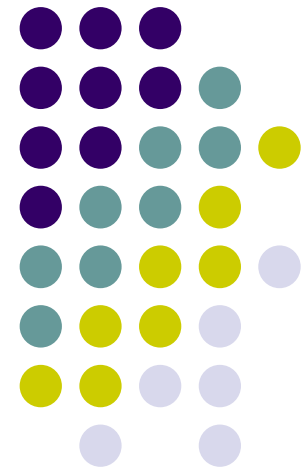


Webサービス検索のための 操作カテゴリ分類手法の提案

早稲田大学大学院 基幹理工学研究科
情報理工学専攻 修士1年
片渕 聡

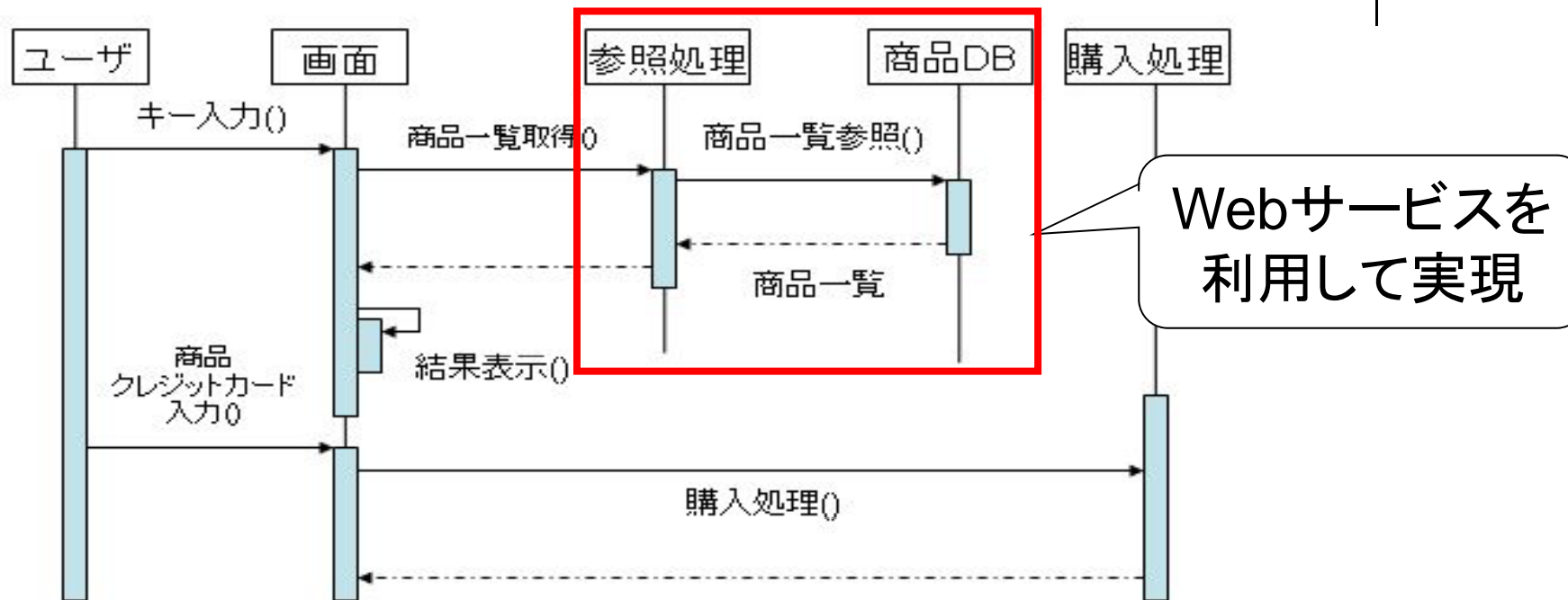




目次

- 背景
- 既存研究
- 提案手法
 - 操作内容による分類
 - 入出力のデータ構造による分類
- 評価
- 関連研究
- まとめ

背景



インターネット上で公開されているWebサービスを利用することで開発コストの低減が可能

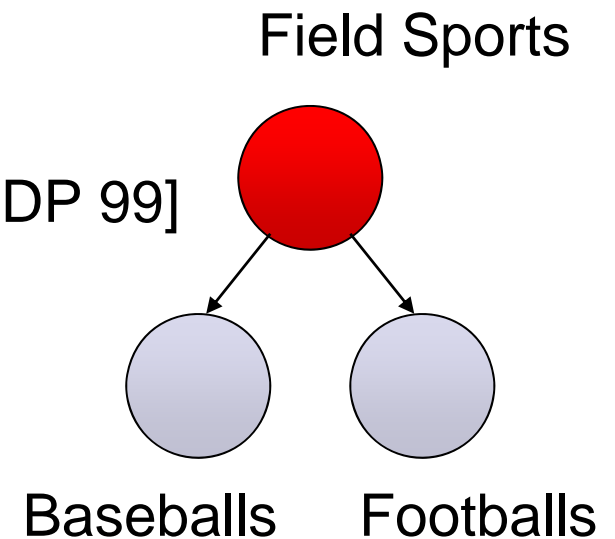
期待するWebサービスを効率よく検索したい



検索における既存研究

- カテゴリ分類を用いた検索
 - 意味のまとまった機能ごとに分類
→ 検索範囲の絞り込みに有効

- **ドメインカテゴリ分類**(UNSPSC)[UNDP 99]
 - 対象ドメインごとにサービスを分類



ドメインカテゴリの問題

1. 操作による効果的な絞り込みができない
2. 振舞いの修正を要する機能を排除できない

既存研究の問題点(1)

