

## 学生SD のためのデータ解析の可能性と重要性 : 学生の仕事力養成のための1つのアプローチ

南, 俊朗

九州大学附属図書館研究開発室 : 特別研究員 | 九州情報大学附属図書館 : 館長・教授

<https://doi.org/10.15017/27252>

---

出版情報 : 九州大学附属図書館研究開発室年報. 2012/2013, pp. 19-28, 2013-09. 九州大学附属図書館  
バージョン :  
権利関係 :



論文

## 学生 SD のためのデータ解析の可能性と重要性 －学生の仕事力養成のための1つのアプローチ－

南 俊朗<sup>†</sup>

<抄録>

大学にとって最も重要な課題は教育力の向上である。大学の一部署である大学図書館にとっても学生支援は大きな課題である。昨今は少子化などの社会的変化に伴い入学してくる学生の多様化が進んでいる。我々はデータに基づいてこのような学生の実像を把握する研究を行ってきた。図書館データとしては特に貸出記録に注目し、授業データとしては出席、宿題、試験、そして学生の自己・授業評価アンケートを用いる。本稿では、大学の教育力向上、特に社会人・職業人としての教育力の観点からこれらの研究結果を総括し、今後の方向性を議論する。

<キーワード> 学生 SD (Student Development), 図書館データ解析, 図書館マーケティング, 授業データ解析

## Potential and Importance of Data Analysis for Student Development －An Approach to the Development of Student's Vocational Capacity－

Toshiro MINAMI

### 1. はじめに

教育機関である大学にとって学生に対する教育力の向上は、その最大の課題である。その一部署である大学図書館にとっても学生への学習支援の重要性は極めて大きい。一方、現在の学生達はかつてないほど多様化している。その背景には少子化や大学進学率の増加、社会状況の変化など様々な要因が存在しているものと考えられる。その結果、多様化に留まらず学生の学力低下という現象が起こっている[13]。十分な基礎学力を備えないまま大学生になってしまった学生に対する高等教育の実現が大学に求められている。

大学における教育機能の主たるものは授業である。授業における第1の当事者は学生であり、第2の当事者は教員、そして第3の当事者が職員や学生を取り巻く様々な人々である。現在多くの大学は授業改善などを目的としてFD (Faculty Development) 活動を導入し、第2の当事者である教員の教育力向上に努めている。またリメディアル教育や初年次教育を強化するなどし、学生への多様な授業提供にも多大な努力を傾注してきた。職員のためには職員 SD (Staff Development) 活動を実施している。しかし、そのような努力にもかかわらず、状況の大幅な改善には結びついていない。

その大きな原因は、授業の第1の当事者である学生側が授業の受講姿勢や学びへの意欲などに関して十分ではないことにあるものと考えられる。大学側としても、学生の心の持ちようも視野に入れた学生教育(学

生 SD, Student Development) への取り組みにこれまで以上の力を注ぐことが求められている。

現在多くの大学において授業改善を目的にFDの一端として学生による授業(評価)アンケートが実施されている。しかし教職員の多大なる労力をかけているにも関わらず、学生SDに注目し、授業の効果をより一層高めようと努めている大学は少ない。授業アンケートに関する研究論文の中で学生SDの観点から考察しているもの(たとえば[25])もまだ少ない。

学生SDを考える上で重要なのは、学力それ自体だけではなく、学びに対する学生の姿勢や意識、知的好奇心や社会で起きていることに対する関心、身の回りの出来事への観察力やその中に存在する課題を発見し、それを解決したいという意欲など、学力や仕事力を支える基盤となる能力である。いわゆるゆとり教育などの結果引き起こされたとみられる「低学力」が問題として注目されることが多いものの、低学力の学生の多くはこのような「基盤力」、あるいは「社会人基礎力[8]」も低く、低学力は基盤力の弱さから引き起こされているケースが多いものと考えられる。

このような観察結果を踏まえると、大学が担う高等教育の目的は、学生を、単に高度な学力を持った人材としてだけではなく、高度な仕事力(職業人力)をも備えた人財として養成し、社会に送り出すことであると言える。実際、就職活動において内定を獲得するためのアドバイスとして論理力、適応力、好奇心や成長

<sup>†</sup> みなみ としろう 九州大学附属図書館研究開発室特別研究員, 九州情報大学附属図書館長・教授 (〒812-8581 福岡市東区箱崎 6-10-1)  
E-mail: minami@kiis.ac.jp

力などが指摘されており[7]、学生 SD の重要性が示唆されている。

我々はこれまで図書館マーケティング研究の一環として資料の貸出データを中心とした図書館データの解析法に関する研究を行ってきた[9-12, 15, 18-19, 21]。その一環として、学部生、修士・博士の学生、教員などの利用者タイプを基に図書や学生の専門度なる概念を定義し、それを利用して学生個人や学部の専門度に関する考察を行い[12]、また図書の NDC (日本十進分類) を利用して学生個人や学部の興味分野に関するパターンの広がりや類似性などを調べた[15, 18-19, 21]。これらの研究は社会人基礎力の中の傾聴力や柔軟性、実行力などと関連した資質と考えることができる。

我々はまた、図書館データ解析と同様の手法を適用することにより、出席や宿題などの平常点、試験点、そして学生自身および授業に対する状況や評価に関するアンケート結果などの授業関連データを分析することにより、授業や学びに対する学生の姿勢などを調べた[13-14, 16-17, 20, 22]。これらの手法は職業人力としての、与えられた課題に真剣に取り組む態度や自分を客観化して評価し、それに基づいて自己の成長を図る能力に関する指標であると理解でき、それに基づいて学生の職業人力を養成したり改善したりするためのヒントを得る手助けになる。

本稿では、これらの経験を土台に学生 SD を意識した図書館・授業データ解析の可能性を議論し、その重要性を訴える。この目的を達成するために本稿は以下次のように構成される。

第2節では、これまで行った図書館データ解析の結果を概観し、これまでに得られた結果が学生 SD にどのように役立てることができるか考察する。

第3節では、これまで行った授業データ解析の結果を概観し、これまでに得られた結果が学生 SD に役立つ知見を検討する。授業アンケートに関しては、従来研究の結果を踏まえて、データ解析の結果がどのように利用可能であるかを詳しく議論する。

第4節では、職業人育成のためのデータ解析の可能性について検討し、今後の研究の方向性やその可能性ならびに重要性について議論する。

最後に第5節で、本稿の議論全体を総括する。

## 2. 図書館データ解析からの知見

本節では、図書館データの解析から学生 SD に関するどのような知見が得られる可能性があるのかを考察するための材料として、我々がこれまでに行った解析結果を検討する。

