

レシピサイトからの 技術的豆知識抽出手法

甲南大学 知能情報学部 知能情報学科

灘本研究室 11471125 米田吉希

背景

何作ろうかな？

カルボナーラ作ろう！
でも作り方は・・・？



背景

楽天レシピ



クックパッド



cookpad

ユーザ投稿型レシピサイトを活用

背景



美味しくなる工夫や注意すべき点等が
記載されているレシピもある

知っている
失敗を防ぐ
ことができる

例えば

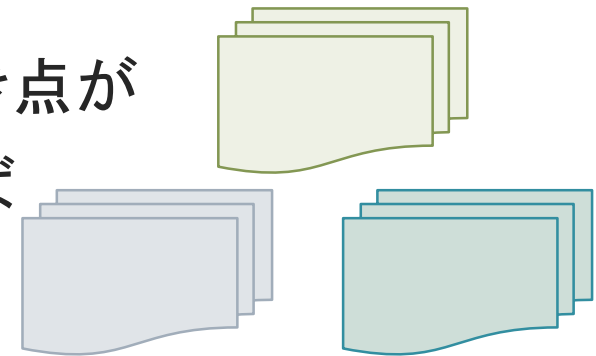
しかし

美味しくなるコツ

卵黄は固まりやすいので、火を止めてから
入れて手早く混ぜることがコツです。

背景

- ・レシピの量が膨大
- ・美味しくなる工夫や注意すべき点が記載されているレシピもあれば
そうでないレシピもある



欲しい情報が見つからない...

役に立つ情報を探すのが大変...



