

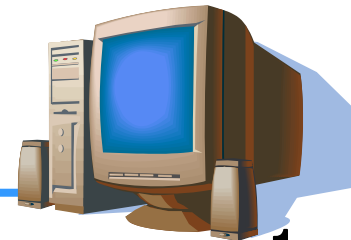
ビッグデータとフォグノードの統合モデル — 英語穴埋め問題自動生成システム

2021年 11月 29日

静岡理工科大学

情報学部 コンピュータシステム学科

工藤 司，山本 剛大，渡邊 智紀



1. はじめに
 2. 関連研究:ビッグデータとフォグコンピューティング
 3. ビッグデータとフォグノードの統合モデル
 4. 適用事例:英語穴埋め問題自動生成システム
 - 4.1 システムの構成
 - 4.2 実装
 - 4.3 評価
 5. まとめ
- 参考文献

1. はじめに

■ 研究の背景:ビッグデータの増大

- 業務システムにおけるデータ管理の課題
 - ✓ ユーザ独自の業務データ整備コストの発生
 - ✓ プライバシー, セキュリティによる「公開」の壁
- ビッグデータの課題
 - ✓ 膨大なデータが公開
 - ✓ データは個々のユーザの業務と整合しない

■ 研究の狙い

- ビッグデータによる**独自の業務データ整備の効率化**
- ビッグデータとfogノードの**データ循環の実装**

2. 関連研究:ビッグデータとフォグコンピューティング

■ ビッグデータの活用事例

1. 深層学習

- ✓ 膨大な訓練用画像, 実装モデルの提供

2. 英文翻訳

- ✓ 自動翻訳, 辞書, 英文校正サイト

3. 統計データ

- ✓ 新型コロナの発生状況, 推移情報の公開

などなど...

フォグ (エッジ) コンピューティング

クラウド
コンピューティング

クラウドサーバ

- データ解析
- 結果共有
- **フィードバック制御**

課題

- ✓ ネットワーク制約
- ✓ フィードバック遅延
- ✓ セキュリティ
(個人情報)