まず 2.1 節で対象とした図書貸出データの概要を示す。次に、2.2 節では、図書が一般読者向けなのか、専

門書なのかを貸出状況から判断する 1 つの試みとして専門度推定を行い、その結果の応用として貸出図書から評価される利用者の専門度も推定する。さらに、2.3 節においてある利用者の興味の状況を貸出図書の日本十進分類 (NDC) の多様性に基づいて調べる試みを紹介する。これらの試みは、図書館の貸出データという、かなり特殊なデータからも、学生の社会人力という、もっと一般的な能力に関する情報が得られる可能性を示している。

### 2.1. 解析対象となる図書館データ

本稿で解析する図書館データは九州大学附属図書館中央図書館における 2007 年度の貸出記録である。各レコードは 1 冊の貸出および返却に関するデータを集めたものであり、貸出した利用者、その所属、身分種別 (学部何年生・修士・博士、教員、職員など)、貸出された図書の ID やタイトル、NDC 分類番号、貸出日時、返却日時などから構成される。貸出レコードの延件数は 67,304 件である。

この貸出レコードによると、開館日数 (貸出が行われた日数) は 348 日である。したがって 1 日当たりの貸出冊数は約 193 冊となる。利用者 (貸出のあった利用者) の総数は 6,118 名である。なお、図書館を利用しない利用登録者や来館していても 1 年間全く貸出をしなかった利用者は本統計から除外されている。その上で計算すると (貸出) 利用者 1 人当たりの年間延貸出冊数は 11 冊である。

貸出件数に占める身分種別の割合を見ると、全貸出件数 67,304 件中 48% (32,609 件) は学生 (学部学生 29,698 件、その他研究生等 2,911 件) による貸出であり、以下、修士学生が 23% (15,800 件)、博士学生が 16% (10,460 件)、その他が 13% (8,435 件) である。入館ゲートでの記録でも入館者の約 6 割が学部学生であり、大学図書館の主な利用者が学生 (特に学部の学生) であることがデータによっても裏付けられる。

貸出全体の半分近くを占める (学部) 学生の貸出データ 29,698 件に限ると、図書を借りたことのある学生数は 2,966 名であり、一人当たりの年間平均貸出延冊数は 10.01 冊である。

学生個人の年間貸出件数を見ると、最大の貸出件数は 208 件 (1 名) である。それに引き続き 181 件、172 件、143 件、141 件 (それぞれ 1 名) となる。貸出件数が 100 件を超える学生は 12 名 (0.4%) である。

逆に少ない方を見ると、最少の貸出冊数は 1 冊であり、468 名 (全対象学生の 16%、6 名中 1 人の割合) が該当する。それに引き続き、2 冊は 413 名 (全体の 14%)、3 冊は 283 名 (全体の 10%)、4 冊は 233 名 (全体の 8%)、5 冊は 196 名 (全体の 7%) となる。

## 2.2. 図書の専門度推定と利用者プロフィール

ある学生がどの程度専門性の高い内容を学んでいるかは学習態度に関するその学生の重要なプロフィール情報であると考えられる。学生の専門度の指標として学生が借りた図書の専門度を採用する。図書の専門度なる指標を定義し、学生が借りた図書の専門度の平均値をもって、その学生の専門度とする。そのためには、まず図書の専門度なる概念を定義する必要がある。我々は  $t$ -値( $t$ -value)と  $a$ -値( $a$ -value)なる2つの指標を考案した[17]。以下順に紹介する。

$t$ -値は学部学生に良く借りられる図書よりも、教員に良く借りられる図書の方が専門度が高いという考えに基づく。具体的にはまず、利用者タイプの専門度を次のように定める。1年次から6年次の学部学生に対して1から6を専門度として与える、修士・博士・教員のそれぞれに8, 9, 10を専門度として与える。それ以外のタイプの利用者の場合はカウントに入れない。ある図書の  $t$ -値は、その図書に関する貸出記録の利用者の専門度の平均値とする。

この定義によると学部学生が主として借りる図書の  $t$ -値は概ね4以下になり、また、修士以上の利用者が主に借りる図書の  $t$ -値は8以上になる。前者の図書を学部レベル、後者を研究者レベルと呼ぼう。それらの中間である  $t$ -値5から7は相当する利用者が極めて少数であるため、このレベルの図書は学部学生と大学院生以上の両方が混合して借りられていることになる。

$t$ -値による図書の専門度の冊数分布を図1に示す。学部レベル、混合レベル、研究者レベルのそれぞれ約40%、26%、34%という割合である。

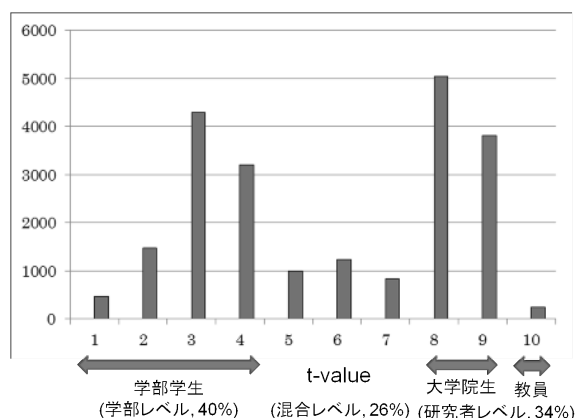


図1 図書専門度 ( $t$ -値) の分布

専門度を測るもう1つの指標である  $a$ -値( $a$ -value)は専門度が低い図書は多くの分野の利用者に借りられる一方、専門度の高い図書は特定の限られた分野の利用者のみに借りられるという考え方に基づく。学部を分野として採用する。対象データにはアルファベット順に AG (農学部), DD (歯学部), DS (芸術工学部),

EC (経済学部), ED (教育学部), LA (法学部), LT (文学部), MD (医学部), NC (21世紀プログラム), PS (薬学部), SC (理学部), そして TE(工学部)の12学部が含まれる。それぞれの図書に関して、まず、それぞれの学部のメンバの貸出冊数総数が全貸出冊数に対してどういう割合になるかを計算する。次に偏りの程度を表す値として割合を2乗した合計値を求める。

特定の1学部だけに貸出が集中した図書はその学部だけ割合値が1となり、その他の学部は0となるため、2乗の合計は1と最大値になる。一方全ての学部が同じ冊数借りた場合、それぞれの学部の割合値は  $1/12$  となる。その2乗値12個の合計は  $1/12$  となり、これが最小値である。この指標では最大値が1となるため、 $t$ -値と最大値をそろえるために、この指標値を10倍した値を  $a$ -値とする。