～操作内容による分類



- 問題1: 操作による効果的な絞込みができない

金融ドメイン

操作(何をするのか)が異なる機能
まで検索にひっかかってしまう

「株価の取得」
「通貨レート変換」
「銀行の検索」
「株の投資」

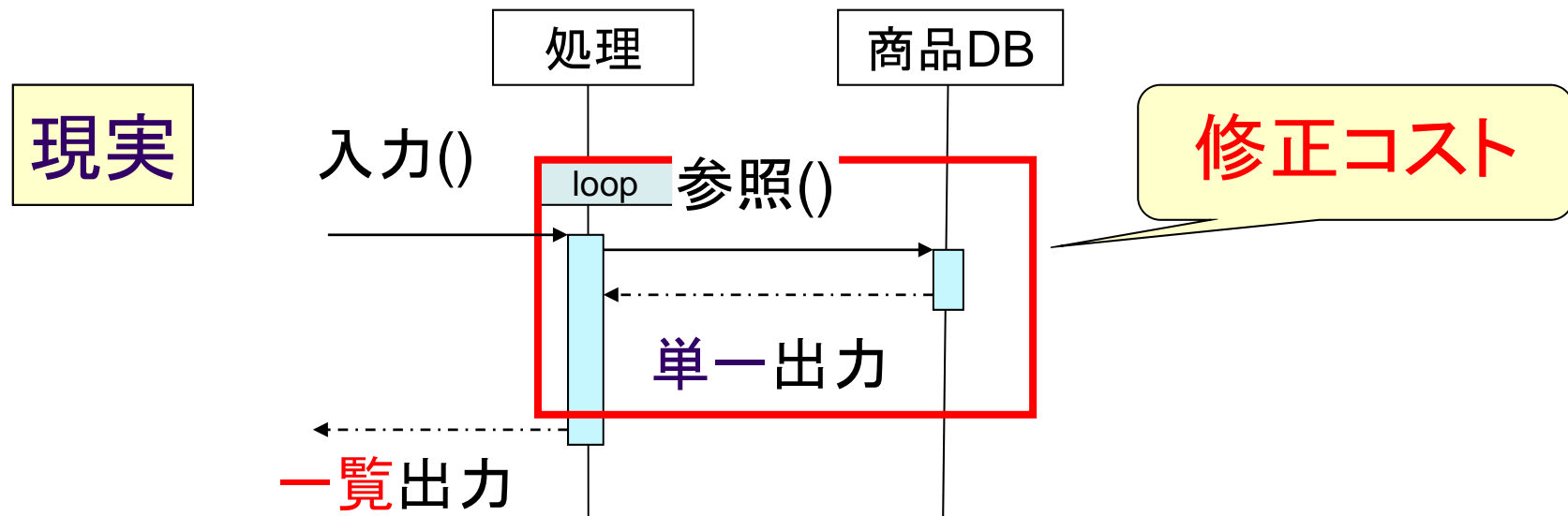
操作内容の異なる機能を検索結果から弾きたい

既存研究の問題点(2)

～入出力のデータ構造による分類



- 問題2: 振舞いの修正を要する機能を排除できない



入出力のデータ構造が異なる機能を検索結果から弾きたい

提案



Webサービスの操作カテゴリ分類手法を提案

1. 操作による効果的な絞込みができない

→ 操作内容による分類

※対象: 出力として情報を提供する機能

購入する機能等
は対象外

2. 振舞いの修正を要する機能を排除できない

→ 入出力のデータ構造による分類

提案手法(1)

～操作内容による分類



- **入力**と**参照データ**の有無によって網羅的に分類

参照データ: Webサービスが持つ情報
例: 商品の種類や値段の情報

操作種別	入力	参照データ
検索	あり	あり
情報取得	なし	あり
変換	あり	なし
該当無し	なし	なし

サービスとして
意味をなさない

検索者が直感的に絞り込みができるように名前を付ける

「**検索**」・「**情報取得**」・「**変換**」の3つの操作に分類

提案手法(2)

