

分散データベースによる学力調査分析システムの構築

An Analysis System with Distributed Database for Surveys of School Assessments

森薫[†] 木幡敬史[‡] 田中佳代子^{††} 玉村雅敏^{‡‡} 金子郁容[†]
Kaoru MORI[†] Takashi KOWATA[‡] Kayoko TANAKA^{††} Masatoshi TAMAMURA^{‡‡} Ikuyo KANEKO[†]

[†] 慶應義塾大学 政策・メディア研究科
[‡] 嘉悦大学 短期大学学部
^{††} 慶應義塾大学 SFC研究所
^{‡‡} 慶應義塾大学 総合政策学部
[†] Graduate School of Media and Governance, Keio Univ.
[‡] College of Business Communication, Kaetsu Univ.
^{††} Keio Research Institute at SFC
^{‡‡} Faculty of Policy Management, Keio Univ.

要旨

平成19年度から全国学力・学習状況調査が実施され、自治体独自の学力調査も実施されるなど、学力調査データの活用へのニーズが高まっている。また、調査を実施した自治体だけでなく、調査に参加した自治体や学校にもデータ活用の機会を提供することが重要である。しかしながら、膨大なデータを処理するためのインフラの整備が進んでおらず、その実現が課題となっている。本稿では、学力調査を対象としたデータ分析システムを、調査に参加した自治体と学校に提供することを想定し、利用者やデータ量の変化に応じて、計算資源をシームレスに拡張する分散データベース環境を示す。さらに、岩手県で実施した学力調査データによる評価実験を行う。

1. はじめに

教育施策の成果と課題を検証し、改善を図ることが求められている。その代表的な取り組みとして、児童生徒の学力や学習状況を把握・分析するために、平成19年度より全国学力・学習状況調査が実施されている。さらに、自治体独自の学力調査も活発に実施されるなど、学力調査データの活用へのニーズが高まっている。また、調査を実施した自治体が結果を分析するだけでなく、調査に参加した自治体や学校がそれぞれの視点で分析し、データ活用の機会を広く提供すること重要である。

現在、学力調査のデータ回収と分析を効率的に実施する Web アプリケーション[1][2]や、表計算ソフトに分析機能を組み込んだツール[3]の利用が進められている。しかしながら、膨大なデータを蓄積・分析するためのインフラの整備が進んでおらず、その実現が課題となっている。

情報システムが扱うデータ量の増大に伴って、計算資源やストレージを増強する必要がある。本システムは、増大するデータ量に応じてサーバ資源を増減させることが可能なデータベース環境において、学力調査のデータ分析機能を提供する。

以下、第2節では実現システムについて述べ、第3節では本システムの評価実験について述べる。最後に第4節で本稿のまとめを行う。

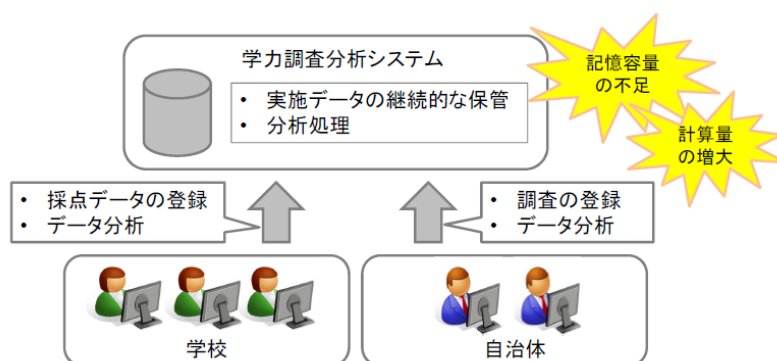


図1. データと計算処理の増大

2. 実現システム

本システムは以下の3サブシステムから構成される。

- 設問データ管理システム
- 採点データ管理システム
- データ分析システム

2.1. 設問データ管理システム

設問データ管理システムは、学力調査の設問構造を登録・管理するシステムである。設問構造を記述した TSV ファイルをアップロードすることにより、登録処理が実行される。設問構造には設問文と選択肢と正答情報が必須の入力項目となっている。また、オプションでメタデータを付与することも可能である。メタデータを付与することで、学習領域に応じた分析が可能となる。設問構造を登録するための TSV フォーマットを表 1 に示す。

