係性を
選択して
情報収集



コンテンツX

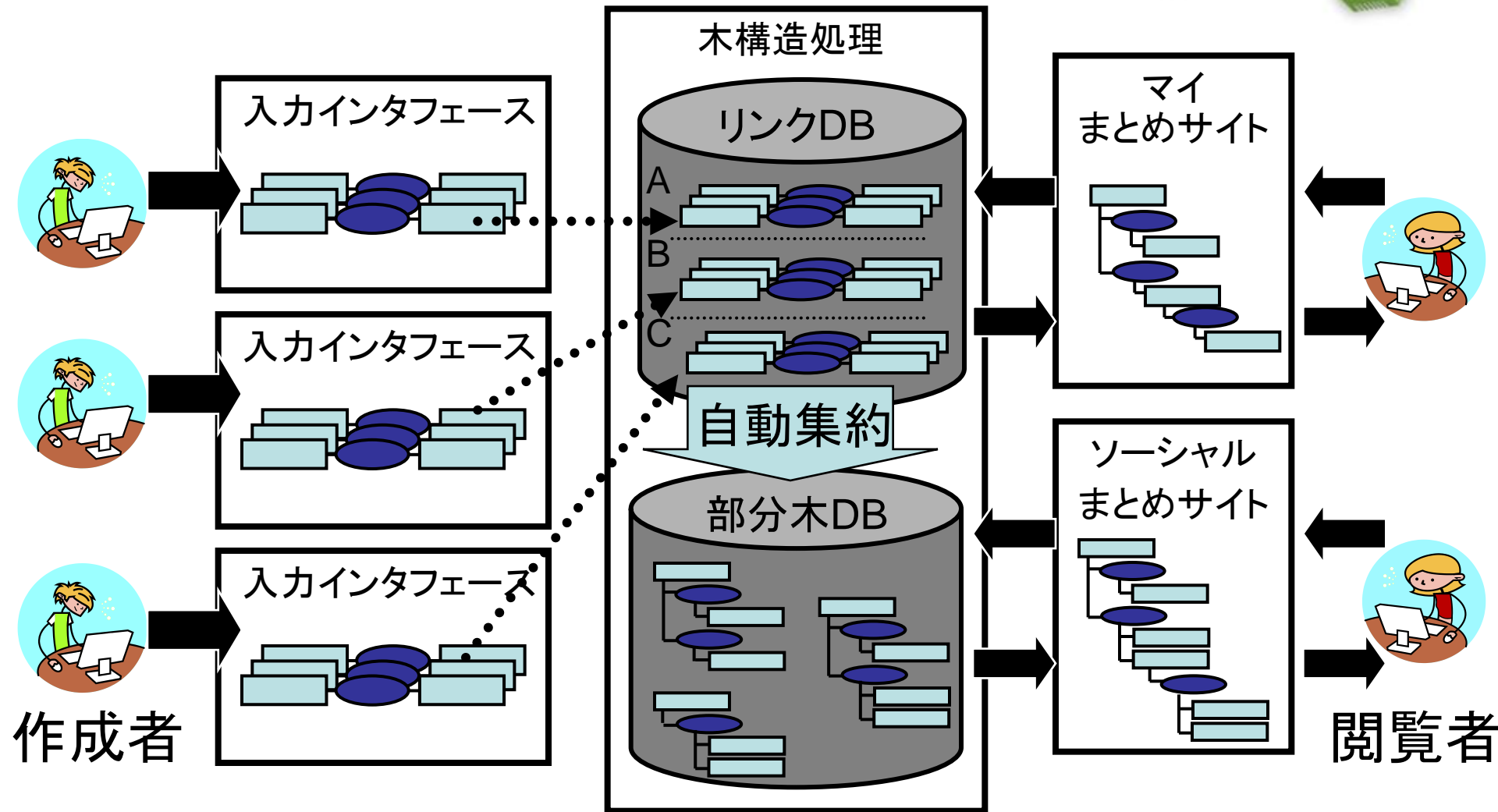
対立

同意

コンテンツY

コンテンツZ

ソーシャルまとめサイト概要



ソーシャルまとめサイト

TOP

既出URL 表示 非表示

+ Rank 5 麺屋 婆娑羅 (めんや ばさら) /大岡山/ラーメン、拉麺、支那そば[食ベログ.com]

× 二郎系,口コミ,山盛り,大盛りラーメン

+ Rank 3 基乃@大岡山 - ラーメン通信

むさしさんトコで詳細が出てて気になってただけど、大崎さんのブログ、とら会 読んで訪問決定〜^^大田区に垂流二郎新店登場っ！ つつーことで、行ってみた。

× 評価,詳細

+ Rank 0 ワンコイン的食べ歩き生活。:つけ麺、始まりました！ 基乃@大岡山 - livedoor

Blog (ブログ)

看板は らーめん つけめん なのに、ずっとラーメンだけで営業していた基乃@大岡山。3月3日から、つけ麺が始まりました。種類は3種類。つけ麺(醤油) 700円、味噌750円、辛味噌800円。麺大盛り無料で、特盛りは50円増しとなっています。

× 評価

+ Rank 0 大岡山 麺屋婆娑羅で中ラーメン / おなかぶよぶよ

職種がエンジニアなので、理系卒が多い職場です。だからなのか、圧倒的にメガネ男子が多い。男子というかおっさん。

+ Rank 0 ぼぶのラーメン紀行: 【266】 麺屋 婆娑羅@大岡山(ラーメン)