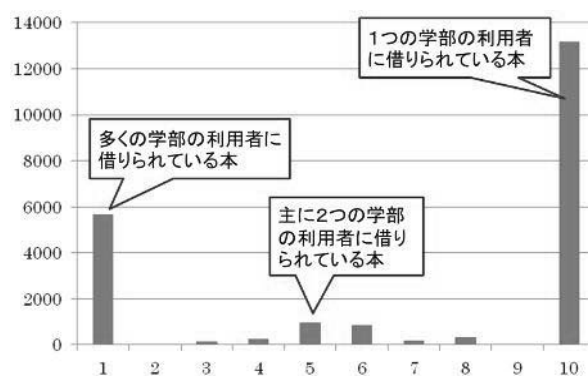


図2 図書専門度 ( $a$ -値) の分布

図書の  $a$ -値の頻度分布を図2に示す。1つの学部の利用者のみに借りられている ( $a$ -値が10) 図書は13,000件ほどあり、これが最大頻度である。一方多くの学部のメンバに借りられている  $a$ -値が1である図書も6,000件近くある。これは6学部以上の学部で同数借りられた場合の偏りに相当する。

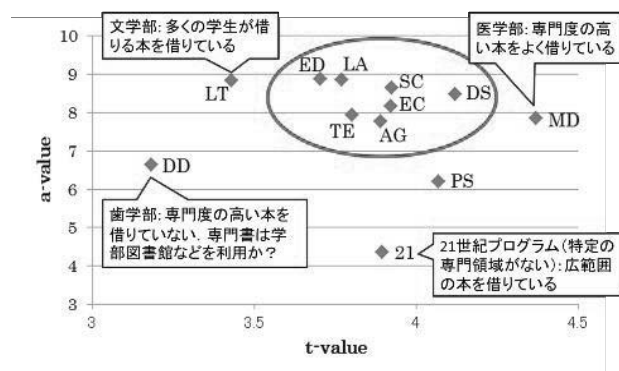


図3  $t$ -値と  $a$ -値による学部の位置づけ

$t$ -値と  $a$ -値は図書の専門度に対する2つの指標として考案されたものであるが、両者の相関関係はほぼ0であり、独立した指標と考えることができる。それぞ

れの定義を利用者の集合に拡張し、学部に対するそれぞれの専門度を調べた結果を図3に示す。多くの学部はt-値が3.6~4.2、a-値が7~9の中央上部に集まっている。それから外れた学部が5つある。

MD（医学部）はa-値としては中庸であるがt-値としては学部の中で最大値となっている。すなわち、医学部の学生が借りている図書は医学部特有という訳ではないが、研究者レベルの高度な図書が多いという傾向を示している。

21世紀プログラムは、いわゆる理系・文系などの既存の枠組みにとらわれず、広い範囲で学ぶことを想定した学部である。a-値が最小であることに、その傾向が表れている。t-値に関しても平均的な値となっているものの1~2年生に多く読まれる図書よりは専門性の高めの図書を読んでいることが分かる。

DD（歯学部）は、両方の値が低めになっている。特にt-値に関しては学部中最低値となっており、比較的低学年の学生に読まれる図書を多く借りている傾向がある。これは、歯学部の学生は専門図書は学部内の図書館を利用し、中央図書館では一般的な図書を借りていることであろうと推測できる。

LT（文学部）は、a-値に関しては平均的であるが、t-値に関してはDDの次に値が低い。これは文学部の学生が多く読んでいると考えられる文学作品などは比較的低学年の学生が多く読み、研究者レベルの利用者の場合は、文学研究などの例外的な場合を除いて、広くは読まれないことに起因しているものと推測できる。

### 2.3. 利用者の興味分野やその広がり分析

図書館の貸出記録データを基に、利用者の興味分野に関する情報を得たい[18-19]。そのために、利用者が借りている図書のNDC（日本十進分類法）番号を分野として利用する。まず、ある利用者が借りた図書のNDC番号（本稿では上位の10分野）の総数を要素とする000の総記から900の文学までの10次元ベクトルを、その利用者のプロファイルデータと定義する。利用者の集合に定義を拡大し学部のプロファイルも同様に定義する。図4に（通常の学部）に学部以外の利用者全体Oを1つの学部のようにみなし13の学部のプロファイルを100%積み上げ棒グラフとして示す。

容易に分かるように、PS（薬学部）やSC（理学部）は分類番号400の自然科学分野の貸出が高い割合を占める。一方、DS（芸術工学部）、LT（文学部）、O（その他）は広範囲の分野から貸出されている。我々の直感と異なり興味深いのは、MD（医学部）は比較的范围の図書を借りているところである。

このような直感的な観察結果を客観的に確認するために興味の広さに関する指標を導入する。本稿では、

情報エントロピーを採用する。これは分類番号に対して、全体におけるその割合をpとするととき、 $-p \log p$ の総和を取ることで計算される。最大値を1とするために対数関数の底を10とする。また、10次元ベクトルの要素の総和、すなわち、図書貸出の総数を興味の強さの指標とすることにする。

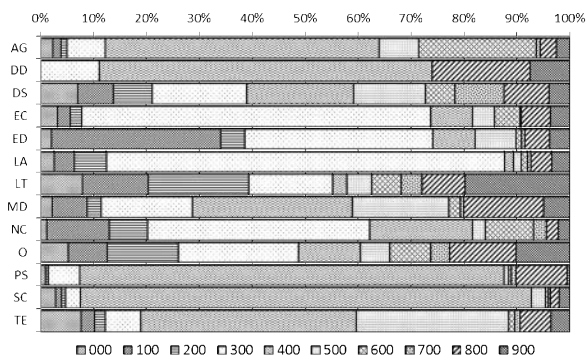


図4 学部のプロファイル比較

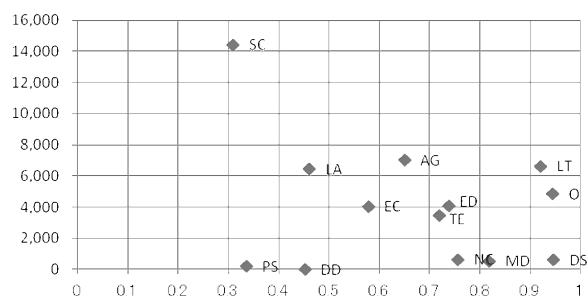


図5 興味の広さ（x軸）と強さ（y軸）による学部の比較

以上の定義に基づき、各学部の興味の広さと強さを図5に示す。興味の広さに関してSC（理学部）が最低値を、PS（薬学部）がその次に低い値となっており、これらの学部の興味が特定の分野に集中しているという図4からの我々の観察と一致する。同様に、DS（芸術工学部）、O（その他）、そしてLT（文学部）の3学部の広さの値が最も大きいことも我々の観察結果と同じである。