表 1. 設問構造を登録するための TSV フォーマット

| 学年 | 科目 ID | 設問番号 | 設問文 | 正答番号 | 選択肢 1 | 選択肢 2 | ... | 選択肢 n | MD1 | MD2 | ... | MDm |
|----|-------|------|-----|------|-------|-------|-----|-------|-----|-----|-----|-----|
| | | | | | | | | | | | | |

2.2. 採点データ管理システム

採点データ登録システムは、学校教員が採点データを記述してアップロードした TSV ファイルが適切に記述されているか検証を行う。記述項目における値の一意性、および、範囲の検証を行う。具体的には、クラスや出席番号に重複がないか、設問の解答番号に登録していない選択肢情報が記述されていないか等を検証する。検証結果は科目ごとに教員に通知される(図 2)。教員は検証結果を確認し、エラーがあれば TSV ファイルを修正して再度アップロードする。

| 学年 | 国語 | 算数 | 社会 | 理科 |
|------|----|----|----|----|
| 小学1年 | | | | |
| 小学2年 | | | | |
| 小学3年 | | | | |
| 小学4年 | ⊗ | ⊗ | | |
| 小学5年 | ⊗ | ⊗ | | |
| 小学6年 | | | | |

図 2. 検証結果の通知画面 (×印はデータ削除のためのアイコン)

2.3. データ分析システム

データ分析システムは、学校から回収・登録されたデータを分析し、表示するシステムである。学校からアップロードされた TSV ファイルを保管し、その情報を分散データベースに格納する。大規模データを対象とした分析システムにおいては、利用者からのリクエストが発行されてから分析処理を開始すると、現実的な応答時間内に分析結果を得ることが困難である。本システムにおいては、バッチ処理によって事前にデータ分析し、その結果を分散データベースに格納する。TSV ファイルの保管には MySQL を利用し、分散データベースには Apache Hadoop の HBase を利用した。

データ分析システムでは平均点や選択肢別の回答状況、クロス集計などの分析が可能である。分析結果の出力結果例を図 3 に示す。これらの分析を、利用者のリクエストが発行されてから処理すると、現実的な応答時間内に分析を完了して表示することが不可能である。そのため、本システムではバッチ処理によって分析結果を事前に計算し、分散データベースに格納する。

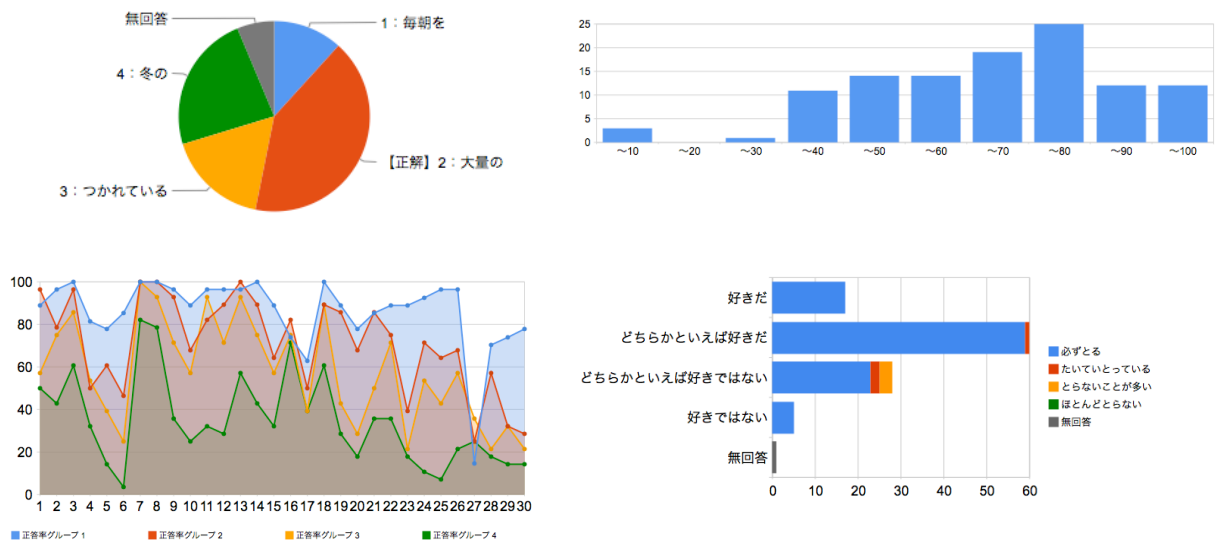


図 3. データ分析の結果例

(左上：回答パターン，右上：正答率ヒストグラム，左下：グループ別正答率，右下：クロス集計)