今まで4回振られているので、いい加減いやになったけど、なんとも悔しいので、またやってきた。それにしてもこの辺は駐車場が少ない。駅ビル建設中なのもあるが、コインパークの絶対数がない。まあ、今日はたまたま停められたけど、停められなくて行かなかったこともあるし。。

+ Rank 4 ライダーのラーメンと共にある日々: 麺屋 婆娑羅(大岡山)

先日店の前まで行ったが臨時休業のため伺えなかった店、今回時間が取れそうだったので再度行ってきた。今回は休みでは無さそうだが、店に着いたのは11:20のため店はまだ空いていない。近くで時間をつぶして11:35に再登場、どうやら最初の客だったようで、入ったときに店内の電灯を点けていた。

× ラーメン

+ Rank 0 大岡山メシハンター「麺類」



- ブクマリスト「見せる」工夫を
 - ブクマ間の関連づけをできるように
 - 選択的にブクマリストを収集できるように

新しいSBMサービス像を提案

続いて、SBMデータの再利用方法

> ずばり、リコメンデーション！

リコメンデーションシステム



▶ 有用な情報を発見するシステム総称

▶▶ ホットエントリ

▶ みんなに人気のあるコンテンツを集計したもの

▶▶ (いわゆる) リコメンデーションシステム

▶ 自分の嗜好に合ったコンテンツを探し出すもの

▶ ブクマ履歴に基づいて新たなコンテンツ発見を！

▶ リコメンデーションの技術要素

▶▶ コンテンツフィルタリング

▶ 内容（文書解析など）に基づく

▶▶ 協調フィルタリング

▶ ユーザの評価（ブックマーク登録など）に基づく

協調フィルタリング(CF)



➤ 推薦システムの多くで利用

➤ Amazon.com, Last.fmなどなど

1. プロファイル-プロファイル間の類似度計算
2. 未評価コンテンツの推薦度計算, 推薦決定

↑ プロファイル

	Alice	Bob	Clai	David
Happening	1	2	2	5
Matrix	5	3	5	2
Gattaca	-	-	3	-
Star wars	4	3	-	1
Speed racer	?	2	5	2

