

レビューの類似性を利用した飲食店推薦

長崎大学 工学部工学科情報工学コース

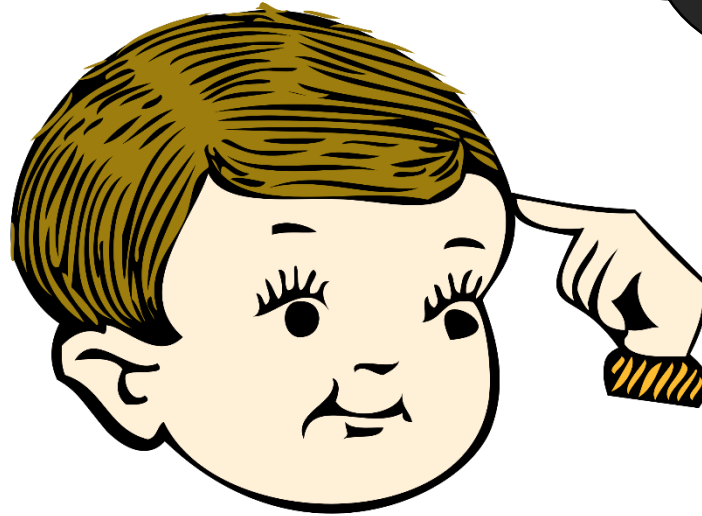
正田研究室

35315055 JEON GOEUN

研究の背景

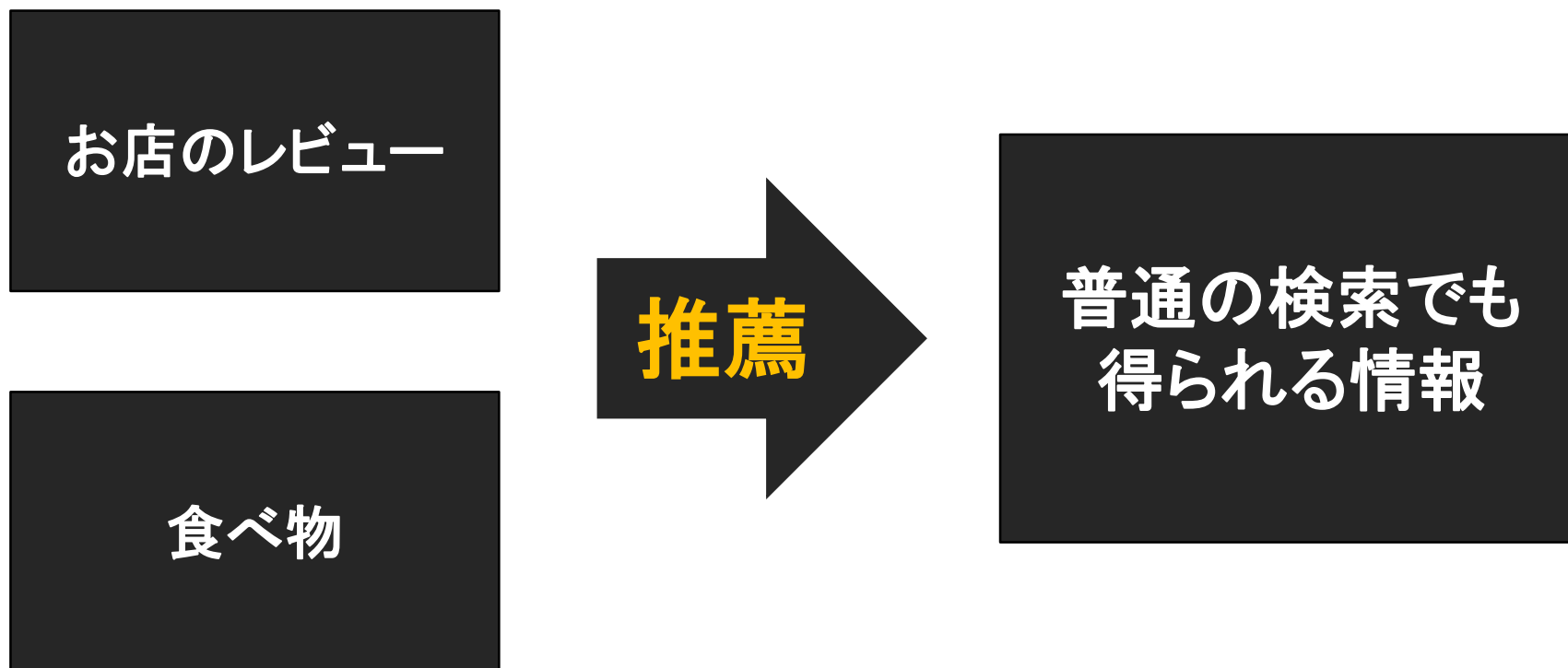
昨日は、そばを
食べたから...