本システムにおけるクロス集計は、調査横断的に実行される。すなわち、単体の調査における項目間の関連性を分析するだけでなく、独立に実施された調査間における項目間の関連性を分析する[1]。この特徴から、時間的な経過に注目した項目間の関連性や、異なる性質の調査における項目間の関連性を分析することが可能となる。例えば、年度初めに実施した調査の項目と、年度末に実施した調査の項目をクロス集計することで、それらの因果関係が有意であるかを分析できる。また、学校においては学校評価に関するアンケートをはじめとして、学校生活や生活習慣を対象とした調査も実施されている。生活習慣の調査と学力調査をクロス集計することで、生活習慣が学力の形成にどのように影響しているかの分析を可能とする。

一方、本システムにおけるクロス集計をバッチ処理で計算した場合、そのデータ量は調査の追加とともに飛躍的に増大する。そのため、分析結果の格納方式として分散データベースを利用した。分散データベースを利用することにより、データ量の増大に応じてストレージ容量を追加的に拡張することが可能となる。

3. 実験

本システムの分析処理において、分散データベースを構成するノードの追加が計算処理のパフォーマンスに影響しないことを検証するための実験を行った。米アマゾン社が提供するクラウドコンピューティング環境の Amazon EC2 の Small Instance (表 2)上に実験環境を構築した。EC2 Compute Unit とは、Amazon EC2 において規定されている処理能力の単位である。また、岩手県で実施された平成 22 年度岩手県学習定着度状況調査のデータ、および、仮想的に作成した学力調査データで評価実験を行った。調査間のクロス集計をするバッチ処理プログラムを起動し、項目あたりの平均計算時間を計測した(図 4)。ノード数を増加してストレージ容量を拡張しても、処理時間が極端に増加しないことを確認した。

表 2. Small Instance のスペック

| | |
|--------|--------------------|
| CPU | 1 EC2 Compute Unit |
| Memory | 1.7 GB |
| HDD | 160 GB |

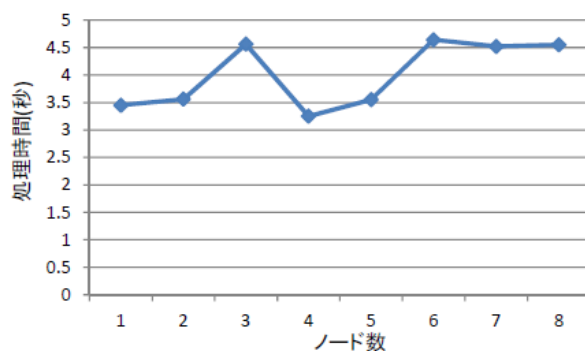


図 4. ノード数と処理時間の実験

4. おわりに

本システムでは、分散データベース上に格納された学力調査データを対象として、調査に参加した自治体や学校のためのデータ分析環境を実現した。本システムは、増大するデータ量に応じて計算資源を拡張することが可能である。また、分散データベースを構成するノードの追加と処理時間について実験を行った。今後は学力調査を実施している自治体と連携して、システム運用の評価を進める予定である。

参考文献

- [1] 森薫, 木幡敬史, 田中佳代子, 玉村雅敏, 金子郁容: “学校評価を対象とした調査間の透過的データ結合機能を有するデータ分析システムの構築”, 教育システム情報学会 第 35 回全国大会 講演論文集, pp.443-444 (2010)
- [2] 森薫, 木幡敬史, 玉村雅敏, 嶋津恵子, 金子郁容: “Web ベースの学力テスト分析システムの構築と評価”, 情報システム学会 第 4 回全国大会 研究発表大会 (2008)
- [3] 佐賀県教育センター: “全国学力・学習状況調査分析ツール提供要領” (2009)