↓ コンテンツ

0.4 0.9 0.1

4.8

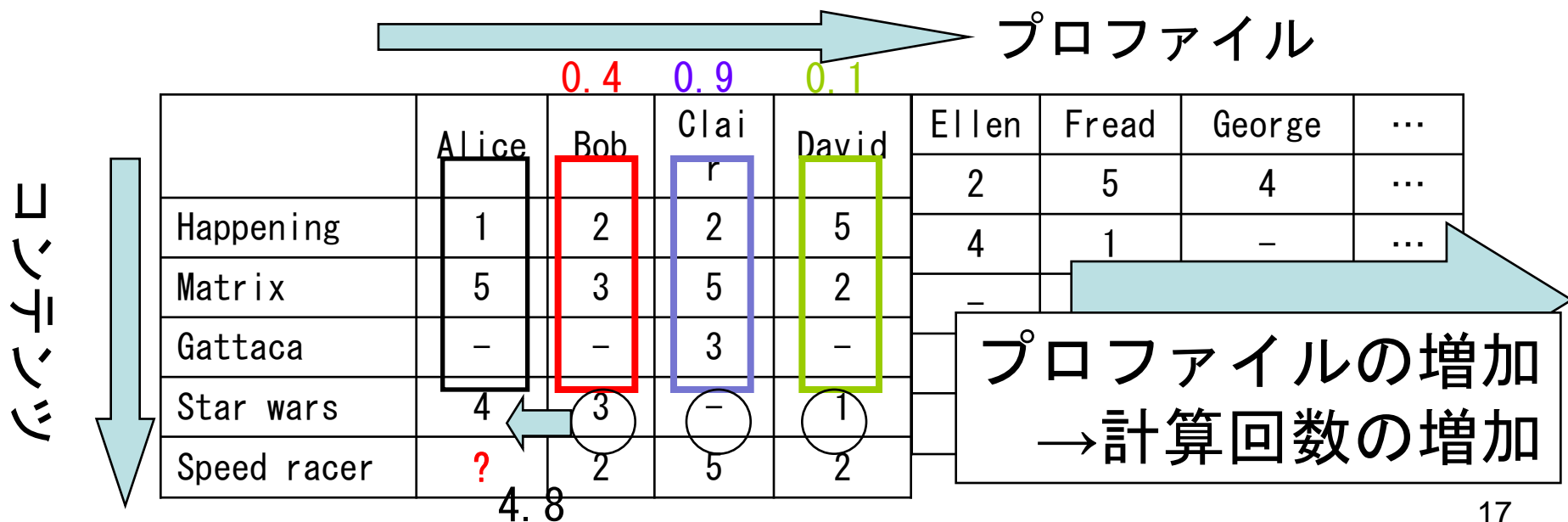
CFの問題



> 規模拡大による計算量の増加

- » プロファイルの増加
→ プロファイル間類似度の計算回数の増加
- » コンテンツの増加
→ 推薦候補の増加 → 推薦度計算回数増加

CFのスケールビリティ確保が重要



CFの分散アルゴリズム

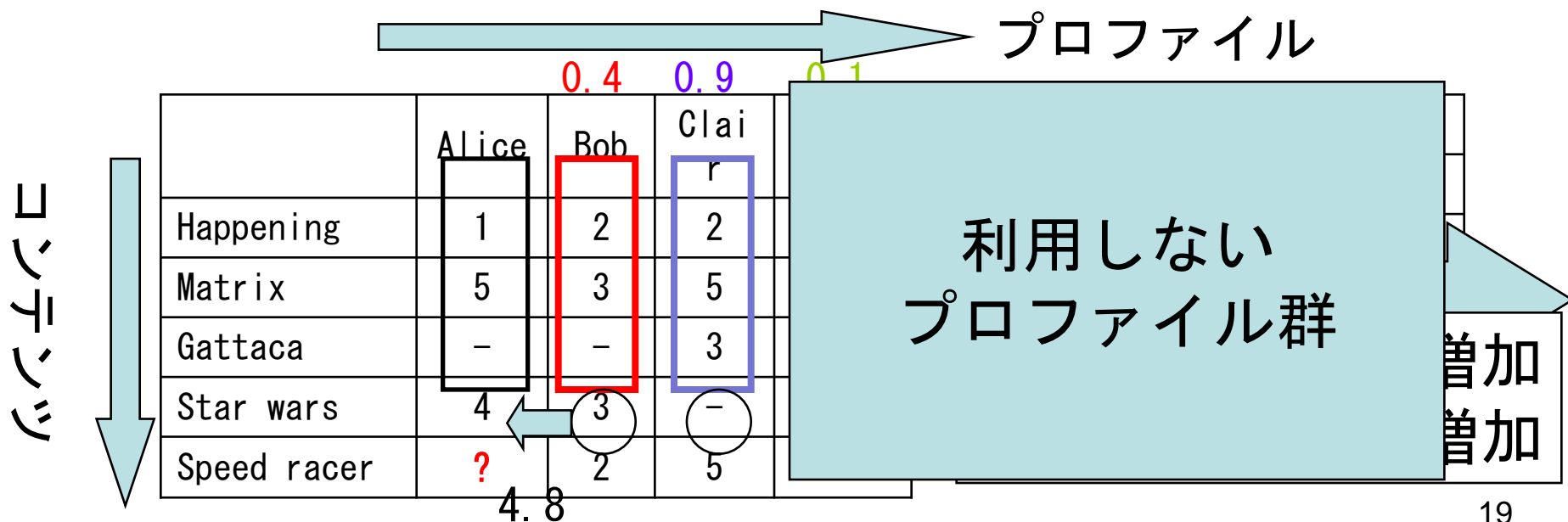


- > Latent class models for collaborative filtering
 - » 分散処理可能な協調フィルタリング
 - » 複数マシンを利用して計算時間短縮
- > Google News personalization:
Scalable online collaborative filtering
 - » グーグルニュースにおける上記論文の実装
- > 大量マシンを使った超並列計算 . . .
まあ、グーグル様ならできるよねえ . . .
 - » そもそも計算量は削減されない
 - » 計算量そのものの削減を可能とするCFはできないか？

利用するプロファイルの削減によるCFの計算量削減という試み



- CFに利用するプロファイルを削減
 - 類似プロファイルのみ用いれば十分な推薦精度
 - 類似プロファイルのみを発見できる方法を提案
 - 少量の類似度計算から類似プロファイル発見木を作成



エコメンテーションのススメ



- ▶ リコメンテーション+エコ=エコメンテーション！
 - ▶▶ 流行らせたいです (^ q ^)
- ▶ エコノミー
 - ▶▶ リコメンテーションシステムに対する要求
 - ▶▶ 推薦の性能（精度，ユーザ満足度→リターン）
 - ▶▶ 導入コスト（手間，お金，マシン場所）
 - ▶▶ コスト対効果がよければ導入される，はず
- ▶ エコロジー
 - ▶▶ CO2削減量も商売武器のひとつ！な時代
 - ▶▶ 稼働マシンを減らして，エコ主張できて，一石二鳥
 - ▶▶ 日本政府もきっと後押ししてくれます（？）

