

大規模ソーシャルグラフマイニングの マーケティング活用事例と今後の展望

株式会社ホットリンク 榎 剛史



ビッグデータ分析技術ワークショップ



1.当社について



自己紹介

榎 剛史 (株)ホットリンク R&D部部長

東京大学 客員研究員

■ 興味領域

- Social Media Mining
- Computational Social Science
- Social Network Analysis
- Natural Language Processing

■ 経歴

- 2006年:修士号(電子情報学)取得
- 2006～2009年:東京電力にて勤務
- 2009年10月:博士課程入学／ホットリンクでバイト
- 2013年12月:博士号(技術経営学)取得
- 2014年～2015年:東京大学 特任研究員
- 2015年～現在:現職





会社概要

社 名
本 社

株式会社ホットリンク

東京都千代田区富士見1-3-11
富士見デュープレックスビズ5階

設 立

2000年6月26日

資 本 金

1,248百万円(2017年3月末時点)

株 式 市 場

東京証券取引所マザーズ

代 表

代表取締役社長 内山 幸樹

事 業 内 容

ソーシャル・ビッグデータの分析・販売事業
クラウドサービス事業
インバウンドプロモーション支援事業など

連 結 子 会 社

株式会社トレンドExpress(100%子会社)
EFFYIS、inc。(100%子会社)
流行特急(100%中国小会社)

連 結 従 業 員 数

120名 (2016年6月末現在)

わたしたちは、
データとAIで
意思決定をサポートする

Big Data



Social



AI

会社概要

ソーシャル・ビッグデータを活用し、
「データとAIで意思決定をサポートする」ことを目指し、
マーケティングに関わる事業を運営・提供しています。

ソーシャル・
ビッグデータ
解析ツール事業

クチコミの
マーケティング活用



ネット上の
リスクモニタリング



SNSアカウント活用
の効率化



ソーシャル・
ビッグデータ
流通・販売事業

24種類の口コミデータ
流通・販売



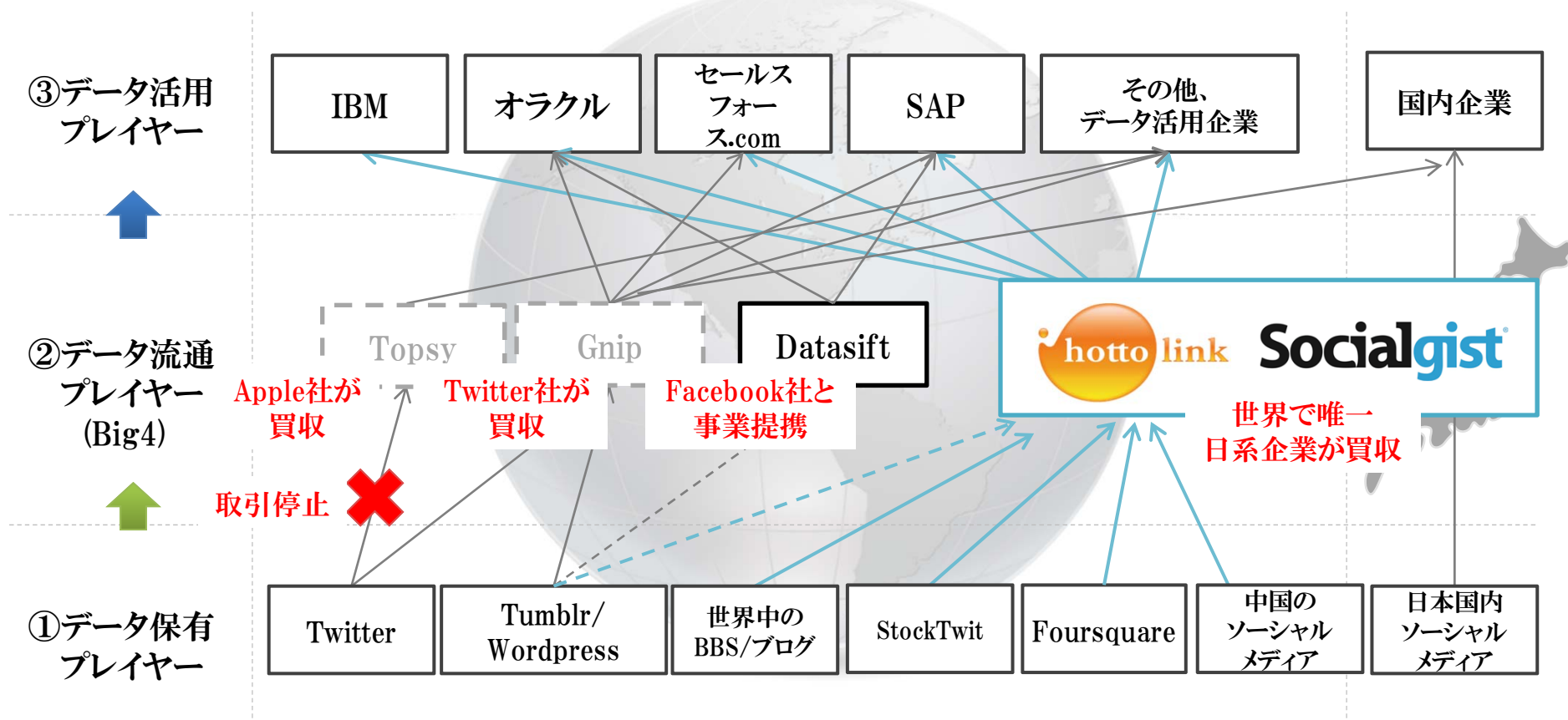
クロスバウンド・
マーケティング
支援事業

訪日中国人の
現状把握



弊社の特色

日本のソーシャル・ビッグデータに加え、世界のソーシャル・ビッグデータ流通のBig4である米国Effyis社(ブランド名:Socialgist)を買収することで世界トップクラスのソーシャル・ビッグデータ流通プレイヤーに！



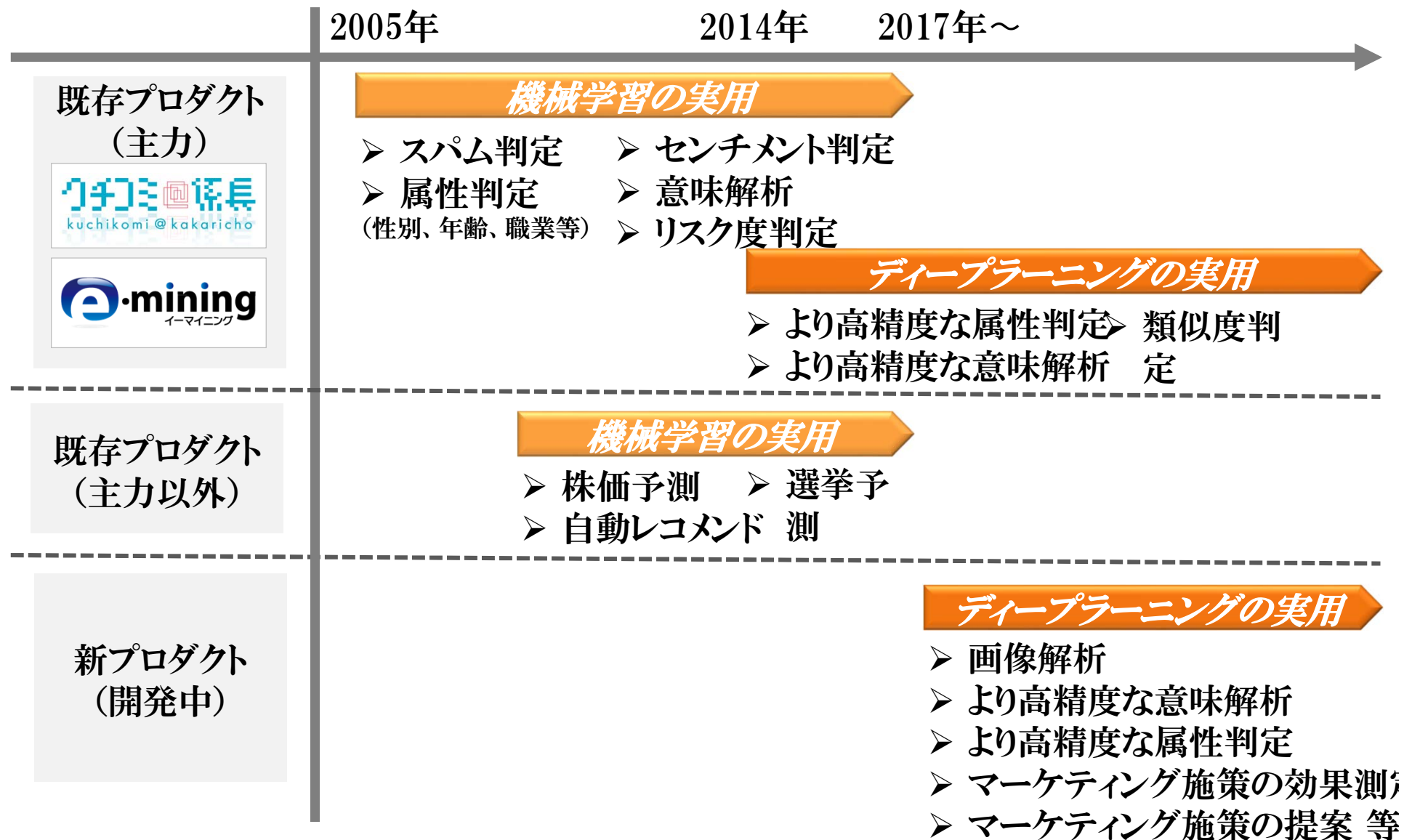
弊社の特色

■ データ量・種類(国内)

