

Query Keyword Extraction from Fuman Data for Collecting Wisdom Tips

楊 靚¹⁾ 北山 大輔²⁾ 角谷 和俊¹⁾

¹⁾関西学院大学 ²⁾工学院大学

背景・動機

ユーザは同じサービスを受けても...

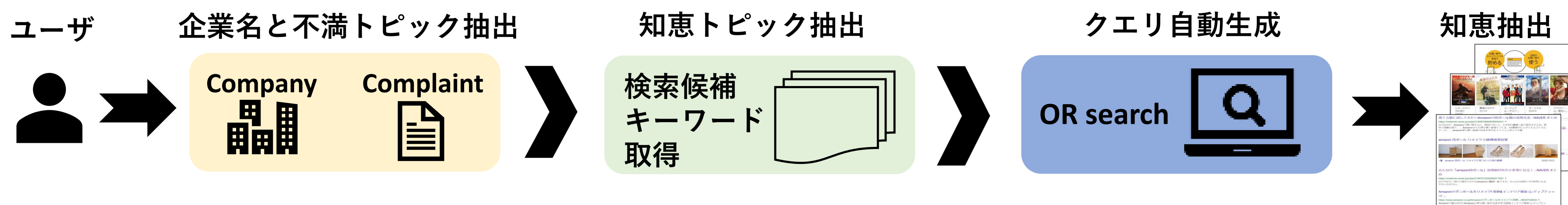
- ①人による考え方や価値観は違うので、不満の内容は様々
- ②ユーザの多様なニーズにリアルタイムに応えることは難しい
- ③サービスに対応する不満の直接解消方法はほとんどない
ユーザの不満を減少させる必要がある

目的

- ①ユーザの不満に対応する知恵を集め、**間接解消方法**の抽出
- ②間接解消方法を提供するためのWeb上の**クエリ自動生成**
- ③ユーザの流動的な不満に応じた**知恵提供**