後半



- > エコメンデーション実現のための
プロファイル選択木生成手法
- > 実装が出来る人がきたー

エコメンテーション

東京工業大学修士課程1年

上村 理

自己紹介



- 上村 理（かみむら おさむ）
 - 集積システム専攻（情報通信系）
 - 修士課程1年
 - 就職先募集中（アルバイトも募集中）

- 研究内容
 - 情報検索
 - リコメンデーション(情報推薦)
 - ソーシャルブックマーク

- 普段は
 - インドアからアウトドアまで
 - ソフト開発のバイトしてます

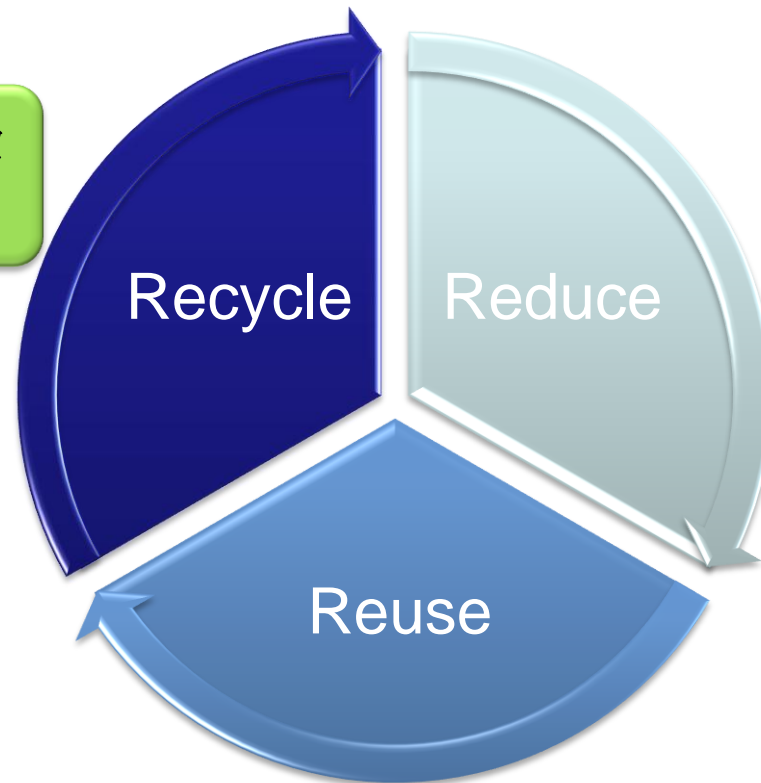