媒体分類	取得開始	取得終了	データ規模
ブログ	2006年11月1日	継続中	数億記事/年
Twitter	2011年1月1日	継続中	
2ちゃんねる	2006年11月1日	継続中	
掲示板	2007年3月25日	継続中	
TVメタデータ	2009年1月1日	継続中	

※直近3年分のTwitterの検索用サーバ(Elasticsearch)のみで、40台強

ホットリンクとAI(人工知能)技術の歴史



ソーシャルメディア分析ツール

くちコミ@係長
kuchikomi@kakaricho

<https://kakaricho.jp/>

● ソーシャルメディアアカウント運用ツール



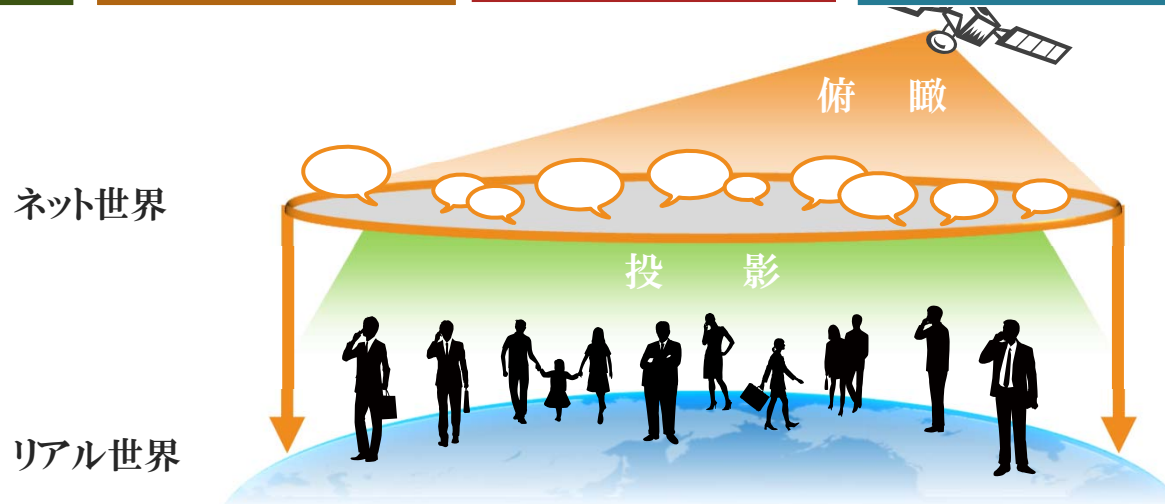
<https://hashtag-ai.buzzspreader.com>

2.ソーシャルグラフマイニング 活用事例



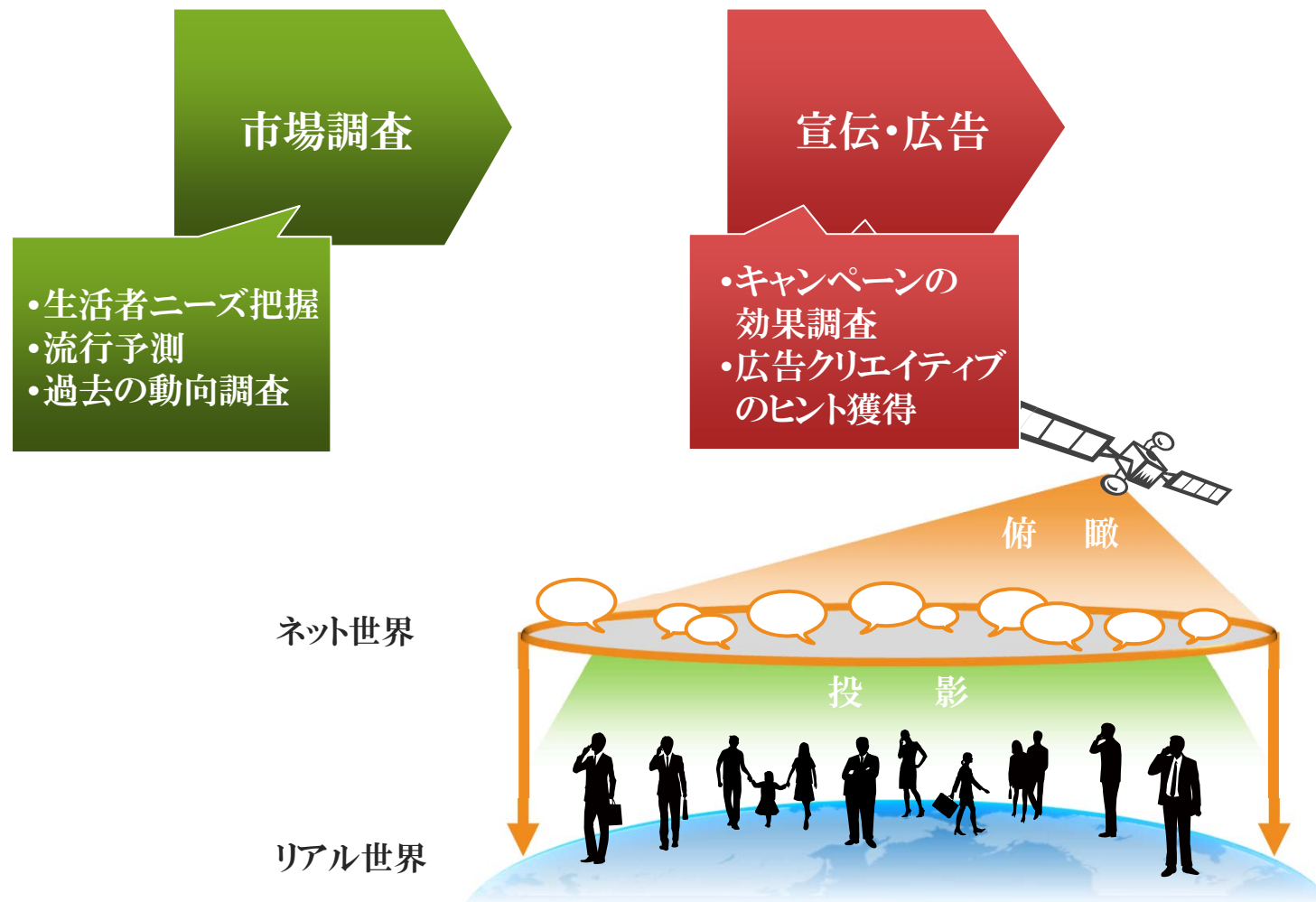
Why Social Media?

マーケティング活動のあらゆる段階に応用可能



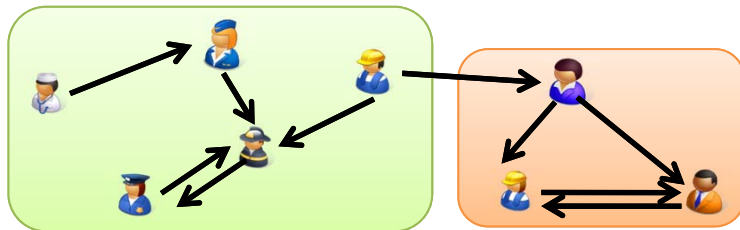
Why Social Media?