～入出力のデータ構造による分類



- 入出力の数が「単一」か「複数」かによって分類

- 単一: 同じ形式のデータが無い

商品	価格	在庫数
商品A	400円	10

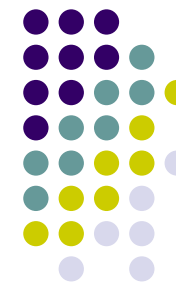
データ形式が
全て異なる

- 複数: 同じ形式のデータを持つ

都市名と天気の一覧データを持つ

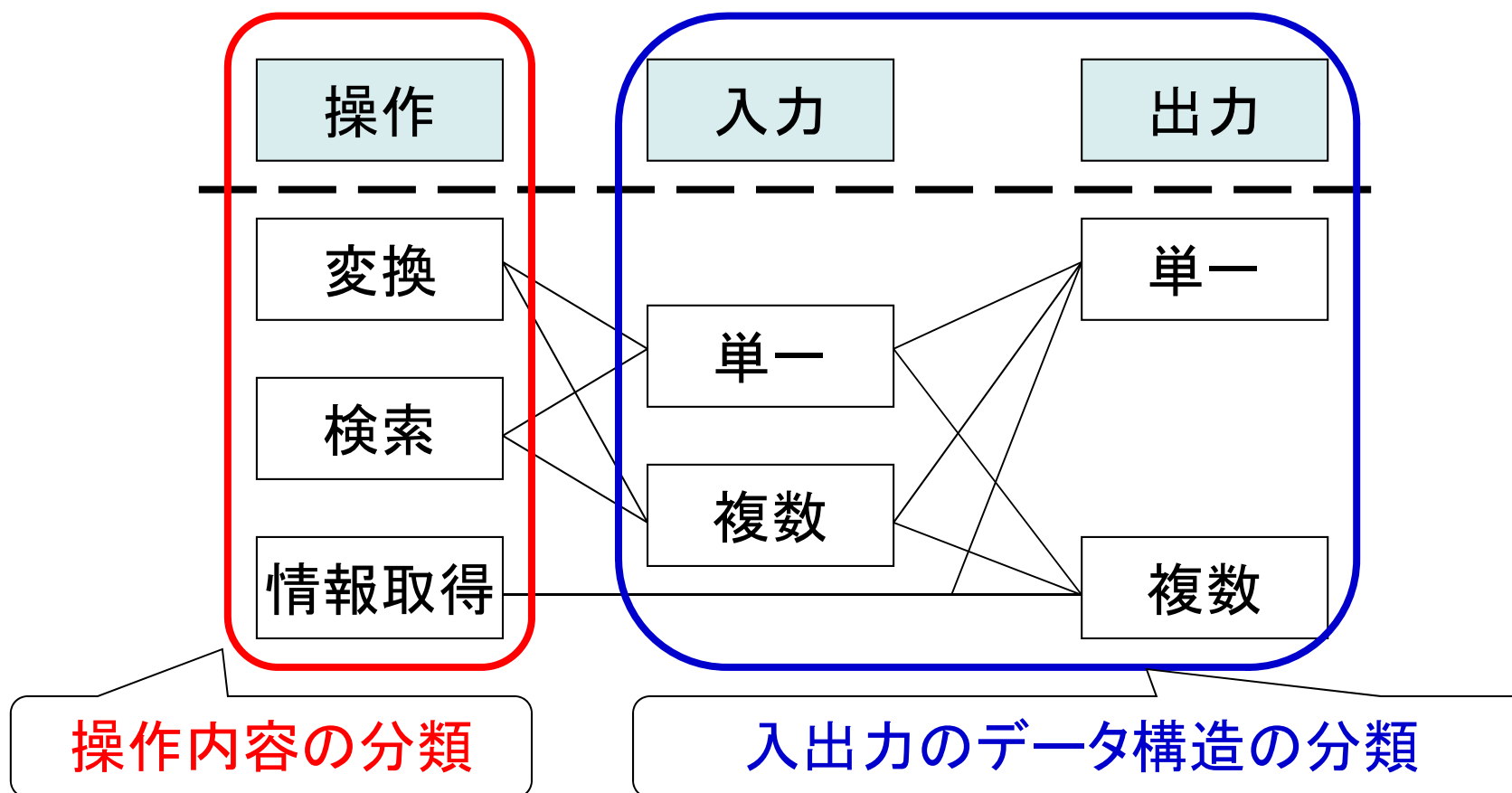
都市名	天気
東京	晴れ
横浜	晴れ
大阪	雨

入力・出力の数が「単一」か「複数」かによって分類



提案手法の最終的なモデル

操作内容と入出力のデータ構造によって分類





評価

- **提案手法**の有効性を**既存手法のみ**の場合と比較
 - 評価指標: Accuracy(検索におけるゴミの少なさ)
Recall(取りこぼしの少なさ)
F値(AccuracyとRecallの総合評価値)
 - データセット(機能)数: 25個(金融ドメイン)

手法	Accuracy	Recall	F値
既存	0.087	1.000	0.160
既存+提案	0.325	0.875	0.474

提案手法によって検索の**絞り込み精度が向上**



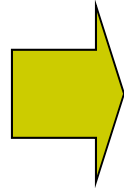
関連研究

- ソフトウェアコンポーネントの検索[H.Washizaki 01]
 - コンポーネントのプログラム構造や振舞いで分類
 - 情報の取得に重きを置いてない
 - 本手法を適用しても「変換」機能に集中してしまう
- Webサービスの自動カテゴリ分類[S.Saha 08]
 - 分類の判定情報を抽出
 - 操作の分類カテゴリを提案する本手法と位置づけが異なる
 - 本手法に併用することで操作カテゴリの自動分類が可能



まとめ

- **Webサービスにおける操作カテゴリ分類手法**
 - 操作内容による分類
 - 操作内容にあったサービスに対する絞り込みが可能
 - 入出力のデータ構造による分類
 - プログラムの構造にあったサービスに対する絞り込みが可能



Webサービスの**検索コストを低減**

- Future work
 - より詳細な操作種別の分類(「フィルタリング」・「ソーティング」等)
 - 情報を提供しない機能(購入する機能・保存機能等)