監視カメラ

フォグ
コンピューティング

クラウドサーバ

- データ解析
- 結果共有

課題

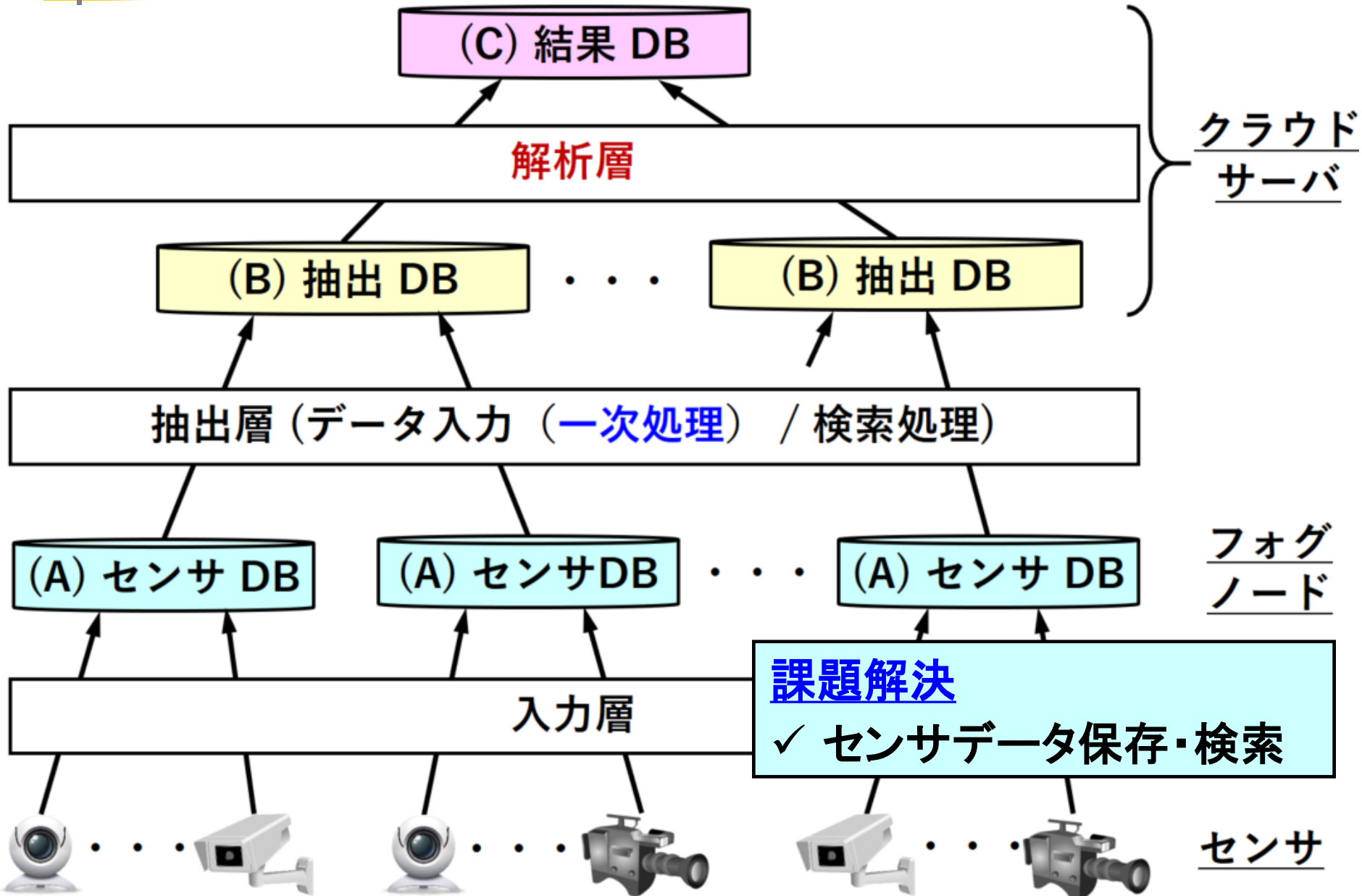
- ✓ センサデータ欠損

フォグノード

- 一次処理
- **フィードバック制御**

監視カメラ

フォグコンピューティング 3階層データモデル



3. ビッグデータとフォグノードの統合モデル

■ 狙い：ビッグデータと個別業務データの循環モデルの構築

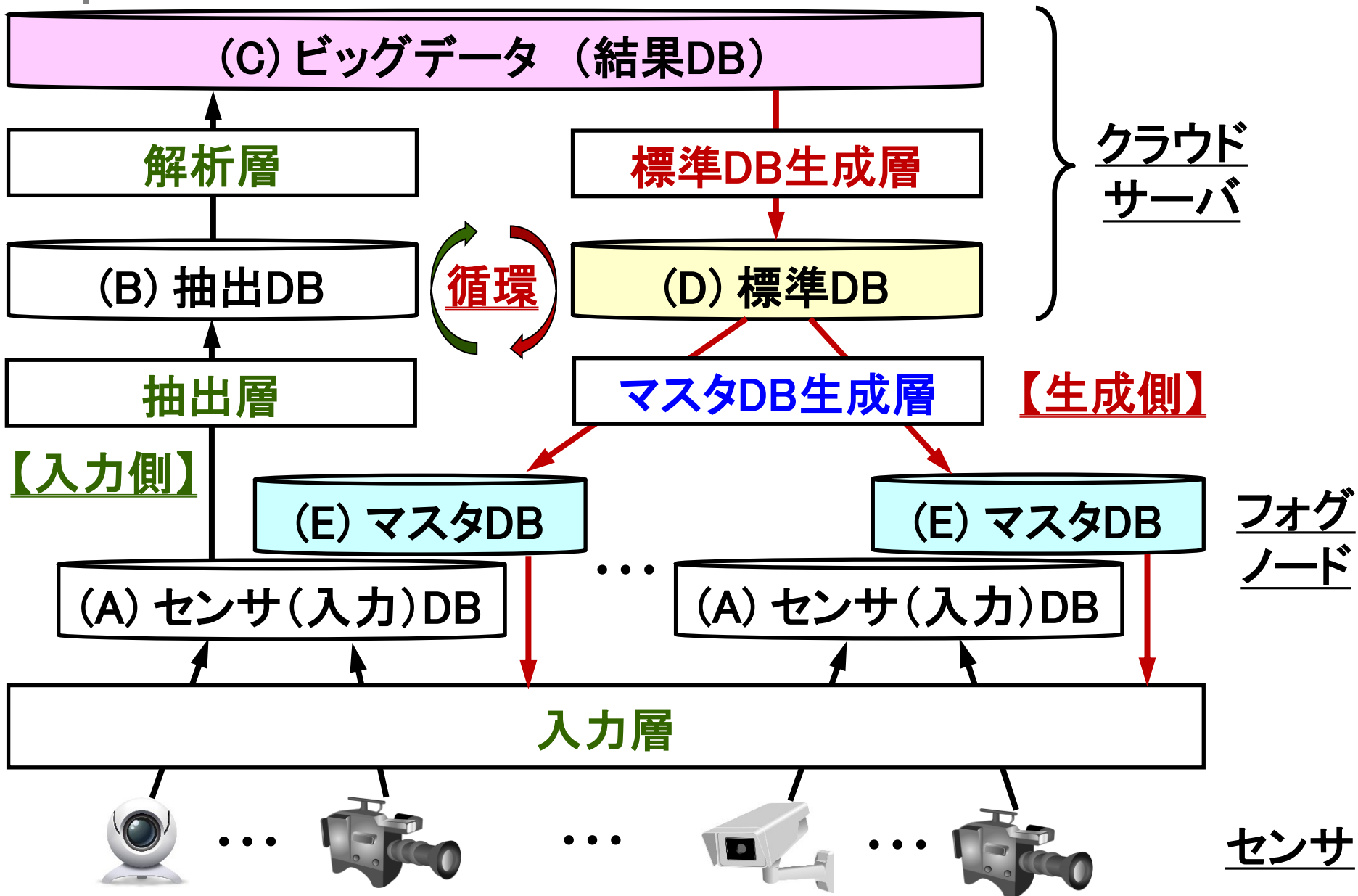
■ 生成側：ビッグデータから業務データを生成

1. 業務で共通的に使用可能な**標準DB**を生成
2. これを、**各業務に必要な形式**に変換して展開

■ 入力側：業務データをビッグデータにフィードバック

1. 非公開の情報は**センサDB**で管理