マーケティング活動のあらゆる段階に応用可能

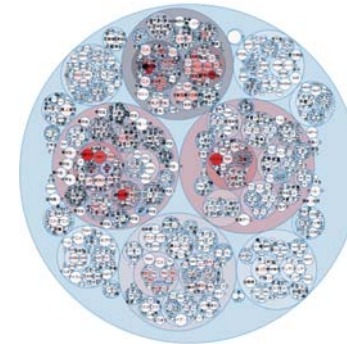


ソーシャルグラフマイニングの活用

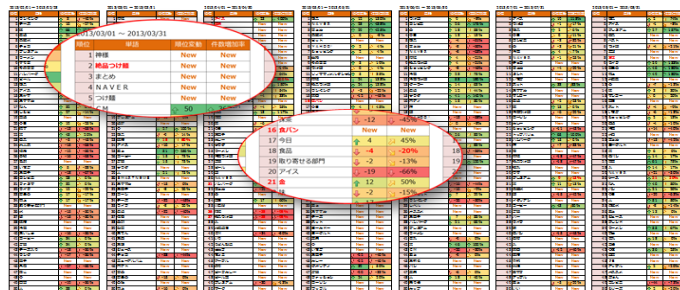
ユーザ属性の推定



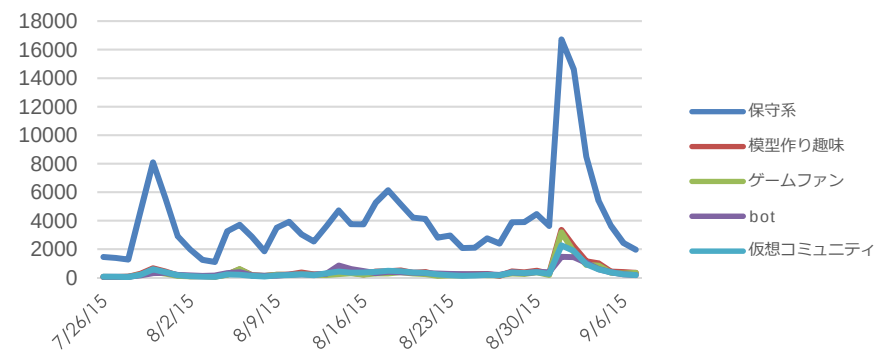
オーディエンス分析 ターゲティング



インサイト発見

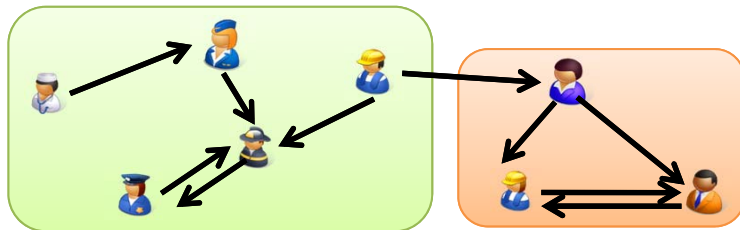


情報拡散分析

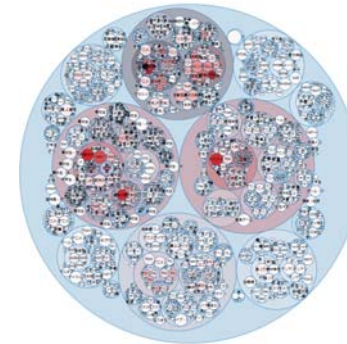


ソーシャルグラフマイニングの活用

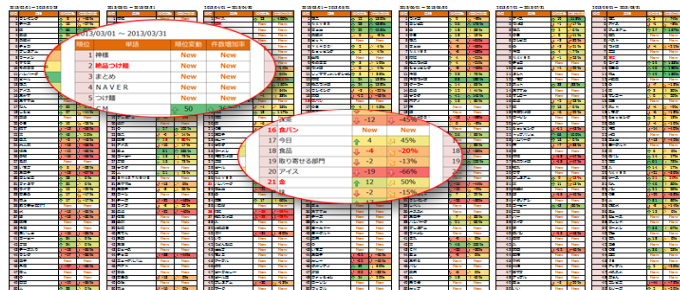
ユーザ属性の推定



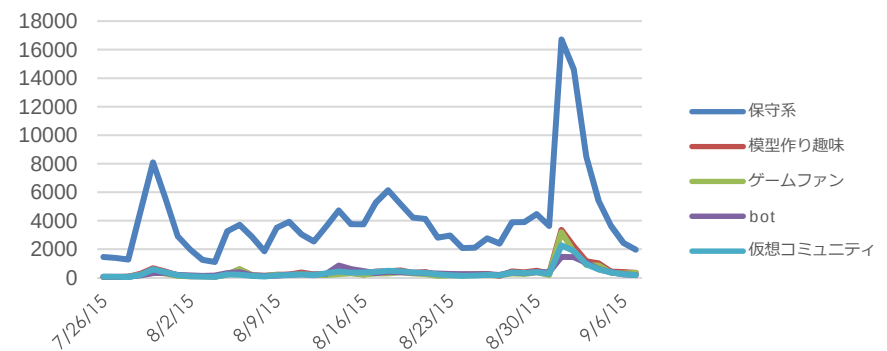
オーディエンス分析 ターゲティング



インサイト発見

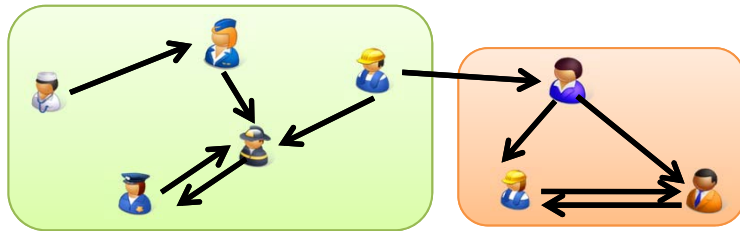


情報拡散分析



ソーシャルグラフマイニングの活用

ユーザ属性の推定



ソーシャルメディアデータからのユーザ属性推定

背景

- 社会調査やマーケティングを行う際には人物属性（デモグラフィクス）が重要である
 - － 性別・年齢・職業・地域
 - － 属性別に分析することで、全体傾向を掴むことが可能

課題

- ソーシャルメディアユーザについて、既存デモグラフィクスを推定できないか？
- ソーシャルメディア上のソーシャルキャピタルに立脚した人物属性を抽出できないのか？

ソーシャルメディアデータからのユーザ属性推定

セグメンテーションの重要性

- ユーザ属性によるセグメンテーションを行うことは、分析やターゲティングにおいて重要である



T層



M3層



F1層

視聴率の集計

社会調査の集計・分析

カスタムオーディエンス

カスタムオーディエンスを選択 | 参照

新しいカスタムオーディエンスを作成

地域 この地域に住んでいる人

日本

次を含める: 国、都道府県、市区町村、DMA、郵便番号または住所を追加

年齢 18 - 65+

性別 すべて 男性 女性

言語 言語を入力...

詳細なターゲット設定 少なくとも以下のいずれかに一致する人を含めます

利用範囲、趣味・関心、行動を追加 | おすすめ | 参照

除外

つながり つながりの種類を追加

オーディエンス

オーディエンスが小さすぎます。

オーディエンス詳細:

- 地域・居住地: 日本
- 年齢: 18歳~65歳

潜在リーチ: 24,000,000人

広告のターゲティング

ソーシャルメディアデータからのユーザ属性推定

背景

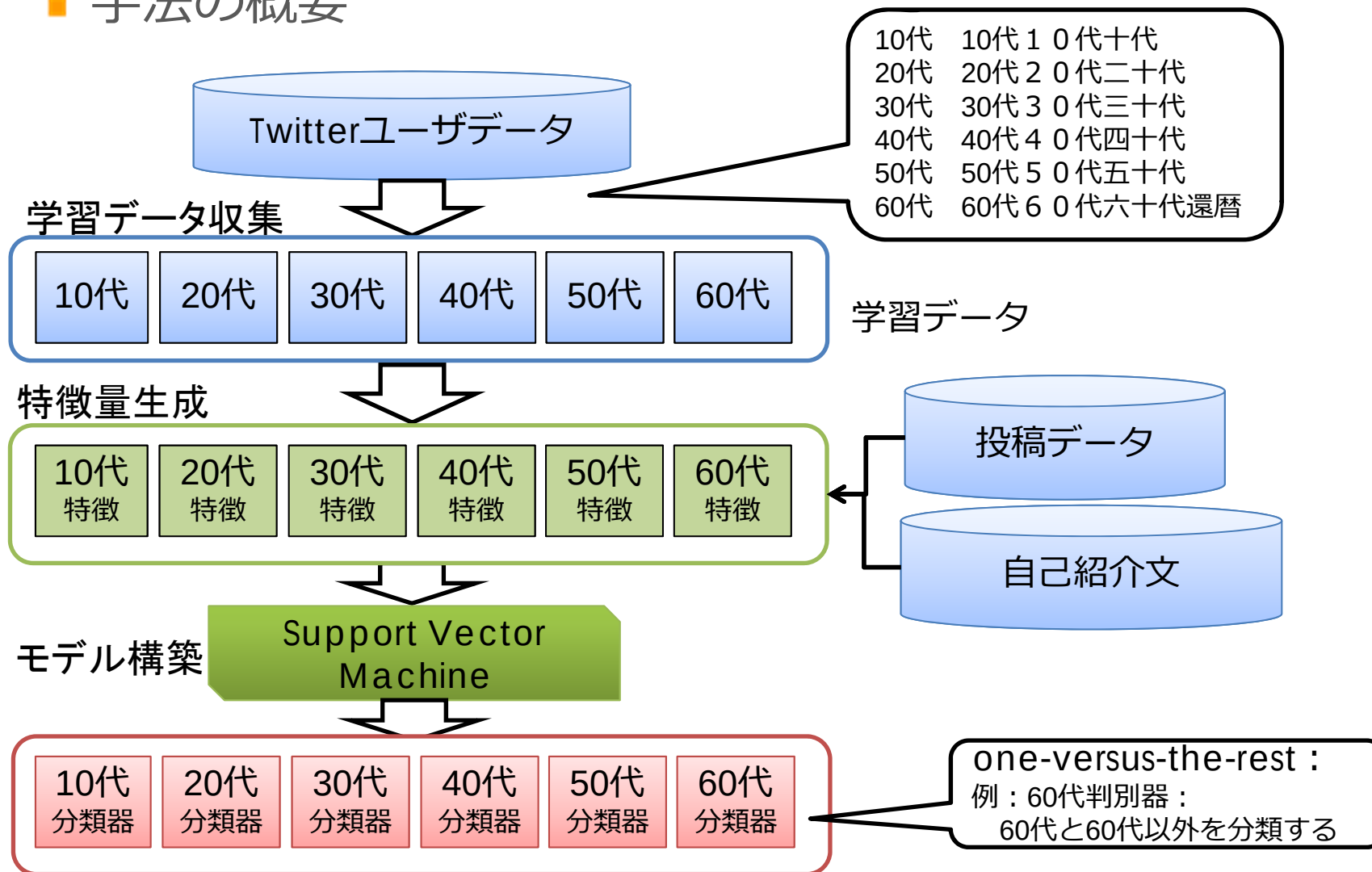
- 社会調査やマーケティングを行う際には人物属性（デモグラフィクス）が重要である
 - － 性別・年齢・職業・地域
 - － 属性別に分析することで、全体傾向を掴むことが可能