背景

おいしくなるコツ

作り方は色々ありますが、全卵を使う簡単なレシピを選びました♪

調理する上で役に立つ情報がない

知っていること失敗を防ぐことができる

しかし他のレシピには

おいしくなるコツ

卵黄は固まりやすいので、火を止めてから入れて手早く混ぜることがコツです。

豆知識

ユーザ自身が調理についての知識の拡大や料理技術の向上が期待できる情報

目的

レシピサイトから豆知識を自動で抽出する

本研究では全料理に共通する豆知識ではなく
各料理に関する豆知識を対象とする



豆知識

時間短縮

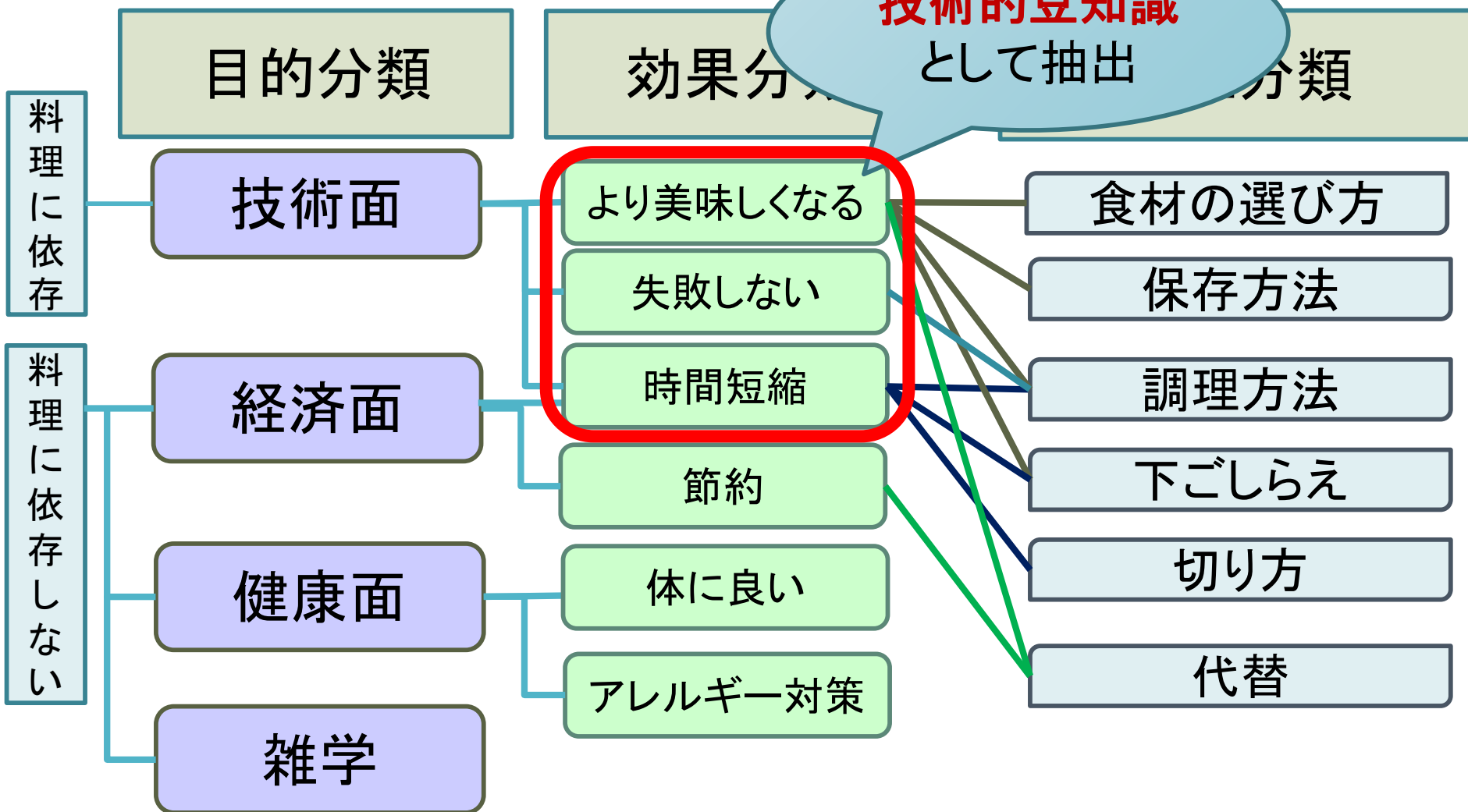
より美味しくなる

豆知識

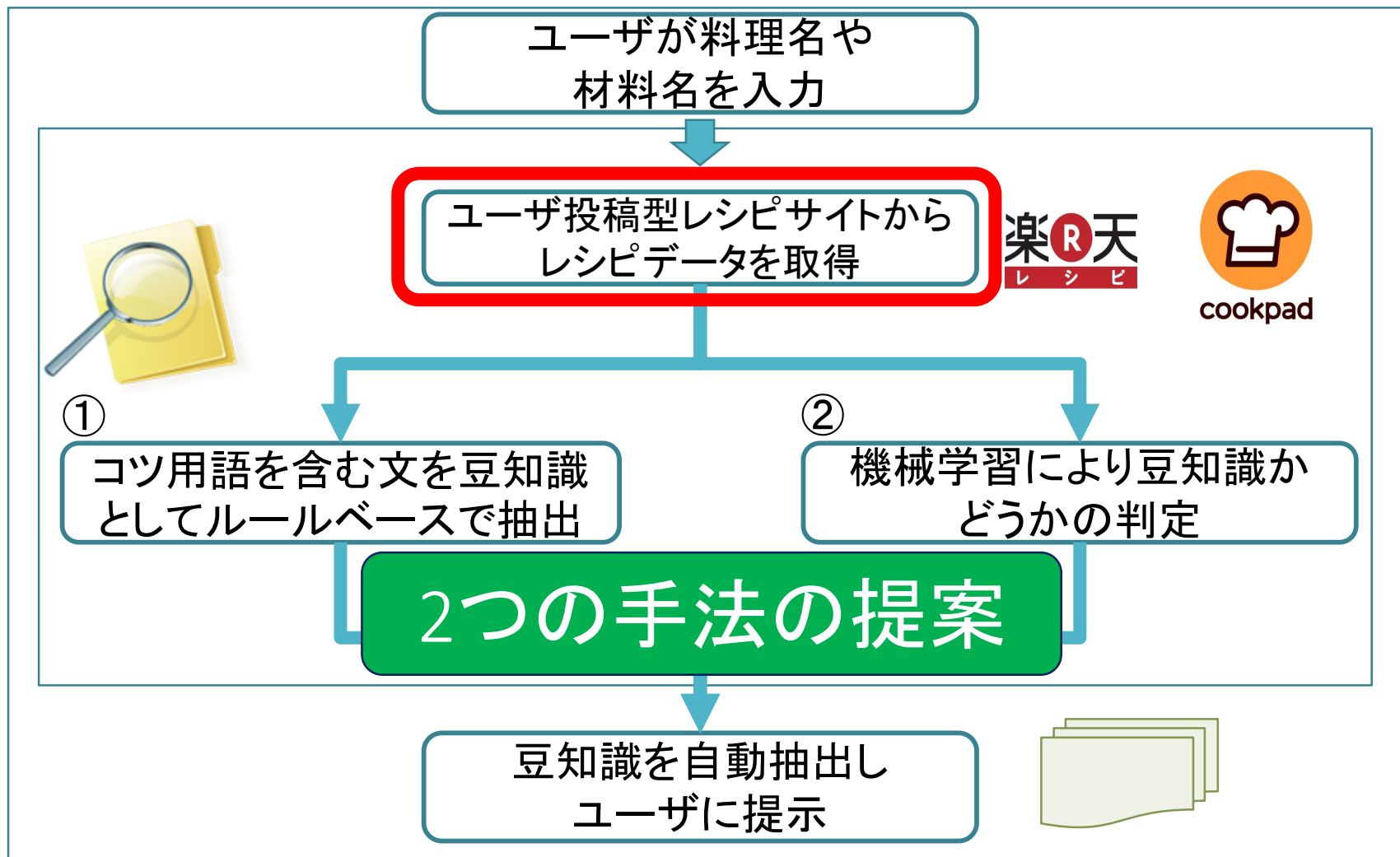
体に良い

失敗しない

豆知識の分類



全体の流れ



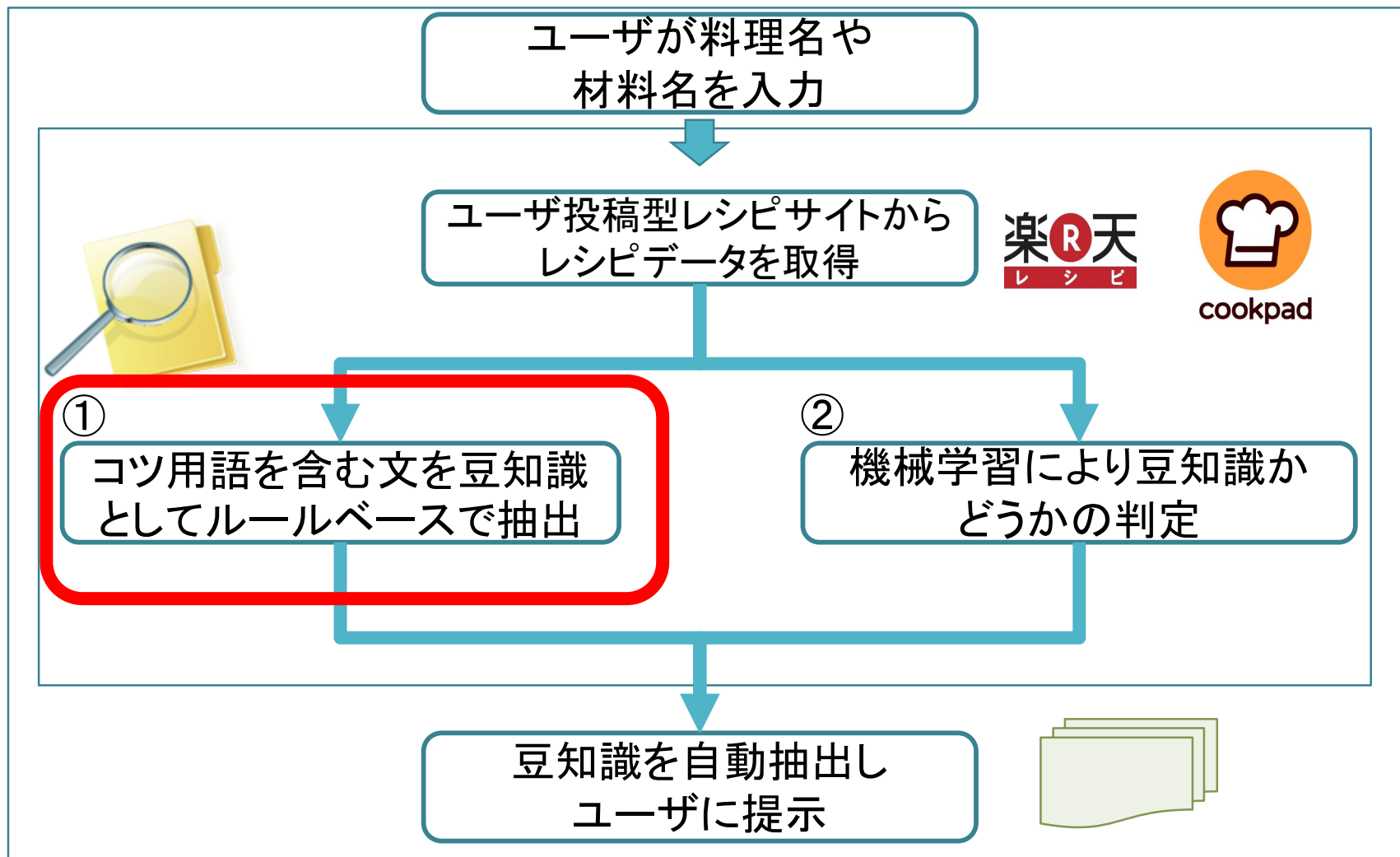
レシピデータの取得



cookpad

ユーザが入力した料理名や材料名の
レシピデータをクックパッドと楽天レシピ
から取得する

全体の流れ



①ルールベース