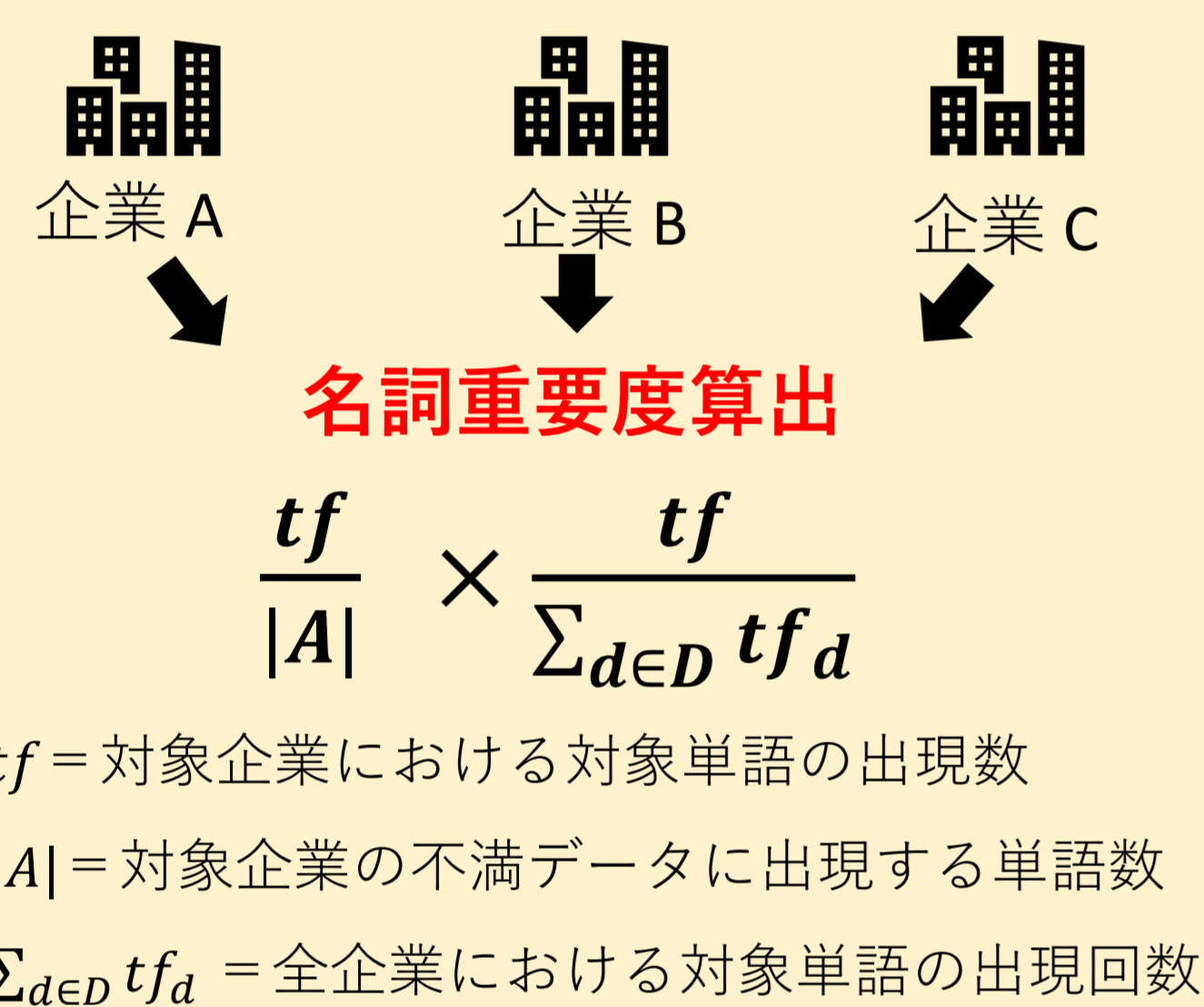


アプローチ



抽出手順

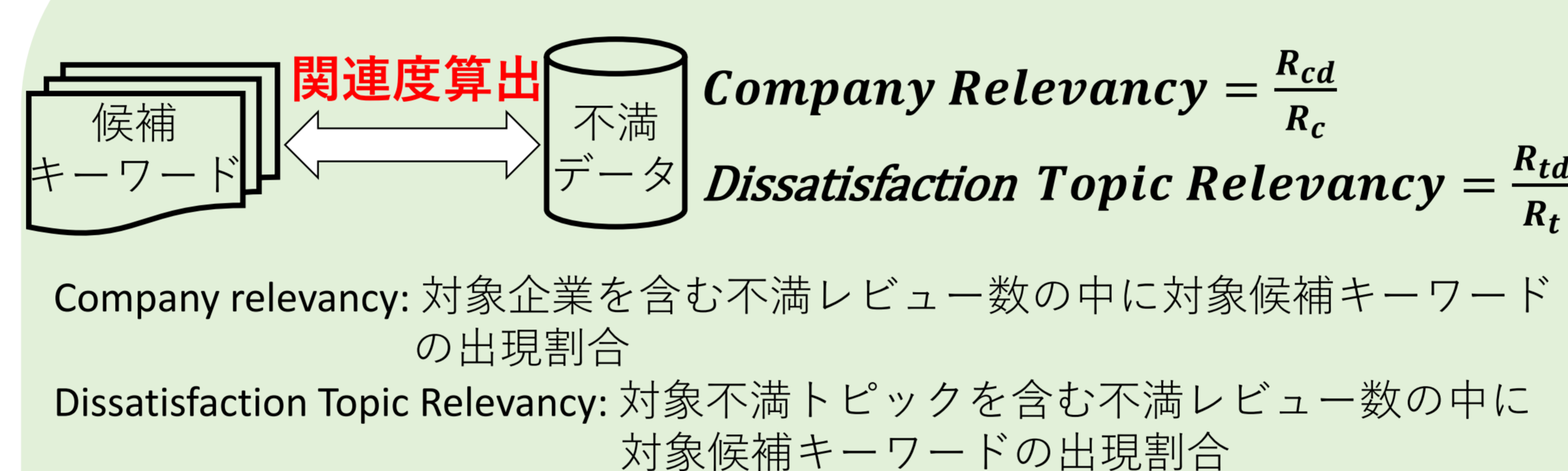
企業名と不満トピック抽出



	企業 A 重要度	企業 B 重要度	企業 C 重要度
名詞X	0.0045	0.0000	0.0000
名詞Y	0.0000	0.0043	0.0084
名詞Z	0.0000	0.0034	0.0000