課題

- ソーシャルメディアユーザについて、既存デモグラフィクスを推定できないか？
- ソーシャルメディア上のソーシャルキャピタルに立脚した人物属性を抽出できないのか？

Twitterデータによる既存のユーザ属性推定

■ 手法の概要



Twitterデータによる既存の人物属性推定

参考：各ユーザ属性の分類に有効な特徴量

男性	女性
自己紹介：男子	自己紹介：女の子
自己紹介：愛しています	自己紹介：女性
発言：腐女子	自己紹介：キスマイ
発言：女子高生	発言：腐男子

大学生	それ以外	50代	それ以外
自己紹介：回生	自己紹介：元気	自己紹介：おじさん	自己紹介：女子
自己紹介：女子大	自己紹介：海外旅行	自己紹介：おばさん	自己紹介：ゲーム
発言：レポート	自己紹介：キ口	自己紹介：読書	自己紹介：在住
発言：サークル	発言：遠足	自己紹介：蕎麦	自己紹介：社会人
発言：履修登録	発言：職場	発言：膝	発言：(t_t)

ソーシャルメディアデータからのユーザ属性推定

背景

- 社会調査やマーケティングを行う際には人物属性(デモグラフィクス)が重要である
 - 性別・年齢・職業・地域
 - 属性別に分析することで、全体傾向を掴むことが可能

課題

- ソーシャルメディアユーザについて、既存デモグラフィクスを推定できないか？
- ソーシャルメディア上の関係性に立脚した人物属性を抽出できないのか？

インタラクションに基づくユーザ属性推定

目的

- ソーシャルメディア上のインタラクションに基づくユーザ属性を作成したい

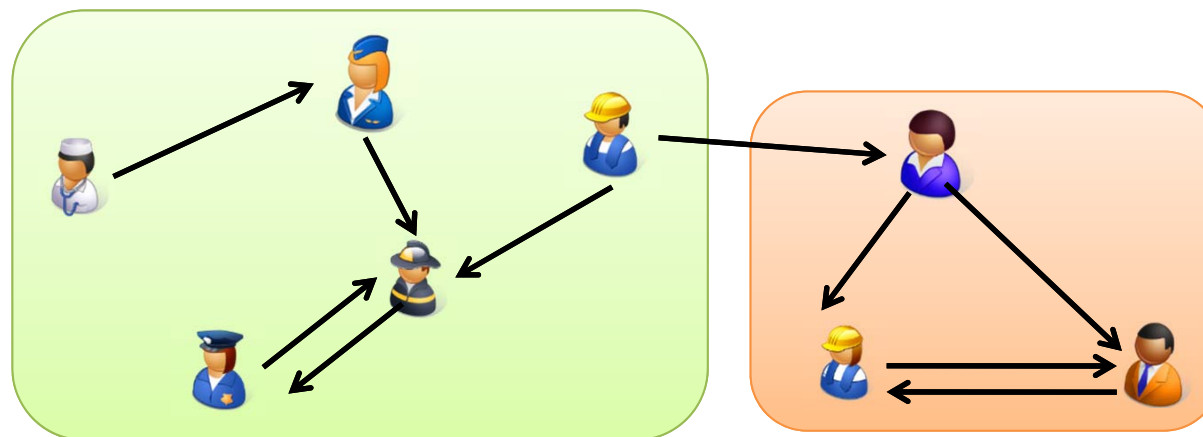
方法論

- Twitter上でのインタラクションからユーザのネットワークを構築する
- ユーザネットワークにコミュニティ抽出の手法を適用してコミュニティを抽出したのち、プロフィール文で特徴付けし、それをユーザ属性の一つとして扱う

貢献

- Twitterユーザ特有の興味・関心を反映したラベルをユーザに付与することができる

インタラクションに基づくユーザ属性推定

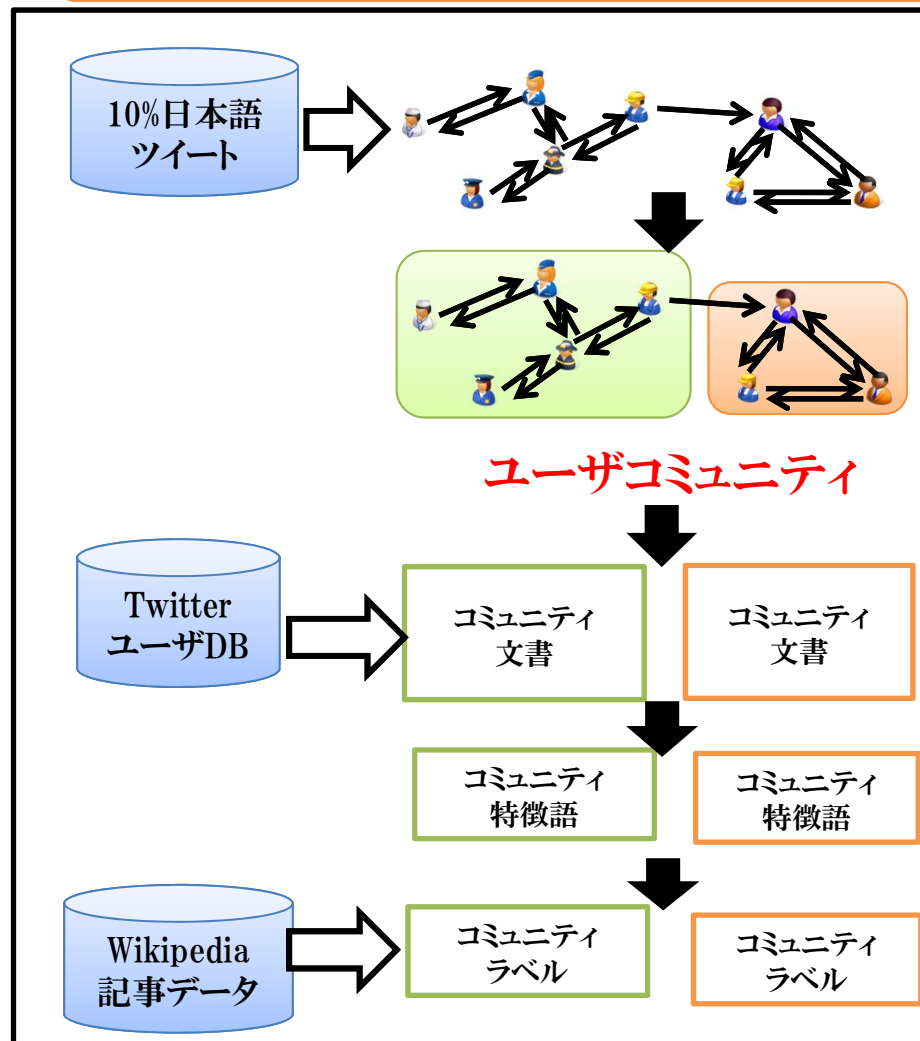


仮定

Twitter上で同じような情報を拡散する
ユーザは同じコミュニティに属する

インタラクションに基づくユーザ属性推定

ソーシャルメディア上のソーシャルキャピタルを分析することで、
定性的に理解可能なコミュニティを構成することができる



1. Twitter上の相互コミュニケーションからユーザネットワークを構築
2. ユーザネットワークからコミュニティを抽出(コミュニティごとにユーザプロフィール文を取得し、コミュニティ文書を構築)
3. 各コミュニティ文書から特徴語群を抽出(特徴群を用いて、各コミュニティにWikipediaタイトルによるラベルを付与)

インタラクションに基づくユーザ属性推定

インタラクションをベースとしたネットワークを用いることで、
定性的に理解可能なコミュニティを構成することができる

種類	人手ラベル	自動ラベル	特徴語					
地域	新潟	新潟市	野球	北越	長岡	向陽	新津	niigata
地域	福島	福島市	郡山	野球	明成	白河	安積	白河
趣味	野球	日本のプロ野球選手一覧	ファン	応援	選手	阪神	カープ	観戦
趣味	ポケモン	ポケットモンスター用語一覧	ゲーム	スマ	ブラ	パズドラ	レート	アニメ

インタラクションに基づくユーザ属性推定

インタラクションをベースとしたネットワークを用いることで、
定性的に理解可能なコミュニティを構成することができる

種類	人手 ラベル	自動 ラベル	特徴語						
職業	エンジニア	Python	haskell	python	エンジニア	vim	microsoft	engineer	ruby
職業	トレーダー	投資信託	投資	トレーダー	fx	株式	トレード	相場	先物
政治・ 思想	ネット右翼	自由民主党 (日本)	安倍	原発	反日	支持	日本	保守	政権
政治・ 思想	左翼	原子力 発電	原発	反対	nukes	racism	tpp	戦争	被曝