今日は
なにを食べる？



研究目的

ユーザーに新しい**情報**を提供すること



研究目的

ユーザーに新しい**情報**を提供すること



使用するデータ

食べログ

日本のグルメ口コミサービス
「食べログ」の公式アプリ
全国80万件以上の飲食店に寄せら
れた2,200万件以上の口コミが無料
で見られる

お店のレビューを用いて
類似度を求める

口コミ ?

投稿する



☆ (866) 認証済

従業員が多すぎて、価格が高くなってる？

人生初の一蘭をなんと博多で(笑)。私はシステムとか全くわからないのですがなんとかかんとか金券を買って席へ。購入したのはラーメン890円キクラゲ140円。主人はラーメン890... [続きを読む»](#)

 62 ? 訪問：2018/05 4.5 1回

 (753) 認証済

博多駅前の地下街にある人気のお店です！一蘭好きな私は、つい行ってしまいま...

博多駅近辺で飲んだ後、〆にラーメンを食べようと思ってきました。お店は、博多駅前の地下2階にあるサンブラザ地下街にあるお店で、前日の夕方にもお店の前まで来ましたが、行列を... [続きを読む»](#)

 155 ? 訪問：2018/04 4.5 1回

 (3042) 認証済

☆★一蘭ラーメン(^_^)3辛濃い味★☆

すっかりチェーン店 何処でも大阪も何店もあり 有難くもない一蘭ですが 味の確認に 苦笑 ココは 食べログ前に何度か来てます〜♪ なんと790円から 890円... [続きを読む»](#)

 374 ? 訪問：2017/08 4.5 1回

🔍 口コミをもっと見る (193件)

使用するデータ

食べ物用語辞書

日本の料理、食材、調味料、調理法等、日本の食べ物に関わる言葉と食品用語を分かりやすく整理したサイト

食べ物用語辞書を用いて
非類似度を求める

日本の食べ物用語辞典 Words of Traditional and Modern Japanese Food

ホーム 五十音順 ▾ 都道府県 ▾ 種別 ▾ 日本料理用語 季節 風習 習慣 地域ならではの食べ方 酒肴 日本の漬物 🔍
日本の食べ物用語辞典とは お問い合わせ 用語一覧 世界の食べ物用語辞典

Pick Up

油そうめん



油そうめんとは・・・油...

[Read More »](#)

舟納豆



舟納豆とは・・・舟納豆...

[Read More »](#)

四方竹



煮貝



[キーワード検索](#)

Search here...

[初めての方へ](#) [お問い合わせ](#) [世界地図](#)

提案手法(1) 形態素解析

形態素解析

代表的な日本語の形態素解析ライブラリー

約2000個の
お店のレビュー

2724種類の
食べ物の説明文

MeCab



レビューの
単語リスト

食べ物の
単語リスト

提案手法(2) cos類似度の計算

- TF-IDF (Term Frequency – Inverse Document Frequency)