一方、興味の広さを数値化したことにより、DD（歯学部）やLA（法学部）がSCやPSに引き続き興味が集中しており、またDDとLAの集中の度合いが同程度であることやMD（医学部）が4番手に広い興味分野を持っていること、NC（21世紀プログラム）やED（教育学部）、TE（工学部）が同程度の広さの興味分野を持っていることなども図から読み取れる。

利用者と学部に関する興味分野プロファイルを利用して異なる情報を抽出する道具としてプロファイルとプロファイルの類似度を利用する。本稿ではベクトル間の類似度として良く用いられているコサイン類似度

を採用する. すなわちプロファイルベクトル A と B の類似度を  $(A \cdot B) / |A| |B|$  とする. ここで, 分子に現れる  $\cdot$  は内積を,  $|\cdot|$  はベクトルの長さ, すなわち自身との内積の平方根である.

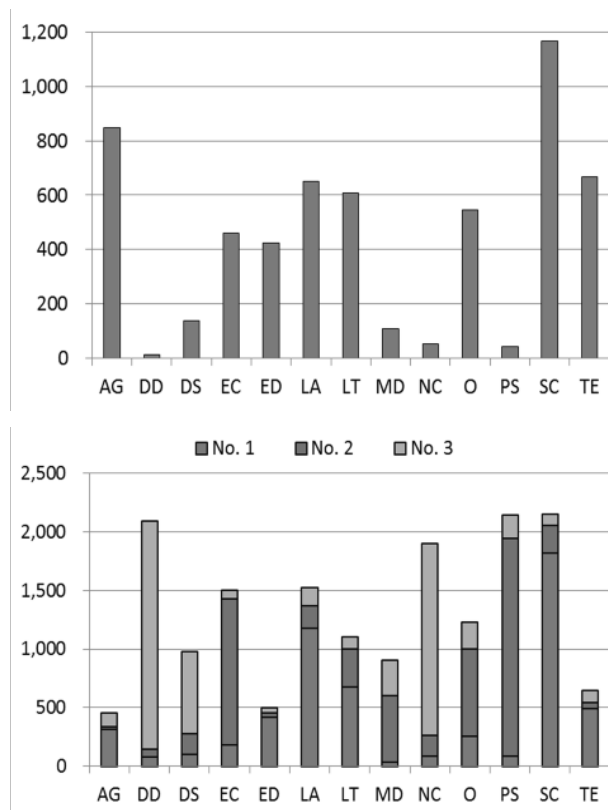


図6 実所属の人数(上)と仮想所属の人数(下)

類似度を利用して, 学生の仮想学部なる概念を定義する. ある学生のプロファイルと各学部のプロファイルの類似度を計算し, 類似度の大きい順に, その学生の第1仮想学部, 第2仮想学部, ... と定める. 第1仮想学部のことを単に仮想学部とも呼ぶ. 図6(上)にその学部の実所属している学生数を, そして図6(下)に, それぞれの学部に対して, それを第1から第3までの仮想学部にしていく(仮想所属の)学生数を示す.

図6より, たとえば AG(農学部)の実所属学生数は約800名であるが, (第1)仮想所属の学生数は400名程度と半分になっている. 同様に EC(経済学部)や TE(工学部)も実所属学生数に対して仮想所属学生数が減少している. 一方, SC(理学部)は実所属学生数が1,200名弱であるのに対して, 仮想所属学生数は1,800名程度と増加している. LA(法学部)も同様に増加している.

図7にそれぞれの学部に対して, その(実)所属学生が仮想所属学部に応じた割合で所属しているかを示す. 基本的には, 実所属と同じ学部を仮想所属とする学生の割合が多いものの, SC(理学部)のように, その割合が小さい例も存在する.

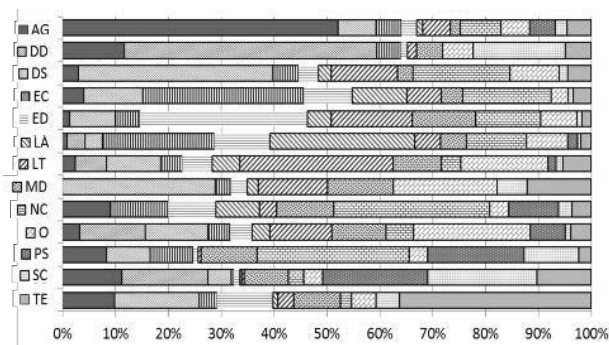


図7 学部別の仮想学部人数割合

MD(医学部)の場合は, 医学部を仮想所属学部にする学生の2倍近くが DS(芸術工学部)を仮想学部としている. すなわち MDの学生の興味分野の多様性は大きく, 医学部学生そして典型的な興味分野でない興味分野を持つ学生が多く存在し, しかも DSの典型的興味分野に近い学生が最も多い.

MDほど極端ではないものの PS(薬学部)も同様な傾向にある. 薬学部の学生の中で典型的な薬学部的興味分野パターンを持つ学生以上に NC(21世紀プログラム)の典型的興味分野パターンを持つ学生が多い. 学部の性格から予想されるほど強くはないものの NC所属の学生は PSの学生と比較して興味の範囲が広い場合が多い. このような学生が NCの典型的興味範囲と類似していることが予想できる.

### 3. 授業データ解析からの知見

学生の学びの場の第1は教室であり, 授業を通しての学びは学生にとっての学習の大原則である. 授業という場における学生の学びへの姿勢や習慣は, 学生にとって, 将来の職場における仕事への姿勢や仕事力に直結するものであり, 仕事力養成の一環として真剣に取り組むべきである. 本節ではこのような認識をベースに, 授業に関する様々なデータ解析から, 学生の学びへの姿勢や積極性, そして好奇心などを読み取る試み[16-17, 20]を紹介する.

#### 3.1. 授業に対する努力・成果分析

本稿では, ある女子短期大学の2年生向けの「情報検索演習」における, 出席や宿題提出などの平常点, 期末考査における試験点, そして期末に実施された学生アンケートへの回答を解析の対象とする. 本授業は, 女子短大における司書資格取得のための必修科目であり, 学生の受講態度はかなり良好であると考えられる.