火をつけたまま、卵液を入れると
一気に固まってしまうので**注意**♪

ベーコンをカリカリに焼くと、
コクが出て美味しくなりますよ！

注意

特徴語

コク

1000件の
レシピを分析

54語の
コツ用語を決定

コツ用語

①ルールベース

コツ用語54語

下処理	下ごしらえ	代用	代替	代わり	無い	足りない	本格
綺麗	おいしい	こつ	おいしく	重要	美味	大事	コツ
グッド	オススメ	風味	おすすめ	いい	良い	コク	引き立
日持ち	スピーディ	失敗	気をつけ	楽	注意	簡単	安い
安く	カンタン	予防	防止	含む	含み	効果	効能
※	ポイント	保存	アップ	冷蔵	冷凍	防ぐ	防ぎ
効率	アレルギー	時短	時間短縮				

ルール

レシピからこのコツ用語54語のうち
 1つ以上含む文を豆知識として抽出

評価実験 (①ルールベース)

目的

有用性を測るために評価実験を行った

条件

- データ取得サイト:
クックパッド
楽天レシピ
- 料理名: カルボナーラ
カレーライス
親子丼
- データ数: 各料理3,500件

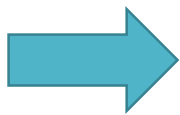
手順

- データ内にコツ用語54語のうち1つ以上含まれているものを豆知識として抽出
- 正解データは人手で作成し、適合率・再現率・F値を求めた

結果 (①ルールベース)

料理名	適合率	再現率	F値
カルボナーラ	0.292	0.671	0.407
カレーライス	0.203	0.638	0.308
親子丼	0.206	0.410	0.274

- 適合率・再現率・F値が3料理とも低い結果となった



コツ用語54語でのルールベースでは技術的豆知識が取りきれてないため。また、豆知識でない文を豆知識として抽出しているため。

考察(①ルールベース)

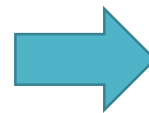
例(カルボナーラ)

良い例

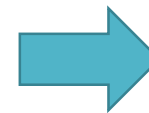
- 卵を加熱しすぎるとパサパサになってしまうので注意。

悪い例

- コツとかポイントは特になく、
簡単だと思います
- うどんからの飛び跳ねる熱湯で
やけどに注意してください。
- 参考簡単タルタルソースID:4320391
失敗しないカルボナーラID:4330142