情報検索論で良く見られる加重値を求めるために用いられるアルゴリズム

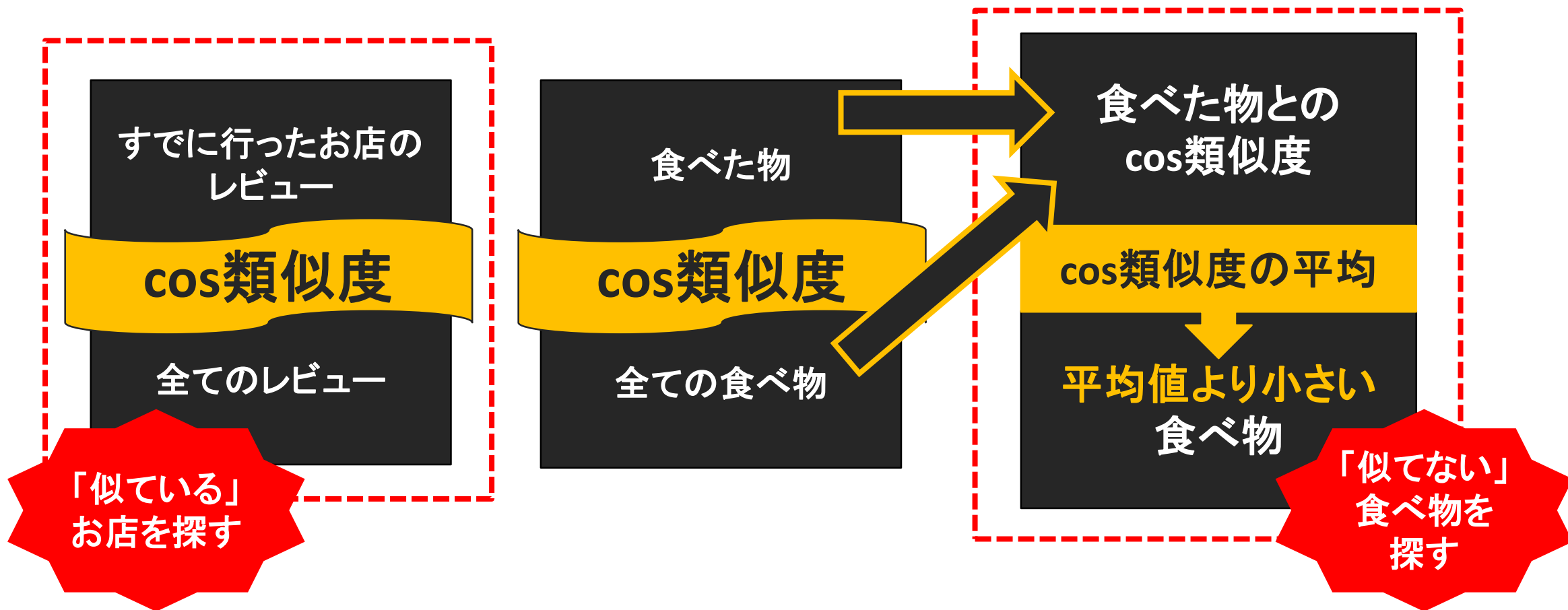
- コサイン類似度

ベクトル空間モデルにおいて、文書同士を比較する際に用いられる類似度計算手法

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

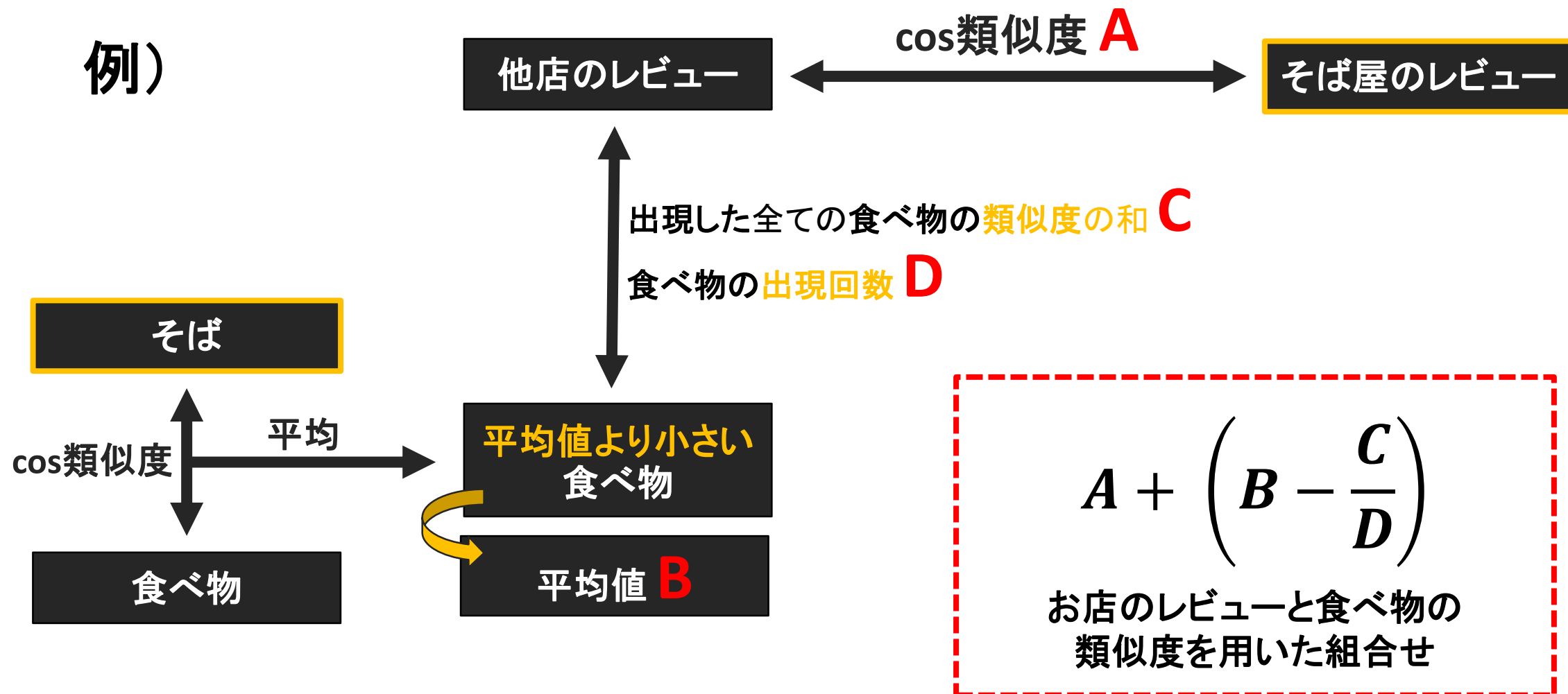
本実験

すでに行ったお店と似ているが、食べたものとは似てない食べ物を売っているお店を推薦する



本実験の実験方法

例)

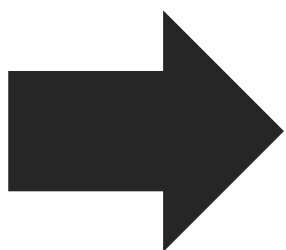


$$A + \left(B - \frac{C}{D} \right)$$

A : すでに行ったお店と他店のレビューのcos類似度
B : 食べ物のcos類似度の平均値
C : 他店のレビュー出現した全ての食べ物のcos類似度
D : 他店のレビュー出現した全ての食べ物の出現回数

$A \geq \left(B - \frac{C}{D} \right)$ のため、

お店のレビューの類似度の比重が高い = 推薦の結果がレビューの方に偏る



$$A + \left(B - \frac{C}{D} \right) \times N$$

Aと $\left(B - \frac{C}{D} \right)$ の値が
同等な値を持つ

$$N = 3$$

本実験の実行結果

すでに行ったお店

「芳水」「栄寿司小吉」「東角」

「そば」「すし」「ちゃんぽん」

食べたもの



N = 0

お店の類似度だけ用いたとき

N = 3

店と食べ物の類似度を用いたとき

評価方法

● DCG (Discounted Cumulative Gain)

正解データのランキングをどれだけ再現できるのかの評価指標

$$DCG_q = rel_i + \sum_{i=2}^p \frac{rel_i}{\log_2 i}$$

✓ 判断基準

意外性のある推薦なのか

お店が食べログの
検索結果で「標準」に
入っているか

✓ 採点基準 (rel の値)