本稿における我々の興味は, 成績そのものだけではなく, 成績や学習成果に対する学生の意欲や学びへの姿勢などの分析にある. そのため, 出席や宿題に関す

る評価点を学生の努力に対する評価（努力点）、試験点を学習成果に対する評価（成果点）と捉える。

授業データ解析の第1段階として努力に対する成果の相関関係を調べる。努力に関する指標としては出席を評価した出席点と宿題の提出状況および宿題の質を評価した宿題点がある。これら2種類の努力点の間の相関係数は0.53であり、ある程度の正の相関が見られる。出席点と試験点、宿題点と試験点の相関を調べると、両者には基本的に大きな差異がない。図8に試験点と宿題点の関連を示す。相関係数は0.0であり、両者には全体的な相関がない。

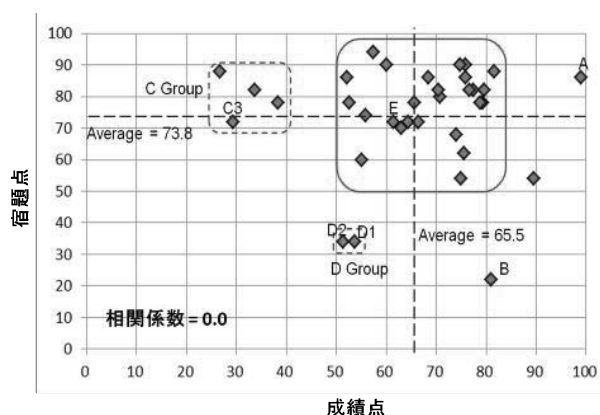


図8 成績点と宿題点の相関

両者の関係をさらに詳しく検討する。多くの学生はそれぞれの平均値の周辺部分（具体的には、成績については50~85、宿題については50~100の領域）に位置している。このクラスから外れた学生が10名程度存在する。

図でAと印つけられた学生は成績トップの学生であり、成績点は100点近くある一方、宿題点は90点に達していない。学生Aより高い宿題点を獲得した学生が何人もいることは、宿題という努力点が直ちに良い成績に繋がるとは限らないことを示している。

この傾向は、出席点に関してはより顕著である。単に多く出席しているだけでは良い成績には繋がらない。この結果は、出席や宿題などの努力を実力結びつける意識や実践が必要であることを示している。

同様の結果は学生4名からなるC群からも窺える。C群の学生達は、出席や宿題などの努力点に関しては概ね平均以上である。それにも関わらず、成績に関しては、それ以外の学生達から離れた最低点グループを構成している。これらの学生達は、一見努力しているように見えながら、それが成績向上に結び付くような真の学習になっていないことを示唆している。

学生2名からなるD群は、最下位ではないものの、試験点と宿題点の両方において低い評価になっている。

これら学生達は出席点に関しては、大きなクラスタに属しており、出席に関しては人並みに努力しているものの、宿題というより意識的な努力に関しては最低レベルの努力しかやっていない。C群と比較すると、出席に関しては同程度であり、宿題という努力点に関しては低いものの、試験点に関しては高い成果を出している。

これらの観察結果を総合すると、C群が形だけの努力をする学生群だったのに対して、D群の学生達は宿題に関する努力が低いために十分な成果が上げられなかった学生群と見ることができる。すなわち、D群の学生達は、宿題に関してもっと努力することにより、成果としての成績点がもっと高くなることが期待できる。自分たちのもっている潜在能力を十分発揮しなかった点で、大変もったいない学生達であると言える。

学生Bは宿題点のみならず出席点でも最低点をとっている学生である。それにも関わらず試験点は、最上位クラスとはいえないものの普通の学生達の上位に位置している。この学生は、潜在的には十分な能力を持っている学生であろう。十分な努力なしでもこのような高い成績が取れるのであれば、さらに努力することで学生Aと同様の一層高い成績点を獲得できるものと予想できる。この学生も自分の潜在能力を十分発揮できていない点で、もったいない学生である。

### 3.2. 自己評価と努力点の相関に関する分析

本稿で解析している授業では、最終回に授業全体を振り返り、授業と学生自身を両方を評価するアンケートを記名式で実施した。記名式とした理由は、どういった学生がどのような評価を行ったかを分析できるためと同時に、学生自身が責任感をもって自身の評価を行って欲しいためである。また、我々の過去の経験から、記名式であっても学生達は意外と正直に自己の問題点などを回答するものであると判断しているためである。

質問項目は全部で12項目あり、そのうち5項目は授業の評価に関する問いであり、6項目は学生自身の評価に関するもの、そして残り1項目は、特に内容を指定せずにコメントを求める問いである。授業と学生自身に関する質問項目の中にそれぞれ総合評価として100点満点の何点をつけるかの問いが含まれている。本節では、学生自身の評価点が前節でも用いた宿題点や試験点とどのような相関関係にあるかを分析する。

図9に学生の自己評価点と宿題点の相関を示す。この相関係数0.5は自己評価点と出席点の間の相関係数0.4よりも大きい。これは同じ努力点の中でも、単に授業に出席するよりも、より大きな努力が必要な宿題に熱心に取り組んだことの方を学生はより高く評価していることを示している。

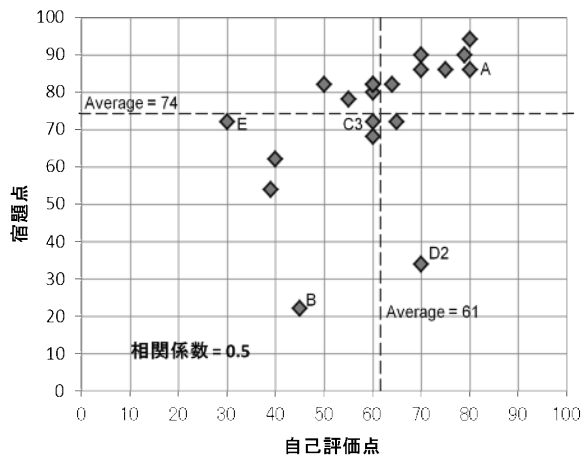


図9 自己評価点と宿題点の相関

図4によると、学生 D2 と B を除いて宿題点と自己評価点の間には大きな相関があることが分かる。これら2名の学生は宿題点が最下位である。恐らくこれらの学生達は、自分の努力が不十分であることを認識した上で、あえて高めに自己評価しているのであろう。

特に学生 D2 の自己評価は平均以上である。既に述べたように D2 (D 群) の学生は平均かそれ以上に真面目に出席しているものの、宿題の提出に関しては熱心でない。学生 D2 は良く出席したことでもって自分自身を高く評価している可能性がある。成績的にも下位であることを考慮すると、このタイプの学生には実質的な学習を行うことの重要性を十分理解させる必要がある。