ミニプレゼンの何か ～エコについて考えてみた～



エコメンテーションにむけて ~3R~



キャッシュデータの
の構造化



計算量の削減

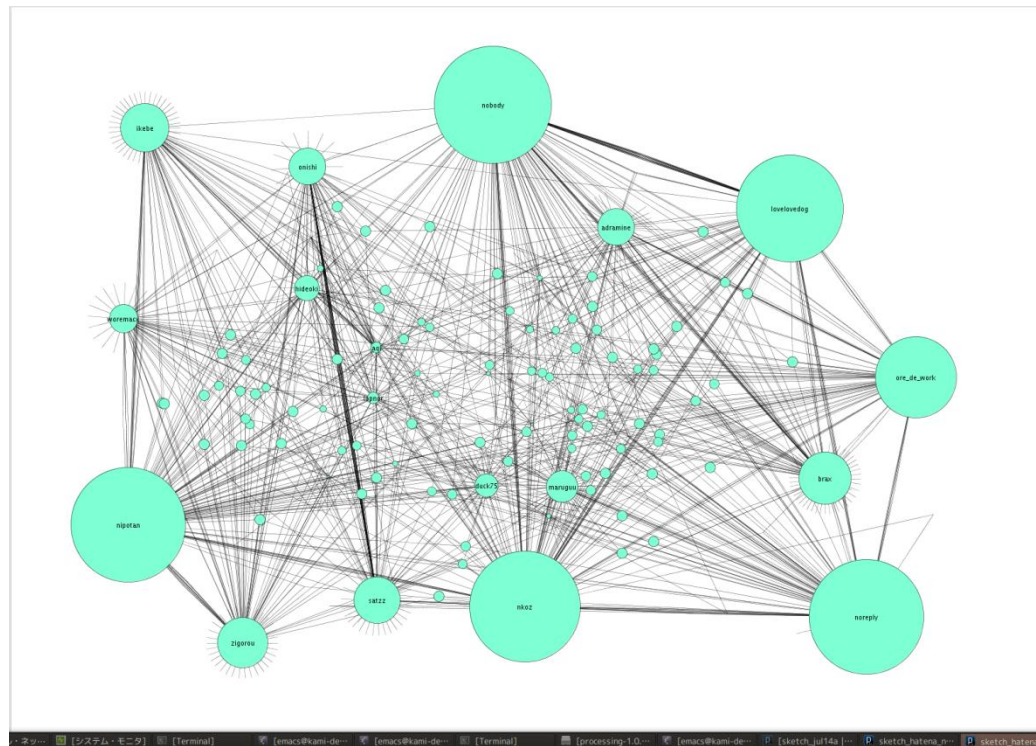
キャッシュの利用

ミニプレゼンの何か ～エコについて考えてみた～

終

はてなSBMの利用

- > お気に入り, お気に入られ
 - » ためしに少したどって描画

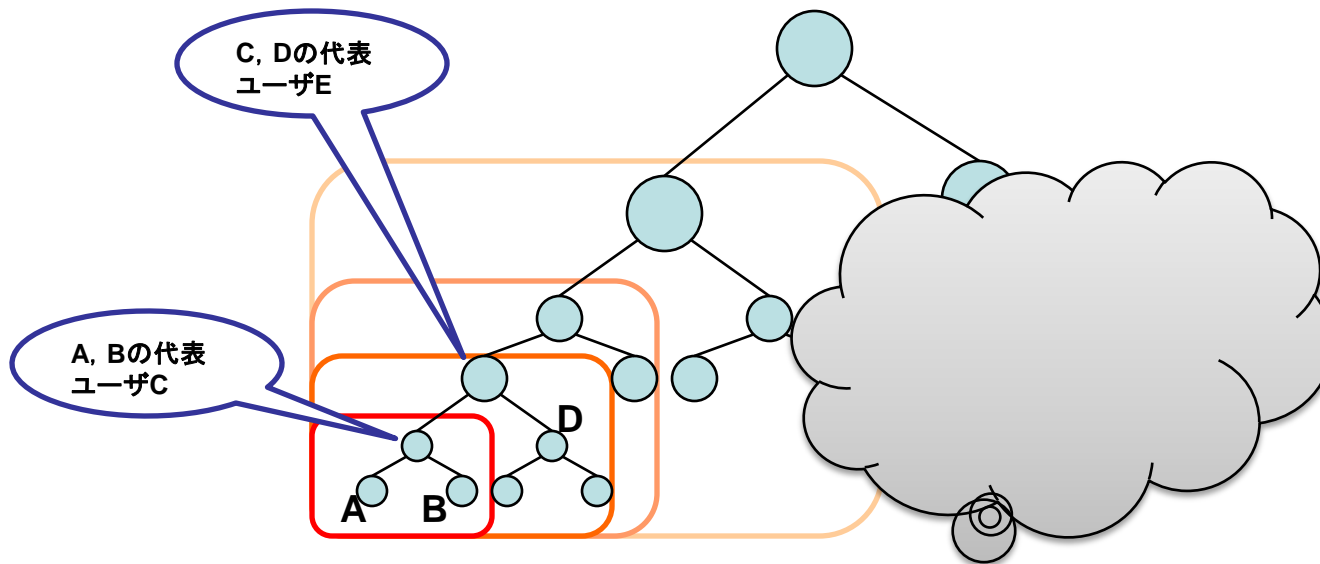


より汎用的なものを目指して



> こんなものができると

- » プロファイル選択木を構築
- » 1hopで高類似度のプロファイルを取得