2. 公開可能な形式（統計など）で**フィードバック**

統合モデルの構成（入力側／生成側）



4. 適用事例：英語穴埋め問題自動生成システム

■ e ラーニング用 英語穴埋め問題の自動生成

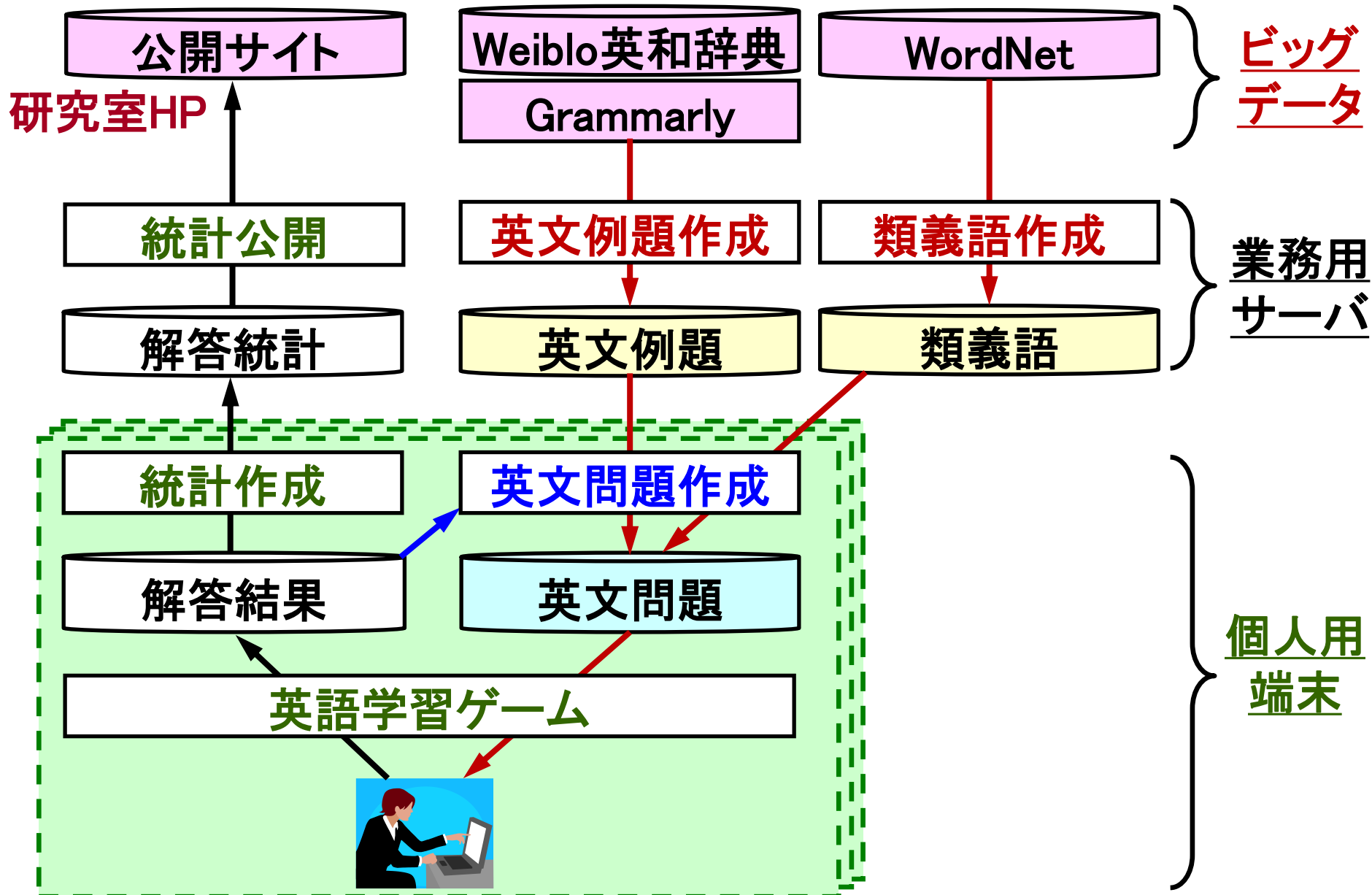
- a. 学習者のレベルに応じた問題の抽出
- b. ユーザ単位の学習・成績履歴の管理(個人情報)
- c. 学習効果のビッグデータへのフィードバック

【事例】

- 問題： Will they the people?
- 意味： それらは人々に影響しますか？
- 選択肢： like, affect, see, save

【参考情報】 正解：affect； レベル 2（準二級以上）

4.1 システムの構成



活用したビッグ・データ

1. Weiblo英和辞典 [2]

英単語の，意味，例文（意味），難易度のレベル。

2. Grammary [3]

英文校正サービスで，**英文の誤り有無が返される。**

本研究では，Google Chromeのアドオンとして，無償版（文法，スペル，句読点チェック）の範囲で使用。

3. WordNet [4]

英語の語彙データベースで，各語彙（単語）は異なる概念を表す**同義語のグループ**に分類される。

➤ 問題の選択肢（誤り）に正解混入を防止

テーブルの構成

【生成側】

1. 英文例題テーブル (標準DB:全ての単語, Weblio)

問題番号	単語	難易度	英文	日本語
14240	brain	2	a clear brain	明晰 (めいせき) な頭脳
14241	brain	2	She has both (*1)	彼女は才色兼備だ
14242	brain	2	Use your brain.	頭を使いなさい

(*1) She has both beauty and a good brain.