### 3.3. 自己評価と試験点の相関に関する分析

図10に自己評価と成績点との相関を示す。驚くべきことに相関係数は-0.1と負値である。自己評価と努力点との相関が比較的大きな正值であったことを勘案すると、学生達は自分の成果というよりも努力に対して評価している傾向が強いことになる。

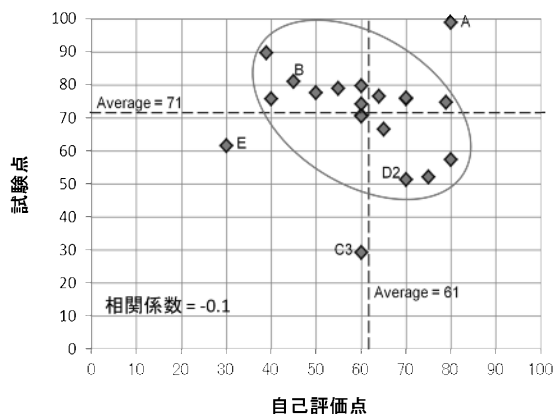


図10 自己評価点と試験点の相関

これは努力はある程度自己認識できるし、客観化する意識があれば、かなり冷静に評価できる一方、成果に関しては、自分で客観的に評価するのが難しいという事情もあるものと考えられる。

図10をさらに詳細に分析する。学生 A は努力点に関しては最上位ではないものの上位グループに属しており、また、成績点に関してはトップであるため、自己評価に関してもトップグループに属していることは納得できる。自負心をもって授業や学習に熱心に取り組んでおり、それを自分でも評価している学生である。

学生 C3 は試験点が最低値であるにも関わらず、自己評価点は平均値程度に高い。すでに分析したように、C 群の学生は出席や宿題提出など外見的な面に関しては平均かそれ以上に努力している。努力の結果が成績という成果につながらなくても努力しているように見えること自体を自己評価しているのであろう。このような学生にも実質的な成果を上げることの重要性を学んでもらう必要がある。

学生 E は自己評価が最低の学生である。この学生の努力点や成果点は平均値にわずかに及ばないレベルであり、特段に低い訳ではない。E は他の学生より厳しく自己評価している。このような学生は向上心が高いものと考えられる。ある程度の努力をし、ある程度の成果を出しているにも関わらず、自分以上の学生の存在を考え、そのようなレベルに至るためにもっと努力をすべきであったと、自分の行動を反省しているものであろう。向上心が高いことは利点であるものの、それが行き過ぎて自己嫌悪など自分自身をネガティブに捉える傾向が強すぎるのは問題である。向上心と同時に自分自身の努力や学びへの姿勢を評価することができるよう導くことが望まれる。

## 4. 学生の社会人力育成へ向けたデータ解析

前節まで学生 SD を目的に学生の現状を図書館や授業データに基づいてモデル化するいくつかの試みを紹介した。本節では、このようなアプローチによるデータ解析によって得られる知見が学生の社会人力（職業人としての能力）育成にどのように役立てられる可能性があるのかを議論する。高等教育は、初等中等教育を卒業した学生を受け入れ社会人として仕事のできる人材として社会に送り出す役割を担っている。社会人力とは具体的にはどのような能力であろうか？

社会人として必要かつ重要な能力に関しては、企業の人事担当者が学生を採用する際に、どこを評価しているかが参考になる。企業が就職活動中の学生を評価する基準は端的に表現すると、「仕事仲間として相応しいか」、あるいは、「自社に貢献してくれそうか[7]」という。この表現自体はかなり曖昧であるものの、そ

の意味しているところは、学力や知識にも増して、学生の行動や心のありようが社会人として仕事を遂行していく上で重要であることを示唆している。

社会人力をさらに具体化すると、好奇心、向上意欲、実行力、論理性、視野の広さ、精神的耐久性や安定性、発想力などがその要因となる（[7]など）。すなわち、仕事や社会に対する好奇心を持って仕事に熱心に取り組み、自分を向上させるためにたゆまず自己研鑽に努め、やるべき事柄を直ちに実行に移すことができ、また、自分の考えを広い視野から論理的に説得性を持って説明でき、しかし一方で自分の考えに固執せず、他人の考えを十分の敬意を払って聞きくことができ（コミュニケーション力が高い）、チームの一員として自分の役割を果たすだけでなく、チーム全体の利益のためにそれ以上の仕事を積極的にできるような人間を社会人力が高いと考えることができよう。これは通商産業省が提唱している社会人基礎力[8]にも通じる。

社会人力をこのように捉えるとき、我々の図書館・授業データ解析によって、学生の持つこのような資質を推定し、把握（モデル化）することはいかに可能であろうか？以下、資質のいくつかを議論する。

**好奇心：** 学生にとっての仕事は勉学であると。第2節で紹介した興味分野の分析結果は、学習に対する好奇心の指標という側面も持つ。興味分野の広がりだけではなく、自分の専門分野を中心に、その関連分野への興味のあり方も分析することにより、将来の職業における専門知識やその周辺知識に対する好奇心に関する予測情報が得られるものと期待できる。

**向上意欲：** 我々は図書館貸出データからの専門度推定に関する研究の中で、学生の実専門度と貸出図書の専門度の差を、その学生の向上意欲の尺度とした[15]。この研究をさらに精密化することにより、職業人としての向上意欲も推定できる可能性がある。

**論理性：** 客観的評価は難しいものの、授業字の作文や宿題へのレポートなどの文章を分析することにより、論理の組み立て力に関して相当の情報が得られる。与えられたテキスト（文章）から、その論理性を自動評価する技術を適用することにより、本項目に関する評価の手間をかなり軽減することができる。

**注意力：** 仕事をしていく上で、顧客からの要請や上司からの指示内容を正確に把握する能力は仕事力の基本である。近年、宿題の提出方法に厳密に従うことのできない学生が増加している。授業での指示などを正確に理解する能力は仕事上の注意力の指標となる。

**スケジューリング力：** 仕事を締切内で終わらせることは社会人の基本である。学生の場合は、授業に遅刻しないことや宿題を締切日時までに提出することなどはスケジューリング力や計画性の良い指標である。

このように見ていくと、授業データからは、与えられたテーマに対する学生のアプローチなど、たとえば、構想力やテーマに関する知識レベルなど、いわばミクロの情報が得られる。いっぽう、図書館データからは、学生の興味分野やその広がり、そして様々な分野に関する大雑把な知識レベルといった、いわばマクロの情報が得られる。ミクロとマクロの情報を併せてモデル化することにより、学生に対するより包括的なイメージを捉えることができる。今後、このアプローチによる一層精密な解析に関する研究が重要である。