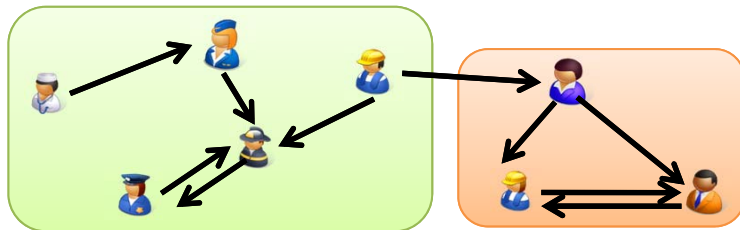
インタラクションに基づくユーザ属性推定

インタラクションをベースとしたネットワークを用いることで、
定性的に理解可能なコミュニティを構成することができる

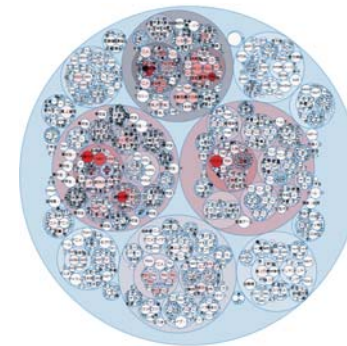
属性種類	属性ラベル
興味・関心	サッカー, 野球, アニメ(女性), アニメ(男性), ゲーム, テーマパーク, 創作(小説, 絵, 歌)
ファン	アイドル, ジャニーズ, 女性声優, 男性声優, ミュージシャン (J-POP, K-POP)
政治思想	自民党支持, 民進党支持
職業	研究者, トレーダー, エンジニア
地域高校	静岡県, 栃木県, 大阪府, 沖縄県
地域大学	東京都, 九州, 中部, 近畿

ソーシャルグラフマイニングの活用

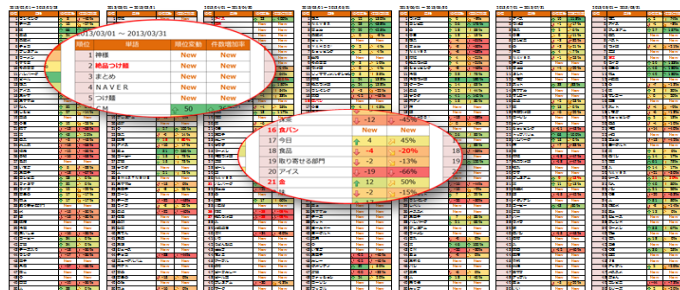
ユーザ属性の推定



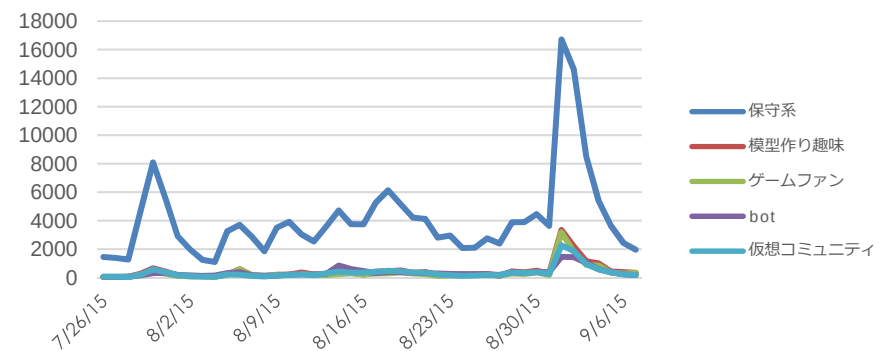
オーディエンス分析 ターゲティング



インサイト発見

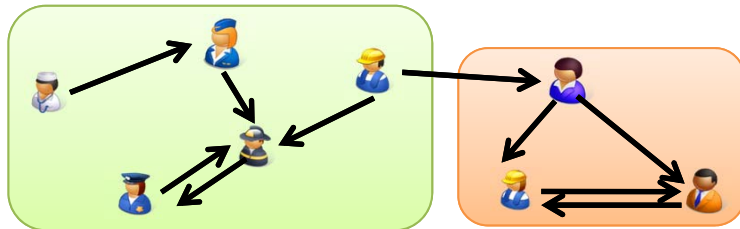


情報拡散分析

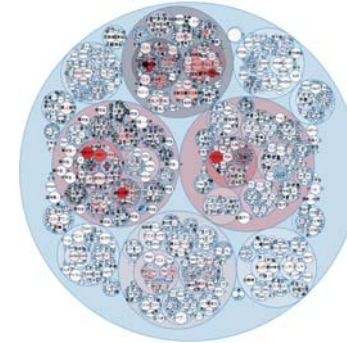


ソーシャルグラフマイニングの活用

ユーザ属性の推定



オーディエンス分析
ターゲティング



応用例：オーディエンス分析・ターゲティング

目的

- ソーシャルメディア上で、自社の製品にエンゲージしているユーザ層を分析する
- それらのユーザ層について、ターゲティング広告を配信する

方法論

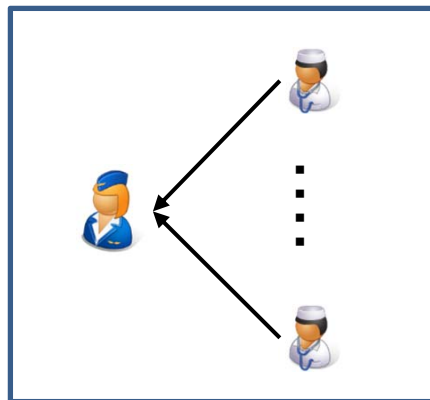
- 自社のアカウントのフォロワーや自社製品に言及したユーザ群の属性分布を明らかにする

貢献

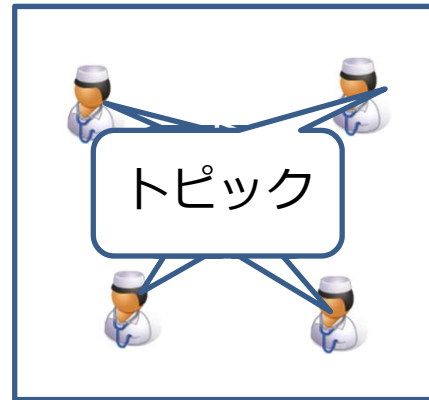
- 既存のユーザ属性によらない、柔軟な調査・ターゲティングが可能となる

応用例：オーディエンス分析・ターゲティング

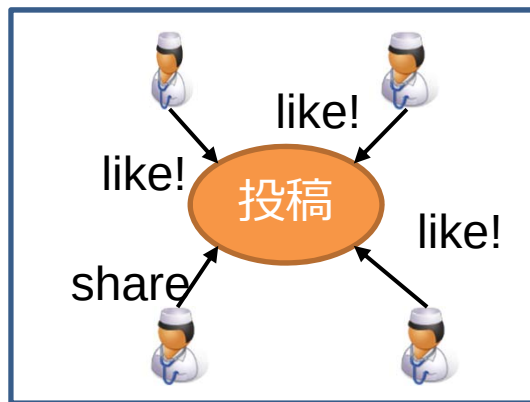
様々なオーディエンス（ユーザ集合）について、
ユーザ属性分布の作成及び可視化



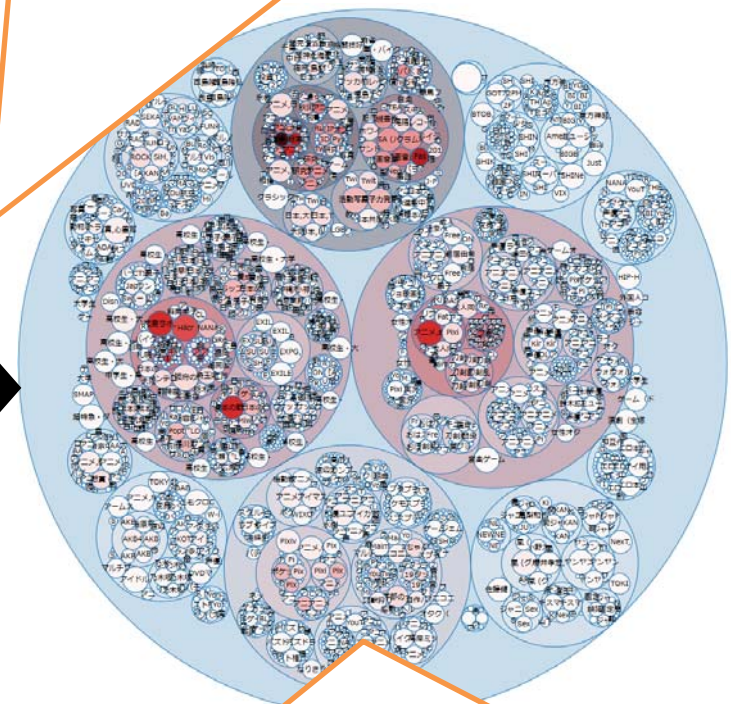
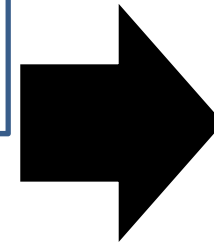
フォロワーリスト