- » 木を上ったり下ったりして柔軟に



基礎検討として



➤ 簡単に実験してみた

➤ 類似度計算

✓ Anti-Folksonomyアプローチ

➤ Normalized Cut(Ncut)で初期の木を構成

✓ クラスタリング手法

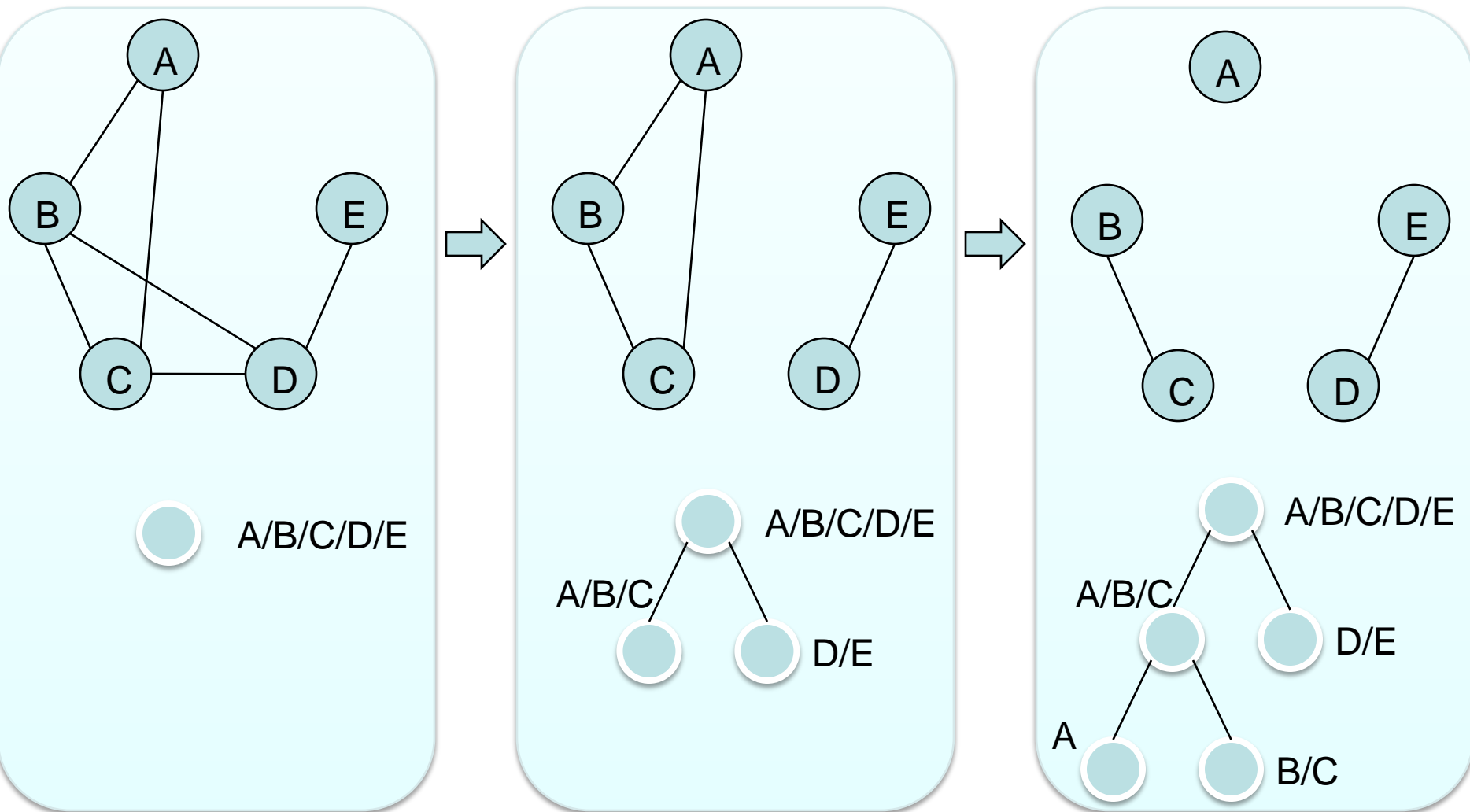
✓ 疎なエッジをカットしていく

✓ 画像処理などで利用

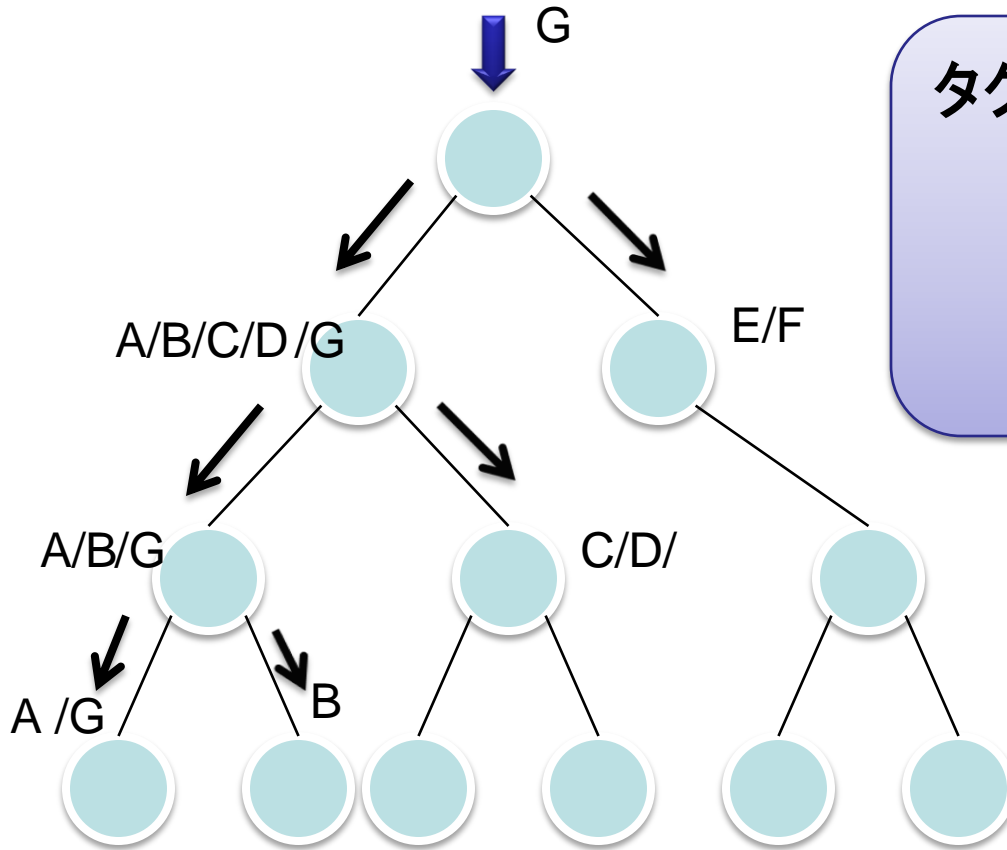
➤ ユーザ追加アルゴリズム

➤ プロファイル取得アルゴリズム

Normalized Cutの流れ