## 5. 結語

教育機関である大学にとって教育機能を最大化することは極めて重要である。そのためには教育機能の中核をなす授業を中心に現状を分析・把握し、それを踏まえた対策を講じる必要がある。従来教育効果を向上させるために教員のためのFD (Faculty Development)や職員のためのSD (Staff Development)活動が盛んに行われてきた。しかし、第1の当事者である学生自身が学ぶ意義を十分自覚し、授業を中心とする教育の機会を最大限に生かすよう自己努力するように持っていくためのSD(Student Development)活動の重要性が十分理解され努力されて来ているとは言えないのが現状である。

本稿では、この学生SDを活発化させることを目的に、その第1段階として学生の活動に関する様々なデータを収集し、それを解析することにより学生の現状、特に学びに対する学生の関心の広さや強さ、そして学びへの姿勢などを分析する試みを紹介した。

第2節では図書館データ解析を、第3節では授業データ解析を取り上げ、データ解析によりどのような知見が得られるのかを具体的に示した。その過程を通して、有益な知見を得るためには、通常良く行われる統計やデータマイニングの標準的手法の適用だけではなく、新しい指標を導入するなどデータの特性に応じた工夫が必要であることも併せて示された。

第4節では、大学教育の大きな目的が、学生を社会人、あるいは職業人として育成し、卒業後仕事の現場で十分に活躍できる素地を身に付けてもらうことに注目し、図書館や授業のデータ解析が職業人としての学生の育成にどのように役立つ可能性があるかを議論した。

本稿と同様に、教育への応用を意図したデータ解析に関しては、従来、KDD (Knowledge Discovery and Data Mining) [1]やEDM (Educational Data Mining) [3, 23, 24]などの分野で研究されてきた。論文[24]ではe-Learningシステムから得られるデータを用いて学生を分類する問題に対するデータマイニングアルゴリズムを比較している。その目的は学生の成績を事前に予測すること

にある ([27]も同様)．学生の学習成果と見なせる最終成績を予測することで、問題のありそうな学生を早期に発見し、対策を講じることができる．

本稿の研究目的も概ね同様であるが、我々の主要な興味は、得られたデータから直接的に成績を予測することではなく、それ以前に学生のやる気や勤勉さなどの心の姿勢や性格などに関するモデル化を行い、それに基づき間接的に成績を予測し、問題のあることが予測できる学生に対して根源的にケアすることを目指している点が異なる．

我々の研究はコンピュータ支援による協調学習 (Computer Supported Collaborative Learning, CSCL) の研究領域とも関連する．論文[26]では、学生間の交流 (Interaction) を構造化することを目的に協調して学ぶための場を扱っている．本テーマに関する我々のアプローチは、まず学生達の学習行動に関する特徴を見出し、その後、得られた知見をベースに学生達の興味や習得知識のレベルのバランスなどを考慮した上で最適な協同学習グループの構築を提案するというものである．

従来の研究と比較した我々のデータ解析へのアプローチは次のような点で相違している．

#### (1) データ解析の目的

すでに述べたように我々のデータ解析は、従来研究で多く行われている授業データに対する直接的なデータマイニングではなく、学びに対する学生の心の姿勢をモデル化することを大きな目的としている．

#### (2) 対象データのサイズ

従来のデータ解析・データマイニングの研究はビッグデータ (Big Data) と呼ばれる大規模なデータを対象としていることが多い．近年は Web システムの発展および普及により、アクセスログなどのデータが大量に得られるようになったことを背景にビッグデータからのマイニングが大きな注目を集めている．

それに対して我々の対象データでは授業データのように、それらと比べると極めて少量である．そのため、少数のデータの値の揺らぎ (ノイズ) が全体の傾向に与える影響が大きくなりがちである．ノイズの影響を少なくするためには、データ以外の背景知識や補完データなどを総合的に駆使して、妥当な解釈を探索するアプローチが有効である．このような考えに基づき、スモールデータであっても有用な知識獲得ができる手法の開発を我々は目指している．

#### (3) 対象データの入手方法

前項とも関連するが、大量のデータを獲得するためには、何らかの自動的手段を導入する必要がある．実際、従来研究の多くは e-Learning システムのログデータを用いるなどにより大量のデータを得て解析するビッグデータ指向である．一方我々が対象とするデータ

は特別なコンピュータシステムなどを仮定しない一般の授業データなどのスモールデータである．したがって、手作業により必要なデータを作成することも可能である．

#### (4) 例外ケース (外れ値) の取り扱い

データマイニングや統計手法を含め従来のデータ解析においては、通常、例外的なデータを除外した全体の特徴を掴むことが大きな目的である．それらから外れたデータはしばしばノイズデータとして扱われる．

一方学生に関するデータにおいては、標準的なクラスから離れた外れ値データであろうとも、それはそれで人格を持った 1 人の学生を表現している．そうである限り、それらのデータも他のデータと同様に尊重されるべきである．そのような外れたデータをも考慮に入れ、それを如何にして教育に活かすことができるかを考えることが我々の研究にとっては重要である．

#### (5) 解析ツールの開発

データ解析に関する従来の研究の多くは既存のツールを様々なデータに適用したものである．たとえばデータマイニングの研究において統計的な手法と同時に相關規則 (Association Rule) を発見する研究は多い [2]．

本稿における授業データ解析の研究は、データやアプローチに独自性があり、従来の解析手法をそのまま適用する訳にはいかない．したがって、我々は設定した解析課題に適合するような解析手法を考案し、それを改善しながらデータ解析を進めて行くスタイルでの研究を行う．これも本稿の研究において重要性の高いテーマである．

なお、我々と同様にスモールデータを対象とした授業データ解析の研究も少数ながら存在する．合田等は毎回の授業で学生に自己評価コメントを提出させ、それを手動でデータ化し、コメント内容から成績を推定する研究結果を報告している [4-6]．

今後の研究方向として次のような課題がある．

(1) 更に新アイデアを考案し、学生の勉学への熱心さ、努力、学びへの態度などをさらに精密化して把握できる手法を開発すること、中でも、学生アンケートの中のテキスト部分からは、多くの情報を抽出できる可能性があり、今後深く研究すべきテーマである．

(2) 本稿とは異なるタイプの授業・図書館・その他のデータを収集し、それらを解析することにより、本稿の研究で示唆された結果を検証したり、比較したりすることが課題である．たとえば、就職活動に関連した既存の研究結果に対して、本稿の手法を用いた解析結果を用いて補強したり補完したりすることにより、有益な知見が得られる可能性が大きい．