トピックに言及



投稿に反応



オーディエンス分析結果をもとに
ターゲティング

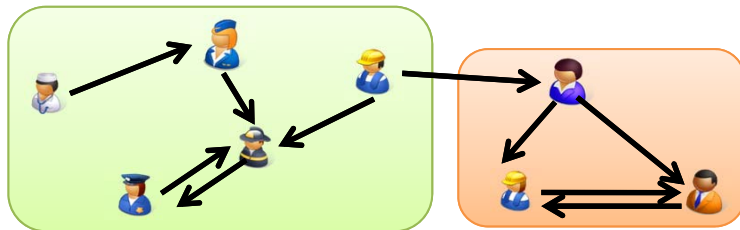
応用例：オーディエンス分析・ターゲティング

- 広告題材：BS-TBS × ホットリンク × 日清フーズ

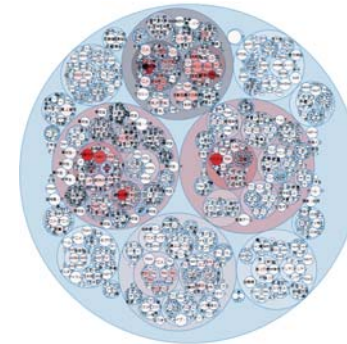


ソーシャルグラフマイニングの活用

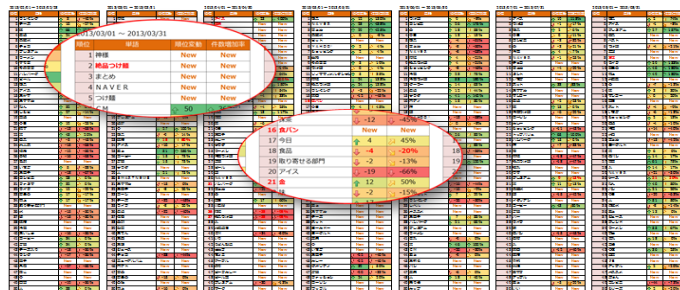
ユーザ属性の推定



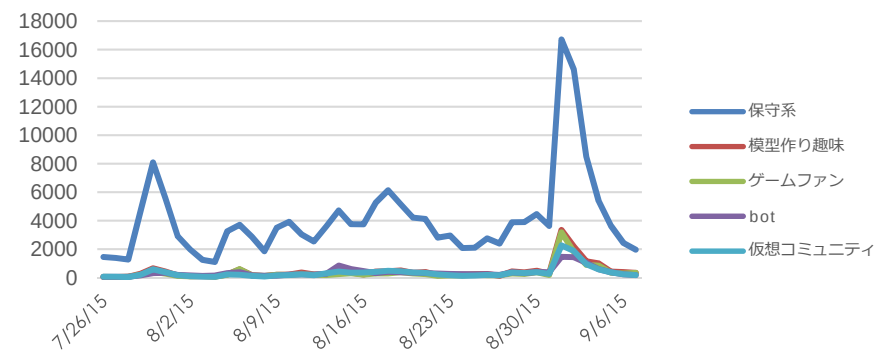
オーディエンス分析 ターゲティング



インサイト発見

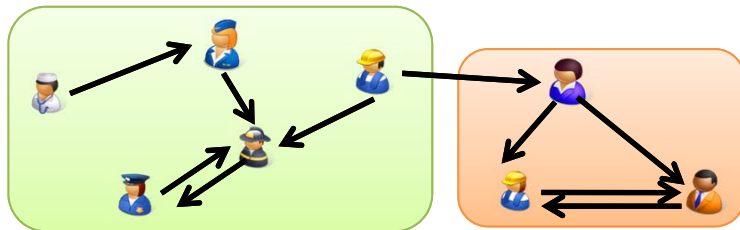


情報拡散分析

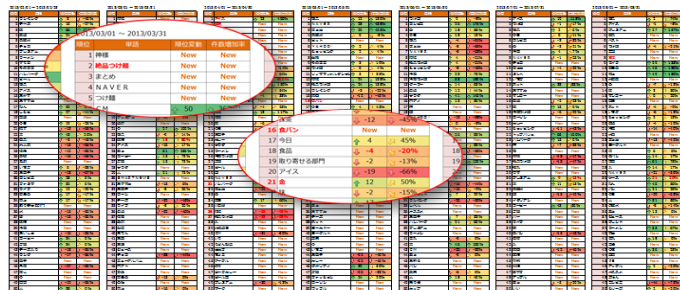


ソーシャルグラフマイニングの活用

ユーザ属性の推定



インサイト発見



応用例：ソーシャルメディアからのインサイトの発見

目的

- 特定のユーザ群でのインサイトを発見したい

方法論

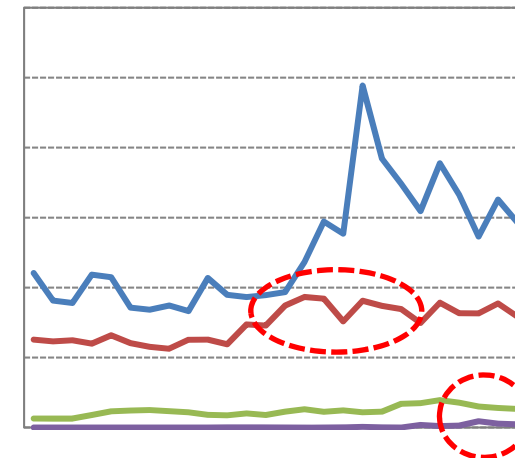
- 一定期間のユーザ群の投稿について、キーワード抽出を適用する

貢献

- 特定のユーザ群をターゲットとする場合のコンテンツ作成やクリエイティブ作成の手掛かりとなる

応用例：ソーシャルメディアからのインサイトの発見

2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31	2013/03/01 ~ 2013/03/31
順位	単語	順位変動	件数増加率	順位	単語	順位変動	件数増加率
1	神様	New	New	16	食パン	-12	-45%
2	絶品つけ麺	New	New	17	今日	4	45%
3	まとも	New	New	18	食品	-4	-20%
4	NAVER	New	New	19	取り寄せる部門	-2	-13%
5	つけ麺	New	New	20	アイス	-19	-66%
	CM	50	36	21	金	12	50%



合計 / デブ活
 合計 / ディズニーメイク
 合計 / パンダメイク
 合計 / 双子コーデ

ターゲットとする属性の投稿に特徴的なキーワードを抽出することで、その属性を持つ人物が興味を持っているもの・ことを知ることができる

応用例：ソーシャルメディアからのインサイトの発見

【ディズニーメイク】

東京ディズニーランドに行くときにするメイク
通常より色味も強く華やかなメイクが多く
ライナーのエッジも強い。



【双子コーデ】(双子コーディネート)

仲の良い友人と、同じ髪型・メイク・服装で合わせること。
イベントなど特別な日にすることが多い。



● 応用例：ソーシャルメディアからのインサイトの発見

EXILEファン高校生のトレンド(2018年2月)

テーマ	特徴語
-----	-----

EXILE関連	exile, 岩田剛典, 今市隆二, アカシア
---------	-------------------------

関連話題	斎藤 工, ローソン
------	------------

トレーダーのトレンド(2018年2月)

テーマ	特徴語
-----	-----

投資関連	ドル, 為替, 円, レート, 通貨, F X
------	-------------------------

仮想通貨	仮想通貨, ビットコイン
------	--------------

● 応用例：ソーシャルメディアからのインサイトの発見

研究者のトレンド(2018年2月)

テーマ	特徴語
-----	-----

研究関連	研究, 数学, ポスト, 大学, 論文
------	---------------------

関連話題	労働, 裁量, エンジニア
------	---------------

アイドルファンのトレンド(2018年2月)

