

多量の学習履歴データの解析を支援する基盤システムの構築

足立 紘 亮^{†1} 上田 敏 樹^{†1} 新 村 正 明^{†2}

学習行動履歴を作成するためには、学内の複数のシステムからデータを収集する必要がある。しかし、データ量が非常に多いことや、システム毎にデータ形式が異なることなどの問題点がある。本稿では、学内に分散するシステムからデータを収集し、解析をするプラットフォームを提案する。このプラットフォームは、MapReduceにより構成する。さらに、このプラットフォーム上でLMSのアクセス履歴を解析するシステムの試作を行った。

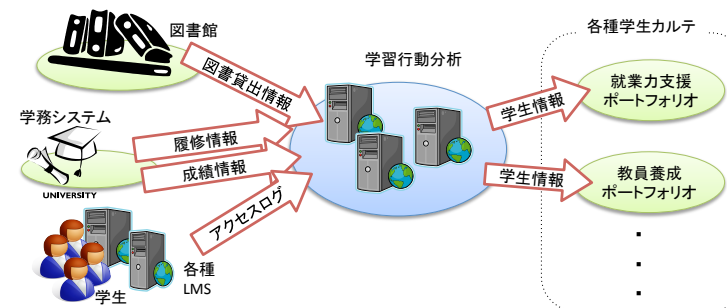


図 1 学生行動履歴の収集

Fig. 1 Learning History Data Collection

Large-Scale Data Analysis Platform for Learning History Data

KOSUKE ADACHI,^{†1} TOSHIKI UEDA^{†1}
and MASAOKI NIIMURA^{†2}

It should be implemented to collect data from different systems in campus to make the learning history of on-campus activities. Since the difference of data format between the systems and the volumes of data, it is very difficult to analyze the activities. In this paper, we propose a platform to collect learning history data from some LMSs and analyze these data based on the MapReduce. And we develop a prototype system of access analysis system on this platform.

1. はじめに

学習者の学習履歴情報は、学習者評価の情報として極めて有用なものであり¹⁾ その収集や評価について数多くの研究がなされている。このような学習履歴情報は、LMS(Learning

Management System) 等によって蓄積されているが、学習過程全体を対象とする場合には、単一のLMSだけでなく、学内に存在にする複数の学習支援システムからデータを収集する必要がある。

また、単に学習履歴の収集だけでなく、受講選択科目・成績等の学務情報や、図書館での書籍貸出等の情報など、学生が学習を進める上でどのような行動をとったのかを表すデータを収集することにより、学習者に対するきめ細かい指導・助言を行うことが可能となる。このような学習過程全般にわたる行動の履歴を、本稿では学習行動履歴と呼ぶこととする。

このような情報収集に関しては、学習行動履歴の収集にとどまらず、IR (Institutional Research) と呼ばれる、学内に分散する情報を収集・整理し、大学運営や企画立案のために提供する仕組みも注目されている。

しかしながら、このような学習行動履歴の収集には、LMSや学務システム等、複数の異なるシステムのデータを処理しなければならないことや、学生生活全般にわたって収集するために複数年度のデータを扱う必要がある、データ量が膨大になるという問題点がある。

そこで、本研究では、多量の学習行動履歴を効率的に処理する基盤システムを提案する。また、この手法を用いて複数年度のLMSの利用状況を集計するシステムの試作を行った。