2. 英文問題テーブル (マスタDB:ユーザ毎に生成)

問題番号	単語	難易度	問題
14242	brain	2	Use your ____ .(頭を使いなさい) (下行へ続く)

選択肢 1 (正解)	選択肢 2	選択肢 3	選択肢 4	字数
brain	overdue	bays	worthwhile	22

【生成側】

3. 類義語テーブル (**標準DB**: 全ての類義語, WordNet)
 - 単語: brain の類義語の事例
(品詞は, 名詞(品詞=1), 動詞(品詞= 2))
 - 正解がbrainの場合は, **選択肢から同一グループの単語を除外**

単語	グループ	品詞
brain	1315218	2
brain	1389468	2
brain	5413129	1
brain	5541806	1
brain	5548525	1
brain	7547610	1
brain	9976200	1

テーブルの構成

【入力側】

1. 解答結果テーブル（入力DB:個人用端末）

学生番号	実行日時	出題番号	問題番号	難易度	結果
1	20210706100000	1	3199	2	1
1	20210706100000	2	3054	4	0
1	20210706100000	3	12278	5	1

注1: 出題番号は1回のゲームの通番,

注2: 問題番号, 難易度は「例題テーブル」のもの

注3: 結果は正解(1) / 不正解(0)

2. 解答統計テーブル（解答統計:業務用サーバで管理）

匿名	レベル	回数	正答数	誤答数
a	2	1	12	8
a	2	2	15	5
a	2	1	17	3

テーブルの構成

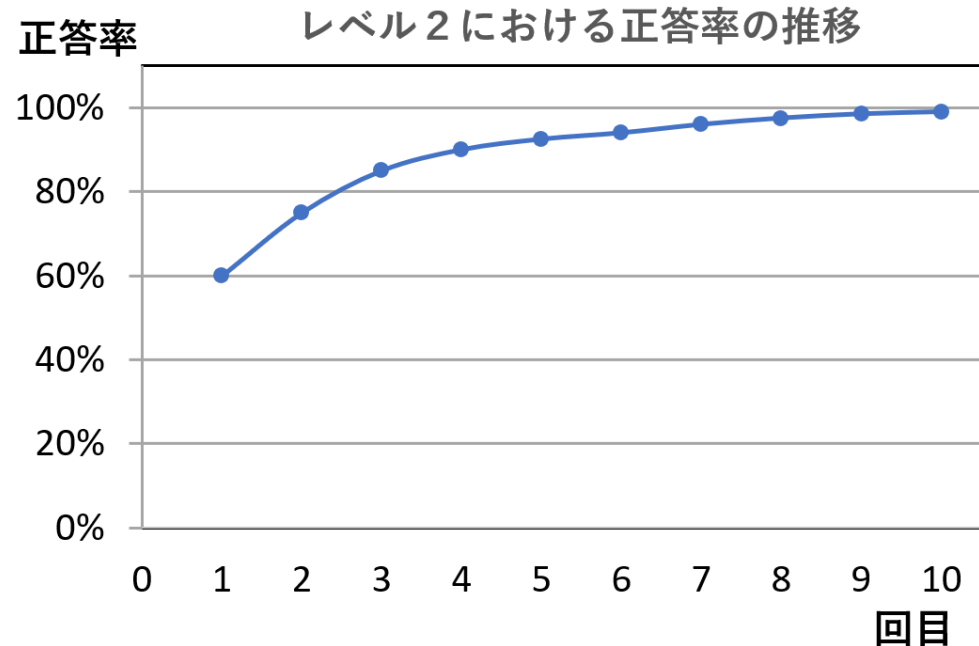
【入力側】

3. 公開サイト（ビッグデータとして公開）

a. 正答率データ（CSV）

レベル	回数	正答数	誤答数
2	1	1200	800
2	2	1500	500
2	3	1700	300
2	4	1800	200
2	5	1850	150
2	6	1880	120
2	7	1920	80
2	8	1950	50
2	9	1970	30
2	10	1980	20

b. 正答率推移グラフ



■ 実装環境

1. 業務用サーバ

- ✓ データベース: MySQL Ver. 8.0.22
- ✓ ブラウザ: Google Chrome Ver. 95.0.4638.69
- ✓ プログラム: Python Ver. 3.7.10
 - MySQL: MySQL Connector/Python Ver. 8.0.18
 - Web: Selenium Ver. 3.141.0 + Chrome Driver

2. 個人用端末

- ✓ データベース: MySQL Ver. 8.0.17
- ✓ データ変換: Excel 2019 + Excel VBA
- ✓ プログラム: HSP Ver. 3.6

■ 英文例題作成 機能の実装

1. Weiblo英和辞典のサイトをスクレイピングし、「英文例題テーブル」を生成.
2. Google Chrome上でGrammarlyを使用して、英語の例文をチェックし、**誤りがある場合はデータを除外.**
 - ✓ 対象:**文になっていない例など.**
 - 例: be lacking in affect (情動に欠ける)

■ 類義語作成 機能の実装

1. WordNet (CSV形式)の類義語(単語, グループ, 品詞)を, MySQLのテーブルに保存.
2. 「類義語テーブル」をビューとして実装.

■ 英文問題作成 機能の実装

Pythonにより、以下の機能を実装.