抽出

知恵トピック抽出



$$Score = \frac{(CR+DR)}{2} W$$

品詞及び感情判定	重み(W)
ネガティブ	0.8
動詞	0.7
形容詞	0.7
固有名詞(地名)	0.7
固有名詞(組織)	0.3
一般名詞	0.3
サ変名詞	0.1

候補キーワード	Score
候補キーワード1	0.0054
候補キーワード2	0.0032
候補キーワード3	0.0021
候補キーワード4	0.0017
候補キーワード5	0.0015
候補キーワード6	0.0009
候補キーワード7	0.0007
候補キーワード8	0.0007
候補キーワード9	0.0001
候補キーワード10	0.0000

抽出

クエリ自動生成

OR search

1. ユーザの多様な不満に応えるため、一つの企業名と不満トピックに対する複数の知恵トピックが生成される
2. 複数の知恵トピックの中から**OR**で一つの知恵トピックを選ぶ、検索を行う
例:

企業A **^** **ポイント** **^** **貯め方** **^** **アップ**

知恵提供



実験

「実験1. 企業名と不満トピック抽出」

- ・同じsub_category下にある6企業を比較することにより、分析対象として3企業それぞれの不満トピックの抽出することができた

企業	抽出個数	閾値	不満トピック (例)
A	186	0.00080	プライム, 配達, 映画, ダンボール, 送料, 注文, 予約, 本, ポイント, 返品
B	144	0.00076	スタンプ, コード, ブロック, コイン, バックアップ, 通話, カメラ, 設定, 投稿, 広告
C	86	0.00080	ニュース, 質問, 出品, 回答, 落札, 記事, タイトル, ID, 雨雲, 取引

データセット

- ・不満調査データセットVER2
- ・sub_category: 「It・Webサービス」
- ・比較対象: 6企業 (A, B, C, D, E, F)
- ・分析対象: 3企業 (A, B, C)
- ・データ量: 各企業1000件、合計6000件不満レビュー

「実験2. 知恵トピック抽出」

- ・上記30個の不満トピックの知恵トピック抽出を行った
- ・企業Aは閾値**0.0043**、企業Bは閾値**0.0020**、企業Cは閾値**0.0033**以下の単語を抽出した

企業	不満トピック (例)	知恵トピック (例)
A	ポイント	チャージ, プレゼント, 貯め方, 使い方, アップ, クレジットカード
B	設定	セキュリティ, android, グループ友達, プライバシー, 初期化, おすすめ
C	プレミアム	特典, メリット, 解約, ソフトバンク, 雑誌

A ポイント	Score
購入	0.0515
サイト	0.0101
期限	0.0090
交換	0.0059
クレジットカード	0.0038
チャージ	0.0022
使い方	0.0011
アップ	0.0009
プレゼント	0.0009
貯め方	0.0000

B 設定	Score
友達	0.0135
どこ	0.0045
通知	0.0038
電話番号	0.0023
セキュリティ	0.0017
android	0.0010
グループ友達	0.0006
プライバシー	0.0005
初期化	0.0005
おすすめ	0.0001

C プレミアム	Score
ヤフオク	0.0115
無料	0.0089
料金	0.0056
コンビニ	0.0038
特典	0.0032
メリット	0.0023
解約	0.0013
ソフトバンク	0.0008
雑誌	0.0002

考察

企業	正解数	抽出件数	再現率
A	45	38	0.86
B	52	36	0.78
C	35	28	0.81

- ・提案手法で不満トピックに対して知恵トピック抽出が可能
- ・スコアで**知恵トピックらしさ**を**ランキング**することが可能
- ・スコアで知恵トピックをランキングする際に**ノイズの除去**が必要

今後の課題

- 不満調査データセットから知恵トピック抽出の検討
- IT以外のドメインのサービスに不満解消方法の抽出の検討 (外食・店舗、公共・環境など)
- ユーザが書き込んだ不満を、自動的解消方法を検索するためのクエリに変換し、知恵提供を行うウェブサイト作成への検討