最小 ← → 最大
「0」 「1」 「2」

長崎市のそば（蕎麦）のお店 47 件

[検索条件を変更](#)

長崎市 × そば ×

さらに絞り込む ランチ 個室 完全禁煙 一人で入りやすい

標準【会員店舗優先】

ランキング

口コミ数順

ニューオープン順

評価方法

N = 0

お店の類似度だけ用いたとき

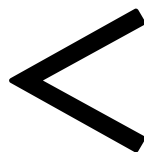
N = 3

店と食べ物の類似度を用いたとき

∴ DCGの値が大きいほど良い推薦

本実験の評価

食べたものが「**そば**」の場合



お店の類似度だけ用いたとき

	お店	番号	①	②	平均
1	亜紗 本店	1.1778	2	0	
2	蕎家	1.1257	0	2	
3	紅灯記	1.881	1	0	
4	天ひろ	1.1364	2	0	
5	蕎家	1.125	0	2	
DCG			3.63	2.86	3.25

店と食べ物の類似度を用いたとき

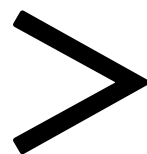
	お店	番号	①	②	平均
1	魚菜や 朝次郎 アミュプラザ長崎店	1.303	2	0	
2	いろは 本店	1.492	2	0	
3	海鮮市場 長崎港 出島ワーフ店	1.46	2	0	
4	ますだ	1.1859	2	0	
5	蕎家	1.1245	0	2	
DCG			6.26	0.86	3.56

①意外性のある推薦なのか

②食ベログでキーワード検索結果の「標準」に推薦されたお店が入っているか

本実験の評価

食べたものが「すし」の場合



お店の類似度だけ用いたとき

	お店	番号	①	②	平均
1	長崎の味処 鮨・割烹さくらい	1.1819	0	2	
2	千成寿し	1.1107	0	2	
3	寿司割烹 松ふじ	1.1755	0	2	
4	コバチ	1.506	2	0	
5	寿司割烹 松ふじ	1.1758	0	2	
DCG			1.00	6.12	3.56

店と食べ物の類似度を用いたとき

	お店	番号	①	②	平均
1	長崎の味処 鮨・割烹さくらい	1.1819	0	2	
2	いわしや	1.497	1	0	
3	こじま 本店	1.568	1	0	
4	寿司割烹 松ふじ	1.1755	0	2	
5	酒菜処 のさ庵	1.404	2	0	
DCG			2.49	3.00	2.75

①意外性のある推薦なのか

②食ベログでキーワード検索結果の「標準」に推薦されたお店が入っているか

本実験の評価

食べたものが「**ちゃんぽん**」の場合

=

お店の類似度だけ用いたとき

	お店	番号	①	②	平均
1	康楽	1.917	0	2	
2	梁川飯店	1.1718	0	2	
3	中華・将龍	1.185	0	2	
4	うまかもん亭	1.1634	2	0	
5	王鶴	1.1305	0	2	
DCG			1.00	6.12	3.56

店と食べ物の類似度を用いたとき

	お店	番号	①	②	平均
1	康楽	1.917	0	2	
2	中華・将龍	1.185	0	2	
3	江山楼 浦上店	1.238	0	2	
4	天天有	1.485	0	2	
5	群来軒	1.213	0	2	
DCG			0	7.12	3.56

①意外性のある推薦なのか

②食ブログでキーワード検索結果の「標準」に推薦されたお店が入っているか

結論

- 3つの食べ物の中で1つだけ評価が悪くなっていた
→ 他の食べ物でも試して評価する
- お店のレビューの中に食べ物の名前があまり出なかった
→ お店のメニューを用いたら食べ物の類似度を高めることができると考えられる
- お店のレビューのデータ量が少なかった
→ 他のグルメサイトなども用いて、それぞれのお店の情報を増やすことでレビューの類似度の質を上げることができると考えられる