2. 学習行動履歴の収集

2.1 学習行動履歴の把握

学内に分散して存在する学生の行動や履修履歴の情報を収集し、学習行動履歴として分析

^{†1} 信州大学工学系研究科

Graduate School of Science and Technology, Shinshu University

^{†2} 信州大学 e-Learning センター

Shinshu University e-Learning Center

することで、単一の LMS からでは得られない、学習過程全般にわたる情報を得ることができる。

また、学習過程全般にわたって収集したデータは、図 1 に示すように、学内の様々な学生支援システムに対して、学生の状態を示す情報として提供することが可能である。

このように、学習行動履歴の収集は、学生の状態の把握や、きめ細かい指導・助言のために極めて有効である。

2.2 学習行動履歴収集の課題

学習行動履歴を収集し、処理する場合、以下のような問題がある。

(1) 多様なデータ形式

学内に存在する様々な情報源から学習行動履歴を収集することから、それぞれ異なった形式のデータを扱う必要がある。

例えば、複数の異なる種類の LMS を運用している場合には、各々の LMS でアクセス履歴を保持する形式が異なる。また、1つの LMS のアクセス履歴であっても、稼働中の LMS ではアクセス履歴をデータベースに保持し、前年度以前の履歴はファイルとして保持している場合には、収集方法を変える必要がある。

(2) データ量

LMS 等で収集される学習履歴は、非常に細かい単位での行動まで把握することができる。しかし、このためには、非常に多量のデータを扱わなければならない。

例えば、約 11,000 人の学生を擁する信州大学の場合、中心的に使用されている LMS は、約 20% の授業科目で利用されているが、年間 400 万件近いアクセス履歴があり、今後も、利用数の増加が見込まれている。また、単年度の情報だけでなく、入学から卒業までの行動を把握するためには、4 年分のデータを処理する必要もある。

3. 解析支援基盤システム

3.1 学習行動履歴収集方式

本研究で提案する学習行動履歴収集・処理の方式を、図 2 により説明する。

(1) フィルタによる形式変換

収集先のシステムにより、保持するデータの項目や形式、またファイル・データベース等のデータ保存手段が異なる。

そこで、収集先のデータから、集計に必要なデータを抽出し、統一形式に整形するフィルタを作成することとする。フィルタは、収集先のシステム毎に作成し、データの保存手段や、

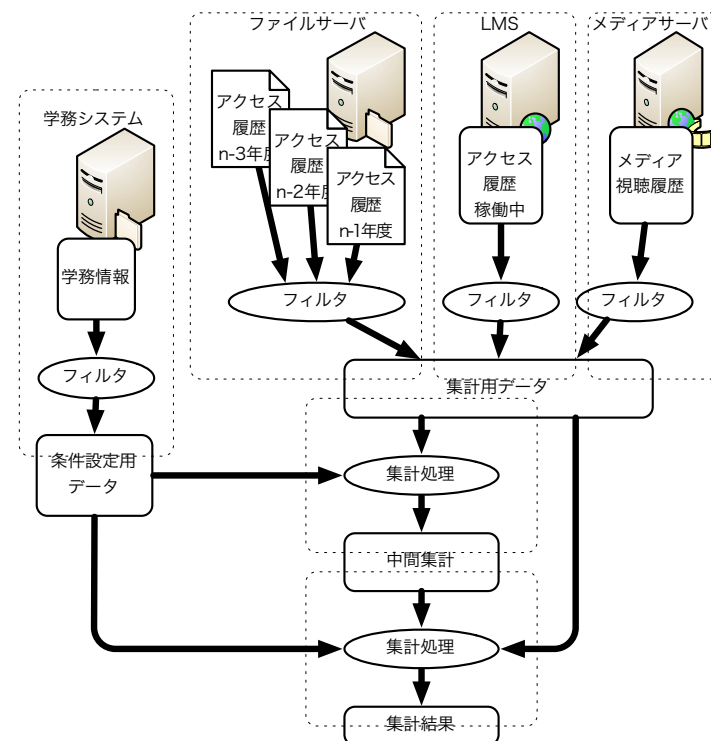


図 2 学習行動履歴の処理手法
Fig.2 Processing Method for Learning History

項目・形式の違いを吸収するものとする。

図2では、アクセス履歴やメディア視聴履歴を集計用データに変換する部分が、これに相当する。

(2) フィルタによる段階的な処理

収集されたデータには、LMSのアクセスログのような学習履歴を示すデータや、科目の選択や図書の貸し出しのような意志決定に基づく行動の履歴が含まれる。さらに、特に後者のデータは、学習履歴の抽出の条件として使用される場合が多い。例えば、ある科目を受講している学生の学習行動履歴の抽出などである。