1. 「英文例題テーブル」から、「英文問題テーブル」のデータを生成
 - ✓ 例題は、ランダムに指定の件数を抽出
 - ✓ 選択肢は、単語をランダムに選択。「類義語テーブル」にある、正解の同義語は除外される.
2. 生成データは、CSVファイルにも保存.
 - ✓ MySQLをインストールしていない端末でも、「英語学習ゲーム」を実行可能にする.

類義語：単語－類義語ビューとして実装

問題番号	単語	難易度	問題
14242	brain	2	Use your ____ .(頭を使いなさい)

選択肢 1 (正解)	選択肢 2	選択肢 3	選択肢 4	字数
brain	overdue	bays	worthwhile	22

単語	類義語	グループ	品詞
brain	brainiac	9976200	1
brain	brainpower	5548525	1
brain	einstein	9976200	1
brain	encephalon	5413129	1
brain	genius	9976200	1
brain	head	5541806	1
brain	learning_ability	5548525	1
brain	mastermind	9976200	1
. . .			

■対象品詞(名詞: 1)

の類義語を検索

➤ 選択肢から除外

入力層の実装

■ 英語学習ゲームの実装

- シューティング・ゲームとして実装, 学習意欲向上を狙う.



問題

正解数/
誤答数

選
択
肢

■ 統計作成 機能の実装

1. 個人の学習履歴: 解答結果テーブル

個人の成績推移, 学習件数の推移などは, Excel VBAを使用し, 「個人用端末」内の機能として実装.

2. 公開用統計

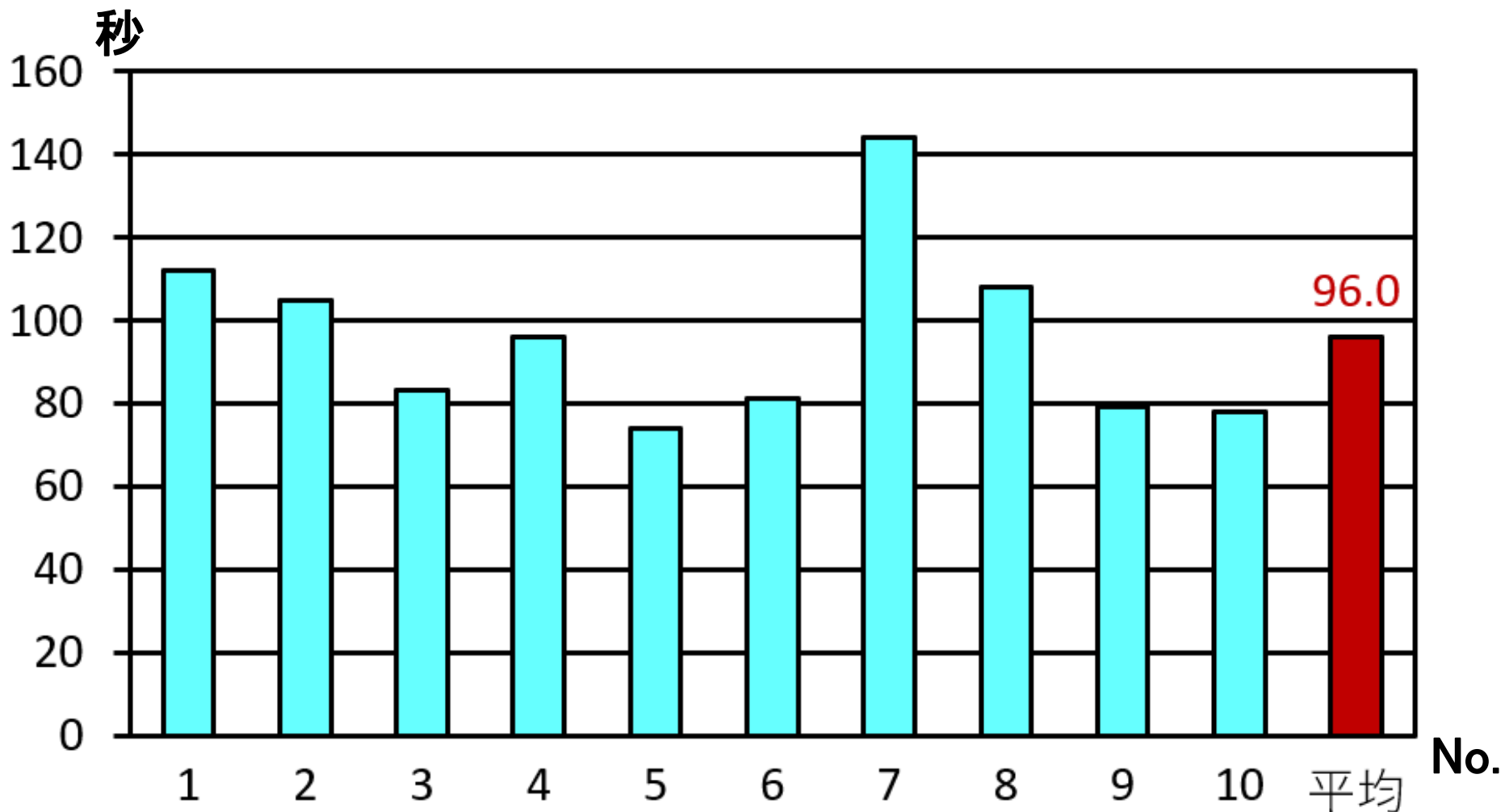
解答結果テーブルを, 解答統計テーブルに反映.