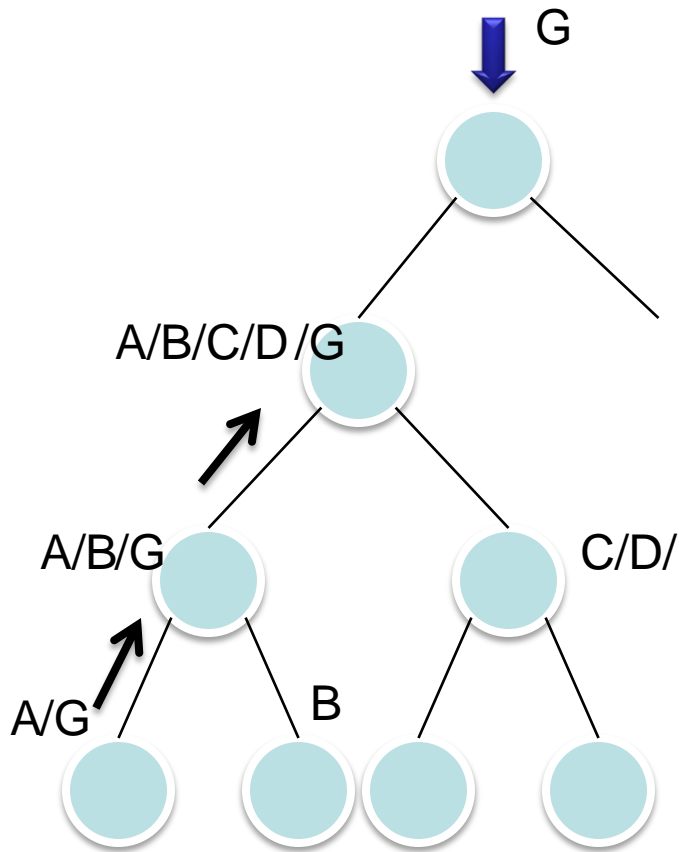
ユーザ追加アルゴリズム



タグ集合の比較

$$\frac{T_i \cap T_j}{T_i \cup T_j} \quad (i \neq j)$$

プロフィール取得アルゴリズム



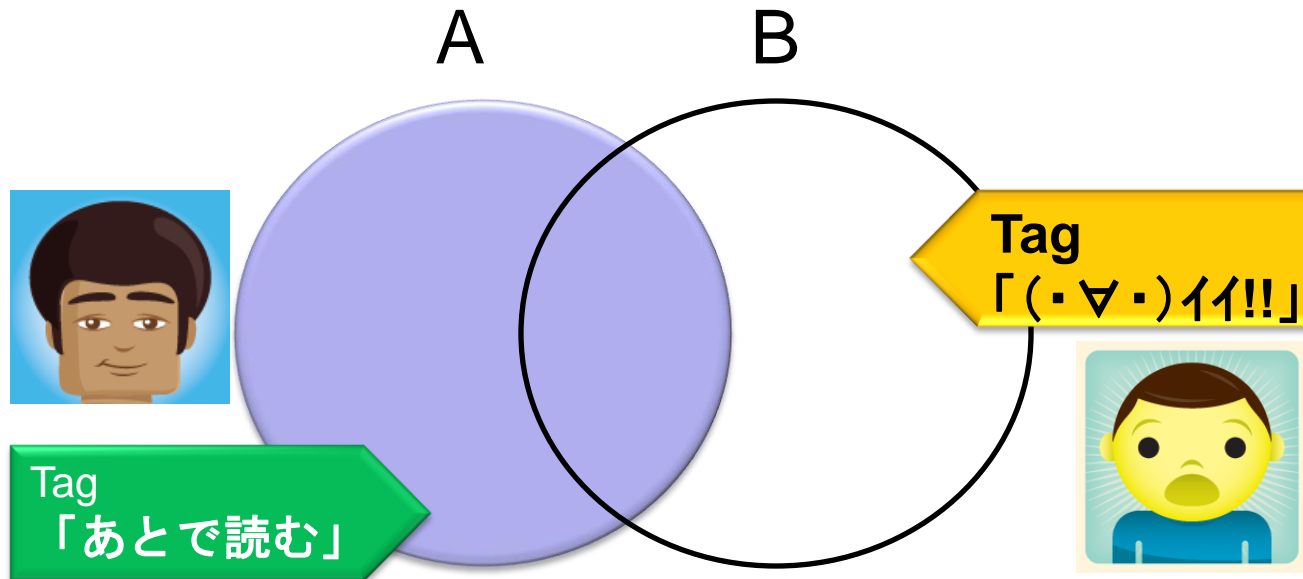
1. ホップ数を決める
2. ホップ数分ホップしたノードのプロファイルを利用

実験方法



> 小規模データセット

- » はてなブックマークのデータを利用
- » 1ユーザ1タグ集合として50人分のデータ