コツ用語を
否定している

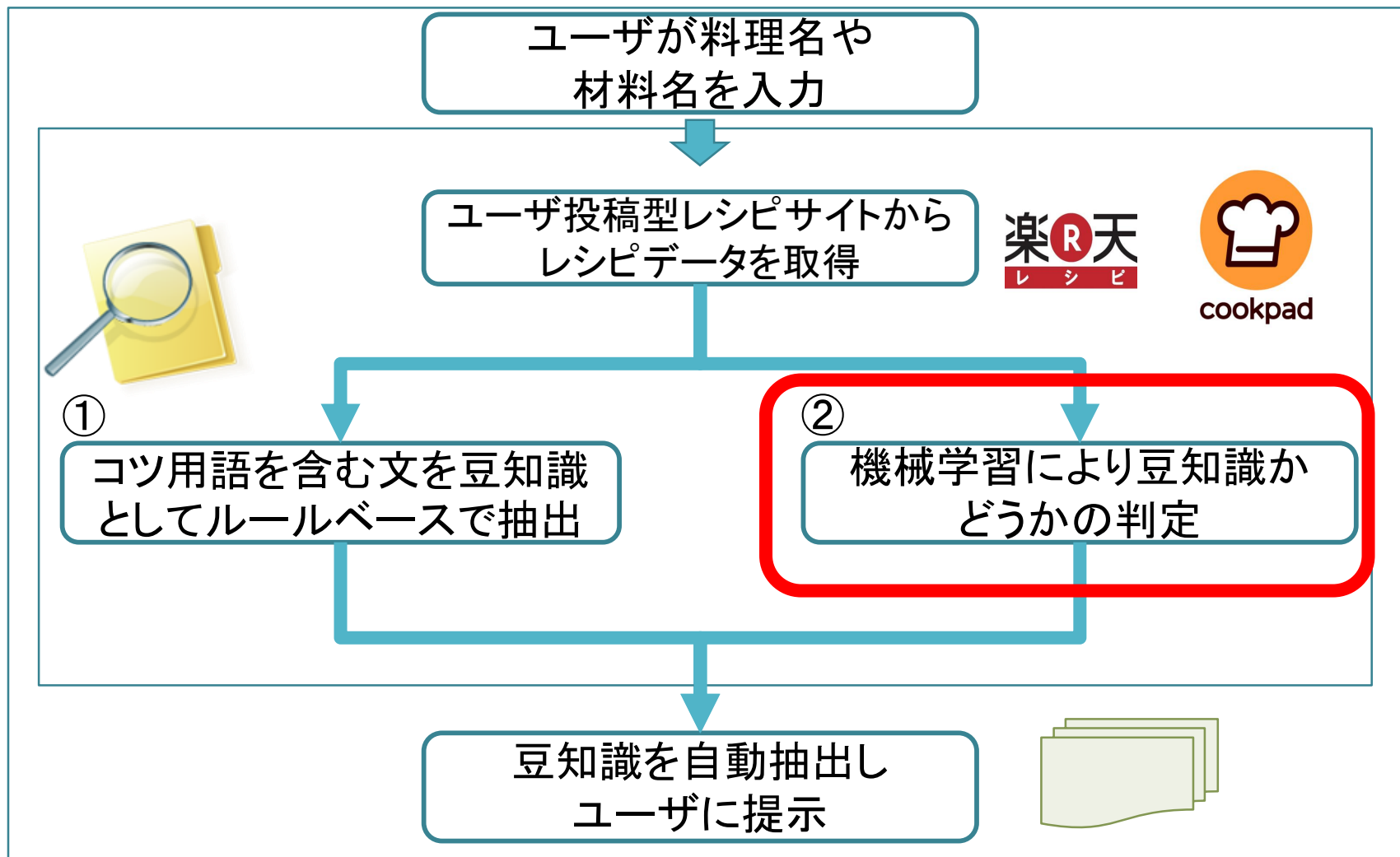


対象料理が違う



他のページを
参照している

全体の流れ



②機械学習

- 学習モデル： SVM
- カーネル： RBFカーネル
- 教師データ： 正例 豆知識文 各料理300件
負例 豆知識でない文 各料理300件
(料理に関係がある文)
- 素性： 教師データに含まれている名詞・動詞・形容詞が
評価データに含まれているか否か

評価実験 (②機械学習)

目的

有用性を測るために評価実験を行った

条件

- データ取得サイト:
クックパッド
楽天レシピ
- 料理名: カルボナーラ
カレーライス
親子丼
- データ数: 各料理3,500件

手順

- 性能評価: 10交差検定
未知のデータ
- 正解データは人手で作成し,
適合率・再現率・F値を求めた

結果 (②機械学習)

料理名	評価手法	適合率	再現率	F値	分類精度
カルボナーラ	10交差検定				0.998
	未知データの分類	0.5	1.00	0.667	0.500
カレーライス	10交差検定				0.995
	未知データの分類	0.486	1.00	0.654	0.486
親子丼	10交差検定				0.741
	未知データの分類	0.673	0.760	0.714	0.695

- 3料理ともに適合率が低く、再現率が高い結果となった



豆知識でない文を豆知識と判断されたため。

考察(②機械学習)

例(カルボナーラ)

良い例

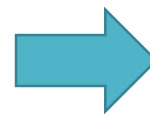
- 卵を加熱しすぎるとパサパサになってしまうので注意。

悪い例

- うどんをバターで炒めてから、パスタソースを加えました
- お好みで粗挽き黒こしょうを最後にふりかけて下さい♪
- こってりが好きならスライスチーズを増やして下さい。



出現単語に
あまり差がなかった



重要単語の
重みを付ける
必要がある

①の手法と②の手法の比較

ルールベース

料理名	適合率	再現率	F値
カルボナーラ	0.292	0.671	0.407
カレーライス	0.203	0.638	0.308
親子丼	0.206	0.410	0.274

機械学習

料理名	適合率	再現率	F値
カルボナーラ	0.50	1.00	0.667
カレーライス	0.486	1.00	0.654
親子丼	0.673	0.760	0.714

ルールベースでのF値は良くなかったが、
 機械学習でのF値はある程度良かった。



豆知識抽出
 手法の改善

まとめと今後の課題

●まとめ

ルールベースと機械学習による技術的豆知識抽出手法を提案し、その有用性を測った。
ルールベースによる抽出の有用性は低く、
機械学習による抽出はある程度良かった。

●今後の課題

実験料理数の拡大

別の手法による自動抽出

技術面以外の項目の豆知識の自動抽出