テーマ	特徴語
-----	-----

アイドル関連	ライブ, 配信, 撮影, アイドル, 予約, アカメ
--------	----------------------------

関連話題	チェキ, ミスオブサークル, オートメッセ
------	-----------------------

● 応用例：ソーシャルメディアからのインサイトの発見

■ 単語のランキング手法

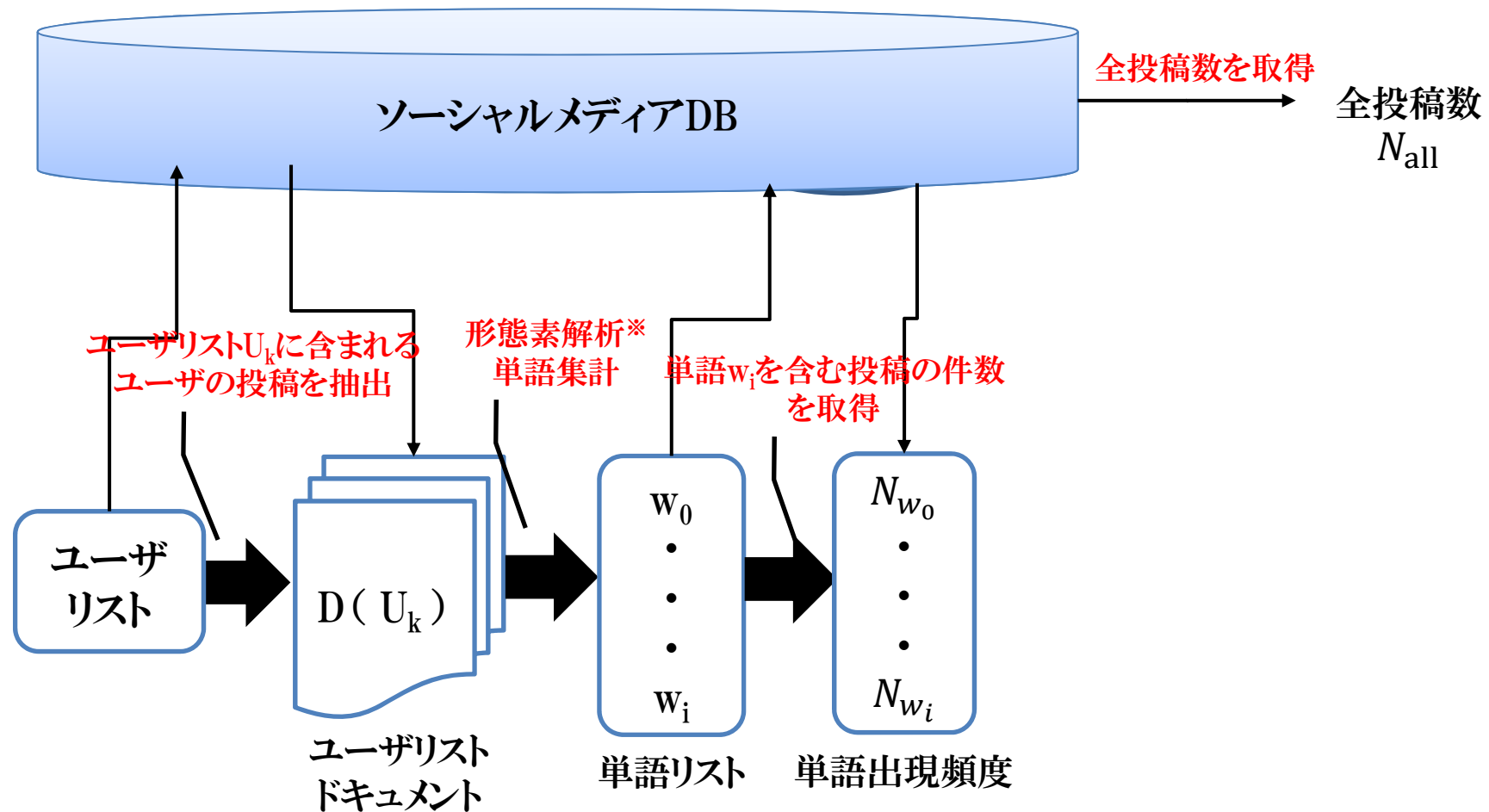
- 入力となるユーザリストの1ヶ月分の投稿を取得
- 全ての投稿を単語に分解(分かち書き)し、頻度を集計
- 単語ごとに χ^2 乗値を算出する
- χ^2 乗値が大きい順に単語をソートする

■ χ^2 乗値

- カイ2乗検定で用いられる値
- 期待値からのズレを表す値

$$- \chi^2 = \sum \frac{(\text{期待値} - \text{観測値})^2}{\text{期待値}}$$

応用例：ソーシャルメディアからのインサイトの発見



応用例：ソーシャルメディアからのインサイトの発見

目的

- 特定のユーザ群でのインサイトを発見したい

方法論

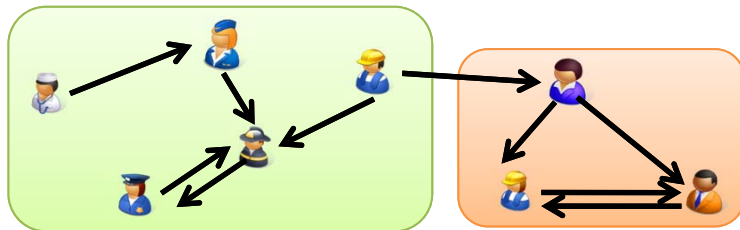
- 一定期間のユーザ群の投稿について、キーワード抽出を適用する

貢献

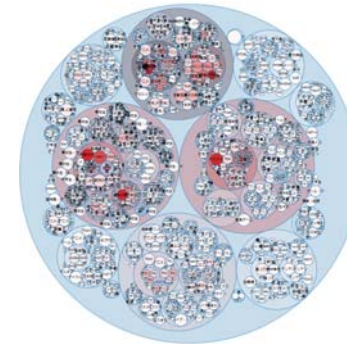
- 特定のユーザ群をターゲットとする場合のコンテンツ作成やクリエイティブ作成の手掛かりとなる

ソーシャルグラフマイニングの活用

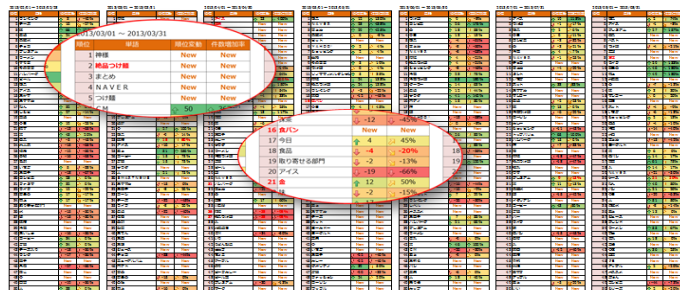
ユーザ属性の推定



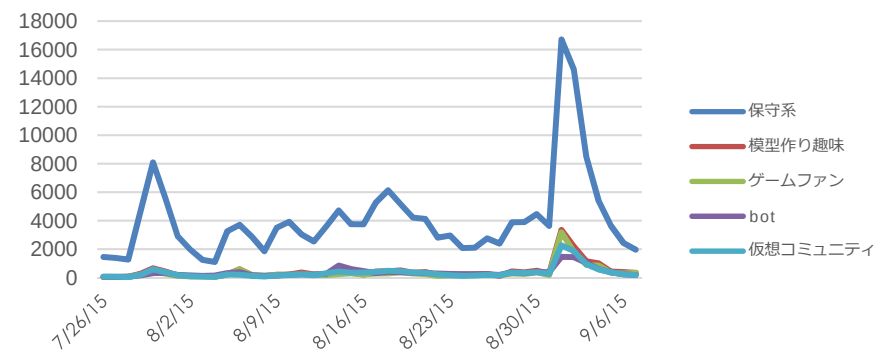
オーディエンス分析 ターゲティング



インサイト発見

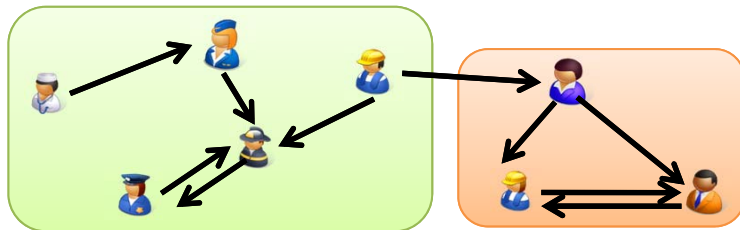


情報拡散分析

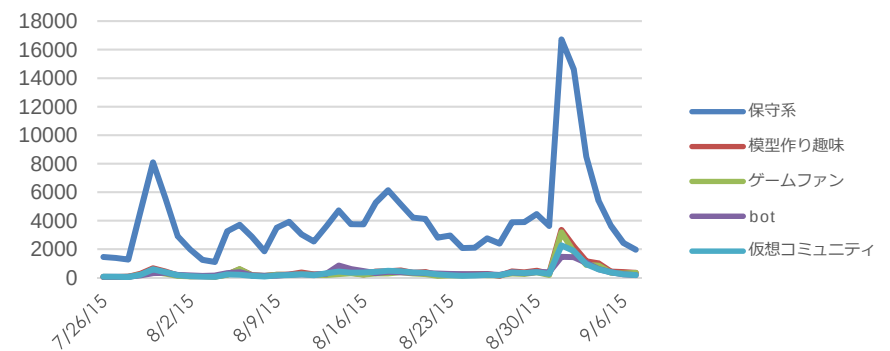


ソーシャルグラフマイニングの活用

ユーザ属性の推定



情報拡散分析



応用例：ユーザ属性を用いた情報伝播分析

目的

- 特定の情報伝播について、どのような属性のユーザが関わっていたかを明らかにした

方法論

- 情報伝播に関わったユーザを時間帯ごとに抽出
- 各時間帯ごとのユーザ群について、各属性ごとの分布を集計する

貢献

- ある話題が多様なユーザに広く話された話題なのか、それとも一部のユーザに偏った話された話題なのかを明らかにできる

応用例：ユーザ属性を用いた情報伝播分析

オリンピックエンブレム 炎上事例

■ 炎上の流れ

- 2020年東京オリンピックエンブレムについて、盗作疑惑が持ち上がる
- 製作者のデザイナー佐野研二郎及びそのデザイン事務所に関連した作品全般に盗作疑惑が持ち上がる
- 2ちゃんねるを中心に、検証が行われ、盗作元と思われる画像が複数発見される
- デザイナー仲間が擁護するも、さらに炎上する
- マスメディアに大きく取り上げられる
- オリンピック準備委員会、エンブレム使用中止を決定