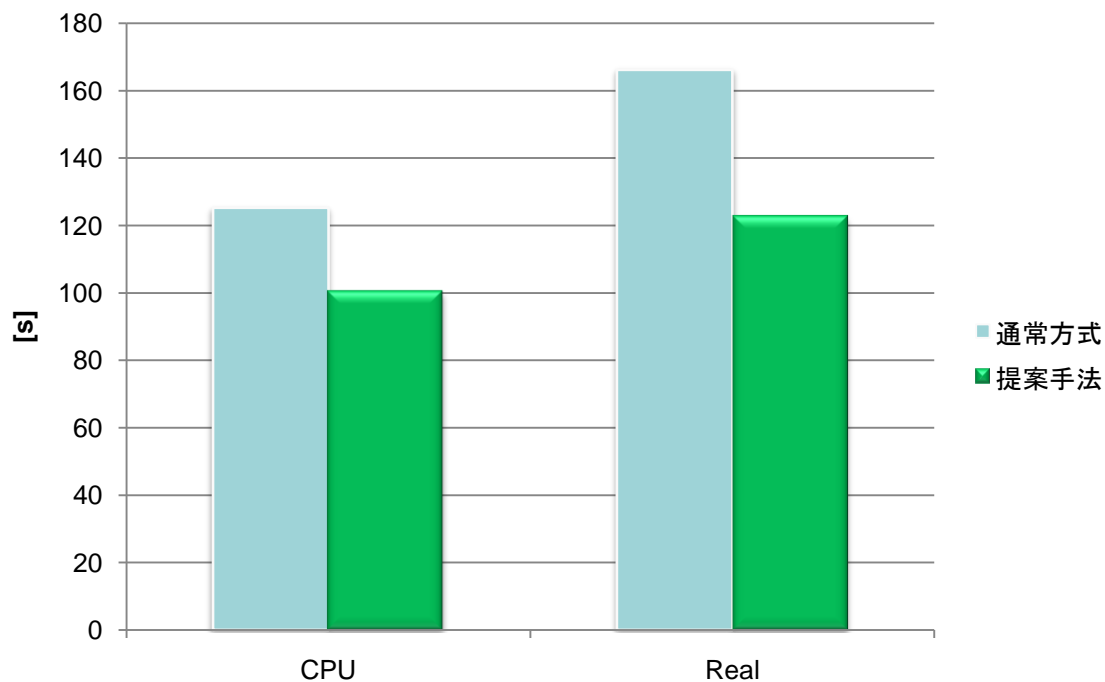


結果

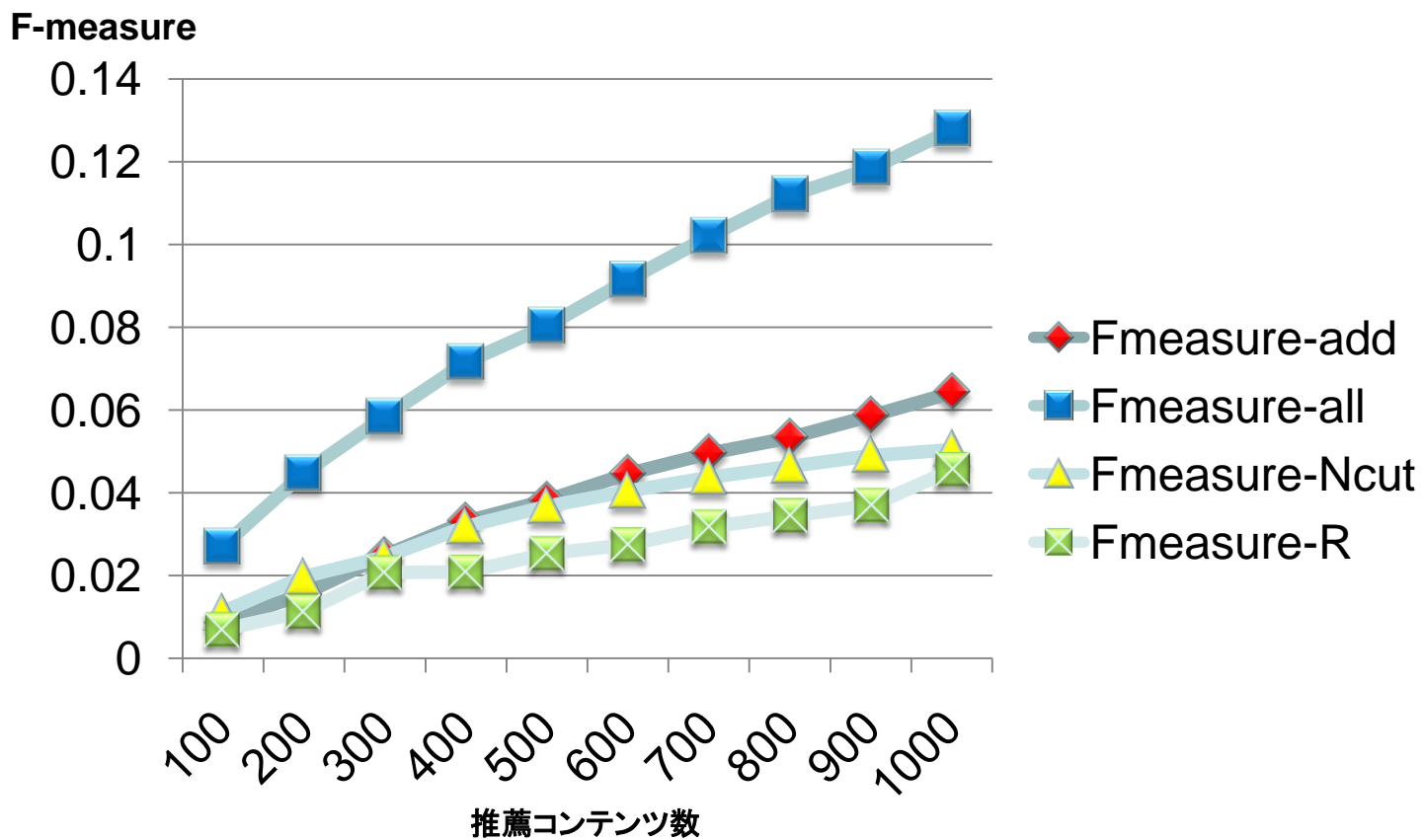


> 実行時間の測定

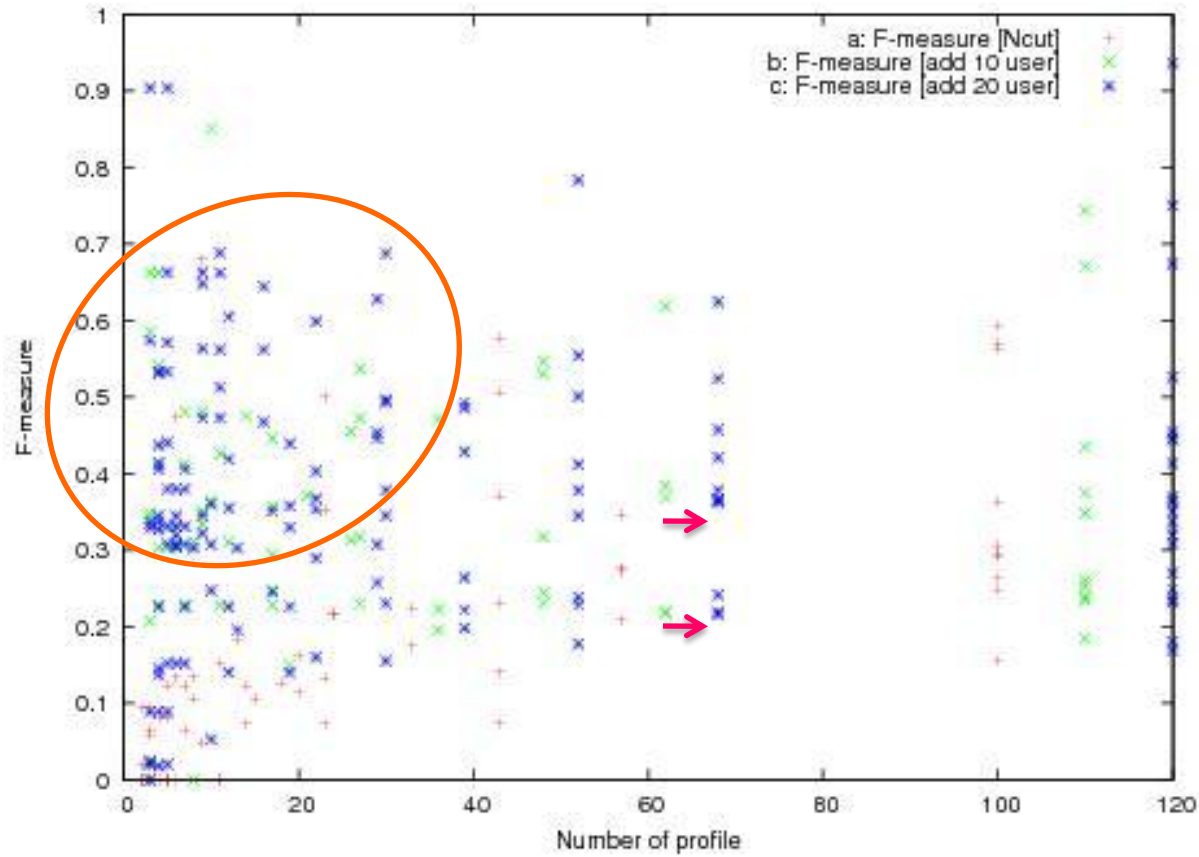
- » 50人に推薦
- » 実時間で約25%削減



結果



結果



問題点と今後



> 実験環境が難しい

» より実環境に近い中規模データセットがほしい

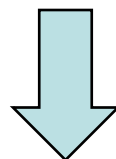
> アルゴリズム改良&追加



リコメンデーション

+

エコ



エコメンデーション