(3) 解析手法を一般化、自動化し、学びに関する学生プロフィール解析システムとして統合化する．

## 謝辞

九州情報大学の大浦洋子教授には、共同研究者としてデータ解析のための統計手法などに関する多くの知識やアイデアをいただきました。深く感謝いたします。また、本研究の一部は科学研究費補助金（基盤研究(C), 24500318）の助成を受けて実施されました。

## 参考文献

- [1] ACM SIGKDD: ACM Special Interest Group on Knowledge Discovery and Data Mining. <http://www.sigkdd.org/>
- [2] R. Agrawal, T. Imielinski, and S. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. 1993 ACM SIGMOD International Conference on Management of Data, pp.207-216, 1993.
- [3] Educational Data Mining Society, <http://www.educationaldatamining.org/>
- [4] K. Goda and T. Mine: PCN, "Qualifying Learning Activity for Assessment Based on Time-Series Comments," Proc. 3rd International Conference on Computer Supported Education (CSEDU 2011), 6pp., 2011.
- [5] K. Goda and T. Mine, "Analysis of Students' Learning Activities through Quantifying Time-Series Comments," Proc. 15th Annual KES Conference (KES'2011), Part II, Lecture Note in Artificial Intelligence (LNAI 6882), pp.154-164, 2011.
- [6] 合田 和正, 峯 恒憲, "PCN 法による自己評価コメントの分析からの改善可能な学生の発見", 教育システム情報学会(JSiSE)特集論文研究会研究報告, Vol.26, No.7, pp.51-58, 2012.
- [7] 井村 直恵, "効果的な就職活動支援," 九州国際大学経営経済論集, 第 17 巻第 2 号, pp.37-43, 2011. <http://www.kiu.ac.jp/organization/library/memoir/img/pdf/keizai17-2-006imura.pdf>
- [8] 経済産業省, "社会人基礎力," <http://www.meti.go.jp/policy/kisoryoku/>
- [9] T. Minami and E. Kim, "Seat Usage Data Analysis and its Application for Library Marketing," Third International Conference on Intelligent Information and Database Systems (ACIIDS 2011), LNAI 6591, pp.238-247, 2011.
- [10] T. Minami, "Potentials of Circulation Data Analysis for Library Marketing --- A Case Study in a University Library ---," The 2011 International Conference on Database Theory and Application (DTA 2011), Springer CCIS 258, pp.90-99, 2011.
- [11] T. Minami, "Book Profiling from Circulation Records for Library Marketing --Beginning from Manual Analysis toward Systematization--," International Conference on Applied and Theoretical Information Systems Research (ATISR 2012), 2012.
- [12] T. Minami, "Expertise Level Estimation of Library Books by Patron-Book Heterogeneous Information Network Analysis -- Concept and Applications to Library's Learning Assistant Service --," The 8th International Symposium on Frontiers of Information Systems and Network Applications (FINA 2012), DOI 19.1109/WAINA.2012.184, pp.357-362, 2012.
- [13] 南 俊朗, 大浦 洋子, "学生の成長を助ける学習支援への模索 --授業データ解析による支援方法発見への試み--," 九州情報大学研究論集, 第 14 巻, pp.39-50, 2012.
- [14] T. Minami and Y. Ohura, "Toward Learning Support for Decision Making --- Utilization of Library and Lecture Data ---," The 4th KES International Conference on Intelligent Decision Technologies (KES-IDT' 2012), Springer Smart Innovation, Systems and Technologies 16, pp.137-147, 2012.
- [15] T. Minami and K. Baba, "Investigation of Interest Range and Earnestness of Library Patrons from Circulation Records," Proc. International Conference on e-Services and Knowledge Management in IIAI-AAI 2012, IEEE CPS, pp.25-29, 2012.
- [16] T. Minami and Y. Ohura, "An Attempt on Effort-Achievement Analysis of Lecture Data for Effective Teaching," Database Theory and Application (DTA 2012), T.-h. Kim et al. (Eds.): EL/DTA/UNESST 2012, CCIS 352, Springer, pp.50-57, 2012.
- [17] T. Minami and Y. Ohura, "Towards Development of Lecture Data Analysis Method and its Application to Improvement of Teaching," 2nd International Conference on Applied and Theoretical Information Systems Research (2nd ATISR 2012), 2012.
- [18] T. Minami, "Profiling of Patrons' Interest Areas from Library's Circulation Records--An Approach to Knowledge Management for University Students --," The Fifth International Conference on Information, Process, and Knowledge Management (eKNOW 2013), 6pp., 2013.
- [19] T. Minami, "Interest Area Analysis of Person and Group Using Library's Circulation Records," IADIS International Conference Information Systems 2013 (IS 2013), 8pp, 2013.
- [20] 南俊朗, 大浦洋子, "授業データ解析による授業改善策発見を目指してー努力・成果・評価の関連性からのアプローチー", 九州情報大学研究論集, 第 15 巻, pp.1-16, 2013.
- [21] T. Minami, "Changes of Interest Range of Students with Circulation Record Analysis," Information conference 2013 (Information'2013), International Information Institute, 4pp., 2013.
- [22] T. Minami and Y. Ohura, "Lecture Data Analysis towards to Know How the Students' Attitudes Affect to their Evaluations," 8th International Conference on Information Technology and Applications (ICITA 2013), 6pp., 2013.
- [23] C. Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005," Expert Systems with Applications 33, pp.135-146, 2007.
- [24] C. Romero, S. Ventura, P.G. Espejo, and C. Hervás, "Data Mining Algorithms to Classify Students," Proc. 1st International Conference on Educational Data Mining (EDM 2008), pp.8-17, 2008.
- [25] 志垣一郎, "学生による授業アンケートと成績の関係," 大阪工業大学紀要理工篇, 第 55 巻第 1 号, pp.1-9, 2010.
- [26] L. Talavera and E. Gaudioso, "Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces," Proc. Workshop on Artificial Intelligence in CSCL, 16th European Conference on Artificial Intelligence (ECAI 2004), pp.17-23, 2004.
- [27] N. Thai-Nghe, L. Drumond, T. Horvath, and L. Schmidt-Thieme, "Multi-Relational Factorization Models for Predicting Student Performance," in KDD 2011 Workshop: Knowledge Discovery in Educational Data, as part of 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2011.
- [28] 吉川政夫, 有沢孝治, 川野辺裕幸, 内田晴久, "構造化された授業評価アンケートの開発," 広島大学高等教育研究開発センター大学論集, 第 43 集 (2011 年度), pp.337-351, 2012.