● 応用例：ユーザ属性を用いた情報伝播分析

データセット

■ 検索キーワード

– エンブレム 佐野研二郎

■ 検索期間

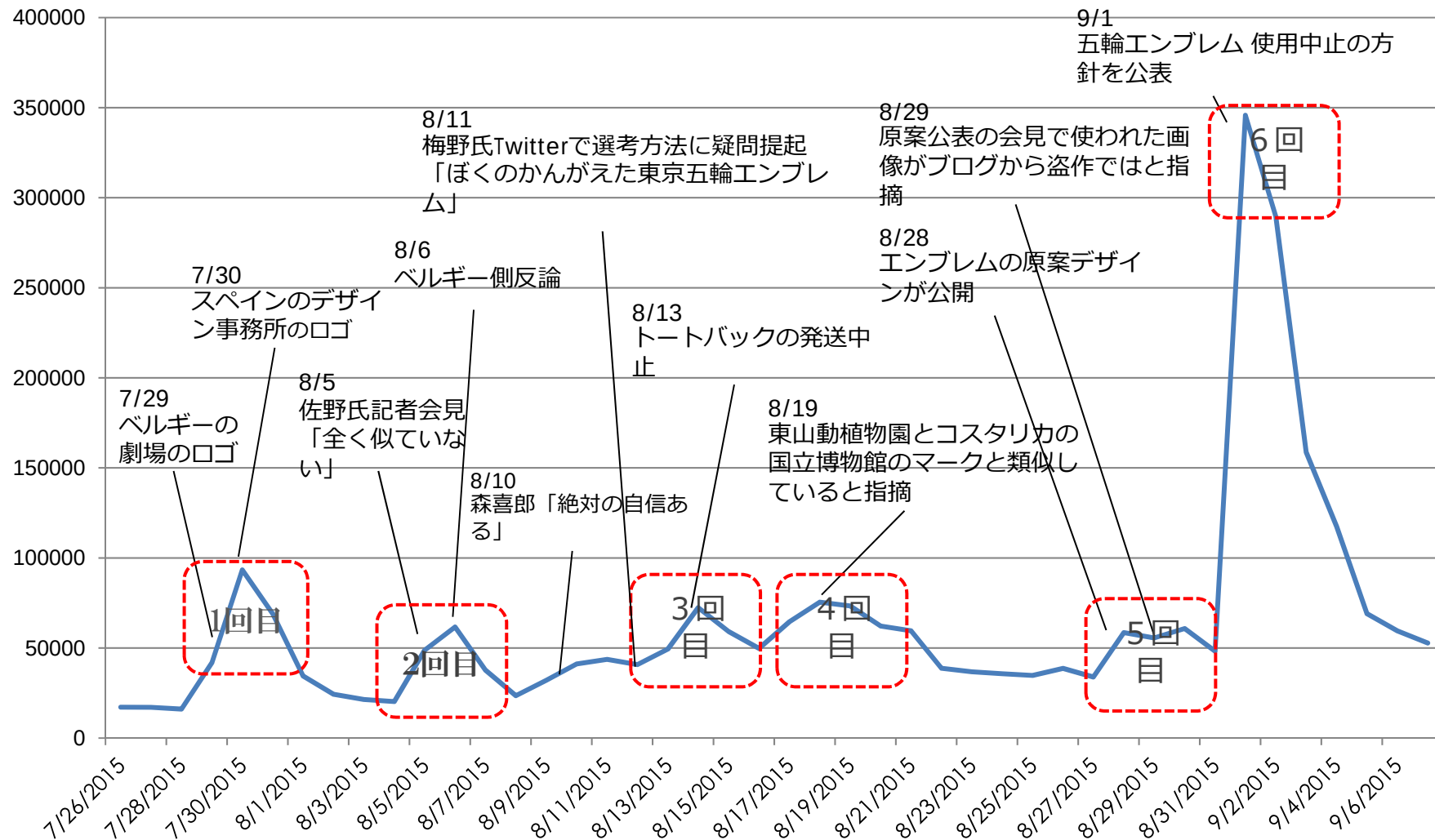
– 2015年7月26日～2015年9月7日

- 投稿件数が最大を記録してから、1週間後まで収集

応用例：ユーザ属性を用いた情報伝播分析

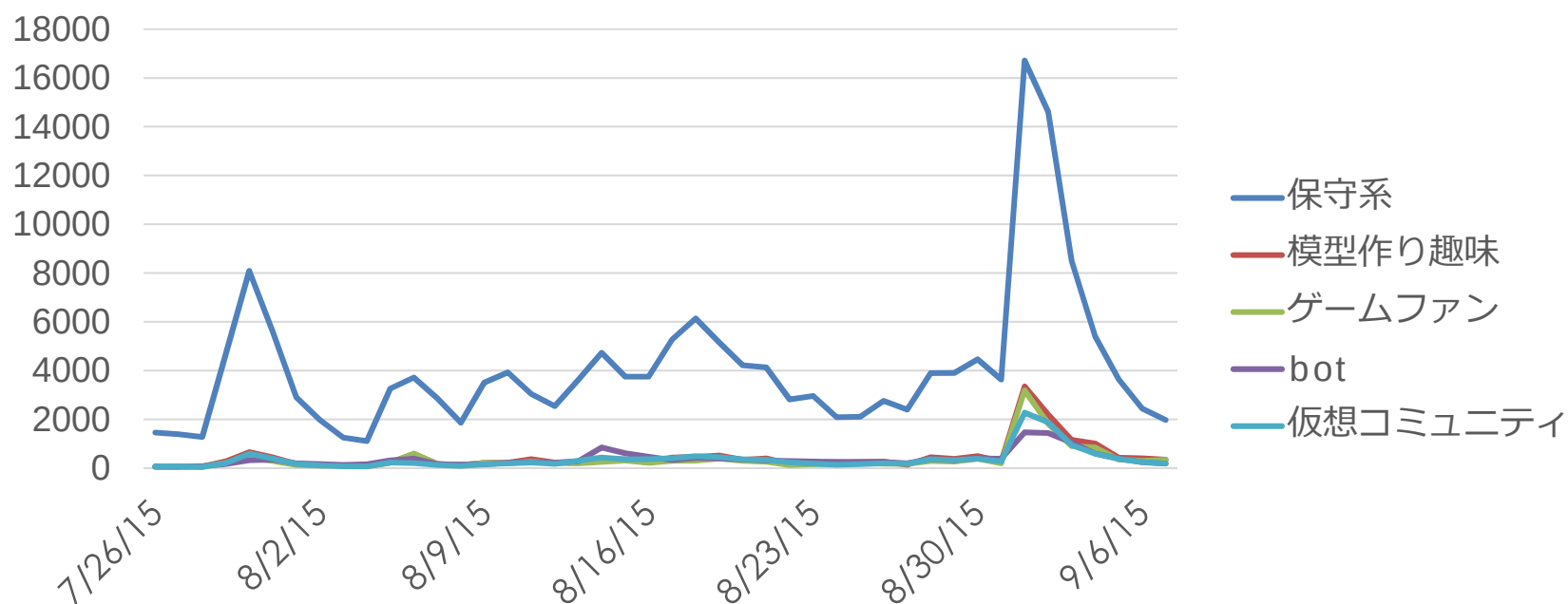
投稿数の時系列推移

バースト



応用例: ユーザ属性を用いた情報伝播分析

コミュニティ別投稿数推移



保守系コミュニティの投稿が支配的である。
また保守系コミュニティのみが大きく変化している時点もある。

応用例：ユーザ属性を用いた情報伝播分析

目的

- 特定の情報伝播について、どのような属性のユーザが関わっていたかを明らかにした

方法論

- 情報伝播に関わったユーザを時間帯ごとに抽出
- 各時間帯ごとのユーザ群について、各属性ごとの分布を集計する

貢献

- ある話題が多様なユーザに広く話された話題なのか、それとも一部のユーザに偏った話された話題なのかを明らかにできる

3. 今後の展望



ユーザ興味推定のためのソフトクラスタリング

動機

- ソーシャルメディアユーザ特有の**インタレスト・属性**を抽出し、それをユーザにラベルとして付与したい
- 一人のユーザが**複数のラベルを持つ**ことを許容したい

アプローチ

- ソーシャルメディア上でのインタラクションに基づくコミュニティ抽出
- 発言やプロフィール文を用いたコミュニティへの特徴付与

展望

- Network Embeddingの活用
- デファクトスタンダードとなり得るソフトクラスタリング手法の開発

ユーザ興味推定のためのソフトクラスタリング

技術的な課題

- 結果の良さと速度の両方を兼ね備えたソフトクラスタリング手法
- 動的かつ高速に変化するグラフについて、時系列変化も考慮できるクラスタリング手法
- (グラフマイニングではないですが)クラスタリング結果をより適切に解釈するための周辺手法の発展(自然言語処理等)

おわりに

- ソーシャルメディア上での様々な事象は、その複雑さ・大規模さから、人間には認識困難である
- 高速グラフマイニングの技術を用いることで、「人間が認識困難である情報」を可視化・俯瞰することができる
- 情報の非対称性は利益を生む(はずである)