このような別のデータの条件に基づくデータ抽出や集計を効率的に行うために、段階的なデータ処理を行うこととする。

具体的には、各段階における処理を、一時ファイルの内容に対して処理を行った結果を次の一時ファイルに書き出すフィルタにより構成するものとし、集計や抽出をいった処理を、このフィルタにより実現する。

この方法により、複数の巨大な一時ファイルが生成されるものの、集計処理をパイプライン化することが可能であるため、効率的なデータ処理が期待できる。

図2では、集計処理を順次実行する部分が、これに相当する。

3.2 MapReduceによるプラットフォームの構築

前項で述べた、フィルタによる形式変換や段階的な処理を実現するために、Googleによって考案された多量のデータを分散並列で処理するフレームワークであるMapReduce²⁾を使用して、プラットフォームの構築を行った。

今回はMapReduceのオープンソースの実装であるHadoopを用いた。また、中間ファイルの保持に関しては、Hadoopによる処理との親和性を考慮し、Hadoop分散ファイルシステムを使用した。

4. アクセス集計システムの試作

本提案に基づく解析プラットフォーム上で、LMSのアクセス集計システムの試作を行った。試作した集計システムによる処理結果の例を、図3、4に示す。これらのデータは、信州大学e-Learningセンターが運営するLMSであるMoodleの2008,2009,2010年度のアクセス履歴、約12,000,000件を処理したもので、集計結果を表計算ソフトによりグラフ化したものである。

図3は、時間別アクセス数の積み上げ集計である。Onは学内からのアクセス、Offは、

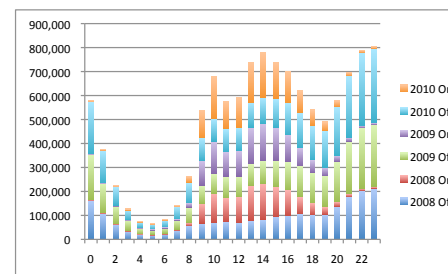


図3 時間別アクセス数
Fig.3 Hourly Access

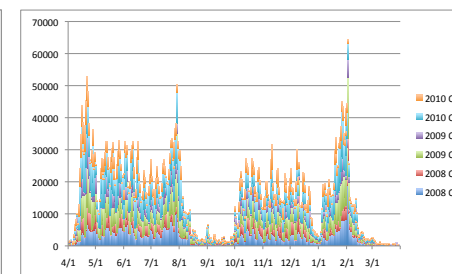


図4 共通教育科目の日別アクセス数
Fig.4 Daily Access of General Education

学外からのアクセスを示す。また、図4は、日別アクセス数のうち、共通教育科目へのアクセスのみを抽出したものの積み上げ集計である。同様に、Onは学内からのアクセス、Offは、学外からのアクセスを示す。

非常に大量のデータであるが、特に問題なく処理できることを確認した。

5. まとめ

学習行動履歴の収集・解析のために、学内に分散する様々なシステムから学習履歴等を収集し、処理するプラットフォームの構築を行った。また、このプラットフォーム上で、アクセスログ解析システムを試作し大量のデータであっても問題なく動作することを確認した。

今回は、解析プラットフォームの構築にとどまったが、今後は、データ解析自身を支援するシステムの構築を行う予定である。

参考文献

- 1) 松居 辰則, 岡本 敏雄.: 学習履歴情報の Digital Portfolio 化とその利用について, 電子情報通信学会技術研究報告. ET, 教育工学, Vol.101, No.397, pp.25-30 (2001).
- 2) DEAN J.: MapReduce, Simplified Data Processing on Large Clusters, *Communications of the ACM*, Vol.51, No.1, pp.107-130 (2008).