■ 統計公開 機能の実装

解答統計のデータをCSVファイルとして出力, **研究室のホームページ**に掲載し, ビッグデータに反映.

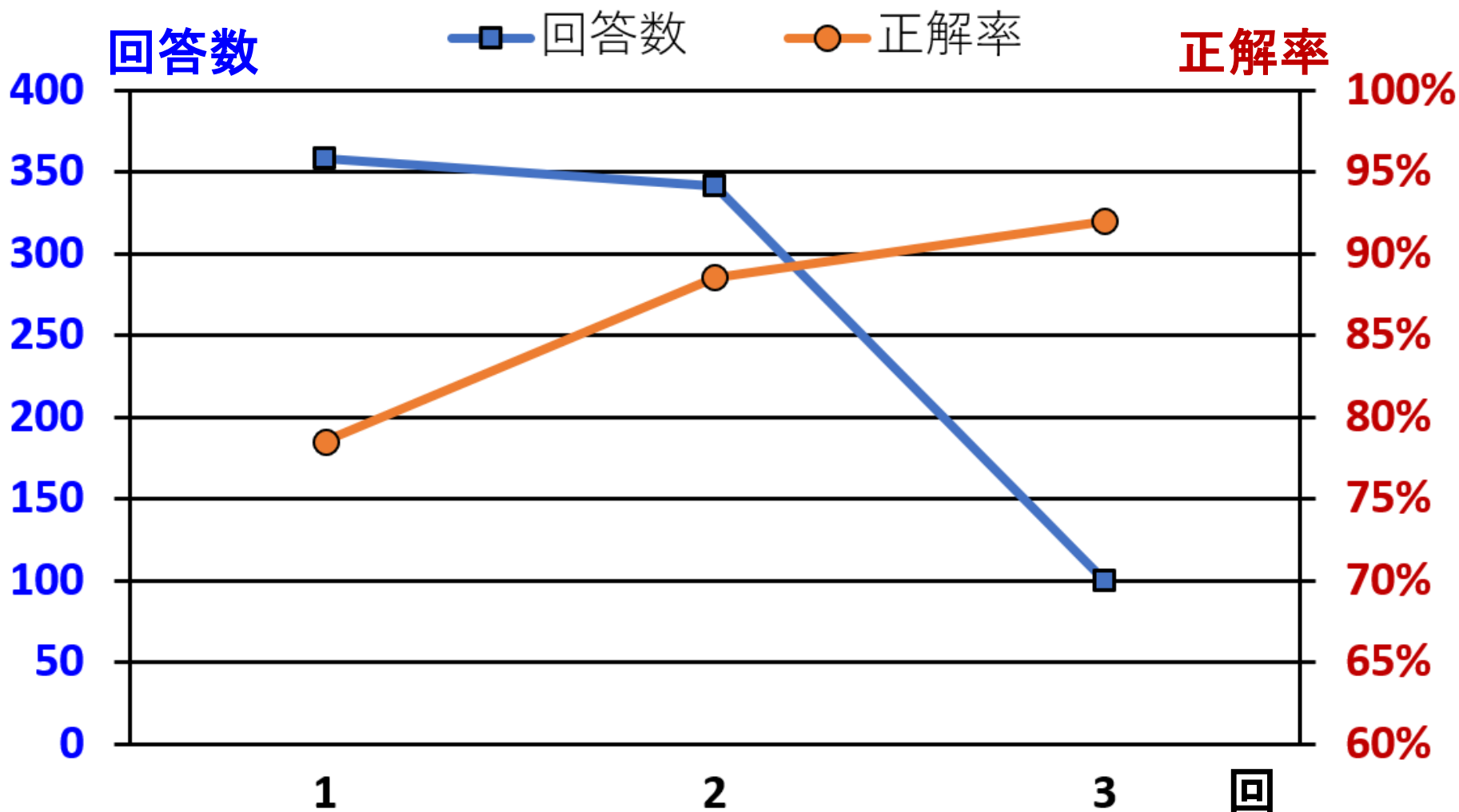
4.3 評価: 問題作成効率の評価

- 人手での問題作成(例題+選択肢): 平均 **96秒/1問**
- 本方式ではシステムが自動作成



4.3 評価: 学習成果の評価

- 同一問題(100問)を繰り返し, ゲームオーバー(回答数)まで実行
 - 学力(正解率)は向上; ゲーム数(回答数)は劣化(=< 疲労)



5. まとめ

■ 研究の背景

- ✓ システム導入で独自業務データ整備コストが発生
- ✓ 一方, ビッグデータとして膨大なデータが公開

■ ビッグデータとフォグノードの統合モデルの提案

- ✓ ビッグデータから階層的に個別業務のデータを生成
- ✓ セキュリティを維持してビッグデータにフィードバック

■ 英語問題自動生成システムに適用

- ✓ 業務データの生成効率化を確認
- ✓ フォグノードの活用によりプライバシーを維持

■ 将来研究: 他の事例によるモデルの有効性評価

■ ビッグデータ

- [1] Princeton University: WordNet, <https://wordnet.princeton.edu/> (2021年11月14日).
- [2] GRASグループ, Weiblo英和・和英辞典, ” <https://eje.weblio.jp/> (2021年11月14日).
- [3] Grammarly Inc., Grammarly, <https://www.grammarly.com/> (2021年11月14日).

■ 使用ツール, プログラム

- [4] Oracle: MySQL Connector/Python Developer Guide, <https://dev.mysql.com/doc/connector-python/en/> (2021年11月14日).
- [5] SFC: Selenium, <https://www.selenium.dev/> (2021年11月14日).
- [6] **Chromium project**: ChromeDriver, <https://chromedriver.chromium.org/> (2021年11月14日).
- [7] ONION software: HSPとは?, <http://hsp.tv/make/begin.html> (2021年11月14日).

■ 関連研究

- [8] T. Kudo: Fog computing with original data reference function, Int. J. Grid and Utility Computing, Vol. 10, No. 5, pp. 574–582 (2019).

本研究はJSPS科研費19K11985の助成を受けたものです.

ビッグデータとフォグノードの統合モデル ー 英語穴埋め問題自動生成システム

ご清聴ありがとうございました

2021年 11月 29日

静岡理工科大学

情報学部 コンピュータシステム学科

工藤 司, 山本 剛大, 渡邊 智紀

