

小林学，大谷真，梅澤克之，後藤正幸，平澤茂一，「出席状況把握システム SAMS とその解析」，『経営情報学会 全国研究発表大会要旨集』，2014 年春季研究発表大会，21-24 ページ，2014 年．

出席状況把握システム SAMS とその解析

SIT Attendance Management System and its Analysis

小林 学¹⁾ 大谷 真¹⁾ 梅澤 克之²⁾ 後藤 正幸³⁾ 平澤 茂一³⁾

Manabu Kobayashi¹⁾ Makoto Oya¹⁾ Katsuyuki Umezawa²⁾ Masayuki Goto³⁾ Shigeichi Hirasawa³⁾

1) 湘南工科大学 工学部

2) 日立製作所

3) 早稲田大学 理工学術院

1) Shonan Institute of Technology

2) Hitachi,Ltd.

3) Waseda University

要旨： 湘南工科大学では学生の授業出席データを収集・管理するシステム（出席状況把握システム;SAMS）を開発し、2013年度から導入した。このシステムはICカード化された学生証で出席を取る端末と、出席状況を登録・管理するWEBシステムからなる。SAMSを用いる目的は、欠席による単位修得が不十分な学生を早期に推定することにより、留年や退学予防、ひいては大学教育改革の推進に役立てることである。本稿ではこのSAMSの概要を述べ、実際に得られた出席情報と単位修得情報を用いて、単位修得に関する推定を行う手法を検討し、評価を行う。

Abstract: We developed and introduced the attendance management system (SAMS) as an academic affairs system of Shonan Institute of Technology. This system consists of IC card readers for taking attendances and the WEB system for managing attendance information. The purpose of SAMS is estimation of students that cannot get enough course credits as soon as possible. We can use this estimation to prevent repeating a course and withdrawing. First, we give an outline of SAMS in this manuscript. Using attendance information, we propose a method to estimate students that cannot get enough course credits.

1 はじめに

湘南工科大学では学生の授業出席データを収集・管理するシステム（湘南工科大学出席状況把握システム;SAMS）を開発し、2013年度から導入した。このシステムはICカード化された学生証で出席を取る端末と、出席状況を登録・管理するWEBシステムからなる。SAMSを用いる目的は、欠席しがちな学生を早期に発見することにより、留年や退学予防、ひいては大学教育改革の推進に役立てることである。後閑らは半期分のSAMSの出欠情報と単位修得情報、講義情報、学生情報を手作業にて解析し、いくつかの有用な傾向を示した[1]。本稿ではまずSAMSの概要について述べ、授業の出欠状況の登録方法について簡単に述べる。次に実際に得られた1年分の出欠情報と単位修得情報に対して自動分類手法を適用する。これにより単位が十分取得できない学生の推定を行う手法を検討し、SAMS及び推定方法の有効性を検証する。

2 SAMS

本節ではSAMSの概要について説明を行う。SAMSを導入するに当たり、☒学生証のICカード化、☒学生証の読み取り方法、並びに☒WEBシステムへの登録方法の3点について主に検討を行った。それぞれについて以下で概説する。

2.1 学生証のICカード化

授業の出席状況を把握することを目的として、湘南工科大学では2012年度の入学生から学生

証のICカード化を行った。このとき今後の汎用性やセキュリティ面なども考慮し、FCF（FeliCa Common-use Format）規格を採用した。FCF規格はFeliCaによるセキュリティの確保及び読み取り装置が普及しているという大きなメリットがある他、様々な大学で導入実績があり、ICデータの共通領域に格納された情報は大学間で連携して利用することが可能である。

本学の学生証のICに登録してある情報は主に1) 学籍番号、2) 氏名仮名、3) 氏名漢字、4) 利用区分、5) 再発行回数、6) 発行年月日、7) 有効期限、8) 学校識別コード、である。

これらの情報は汎用の非接触ICカードリーダーによって簡単に読み取ることができるため、PC端末のパスワードの代替や、部屋の電子キーなど、出欠以外の目的に利用することも可能である。またソニー社のPaSoRiなどを用いると、独自にICカードリーダーのプログラミングを行うことも可能となり、学生証を用いたオリジナルソフトウェアを構築することも難しくない。

2.2 学生証読み取り端末

学生証を読み取る端末（以下SAMS端末と略す）にはCSE社のモバイル機器PITを用いている。この端末は非接触でFeliCaを読み取ることができ、またOSにWindows CEが搭載されている。さらにクレードルによって簡単に充電を行うことが可能である。今回この端末で利用する出席確認ソフトウェアの開発は(株)ISTソフトウェアに依頼を行った。

SAMS 端末は各学科事務及び教務課に複数台常備し、各教員が講義に持参する形式で利用する¹。

導入計画時に最も問題となったのは、SAMS 端末と後述する学務事務システムとの連携である。利便性及び信頼性向上のためには SAMS 端末に各講義情報を伝える必要があるが、講義情報は学期が始まってもしばらく変更される可能性がある。また全講義情報を前もって SAMS 端末に入れてしまうと、多数の教員名や講義名の中から担当講義を選択する手間が生じる。

以上の観点から、各教員が USB メモリを用意し、その中に担当教員のみの「講義情報ファイル」を入れておく方式を採用した。「講義情報ファイル」の各行には 1) 教員 ID, 2) 教員氏名, 3) 講義 ID, 4) 講義名, 5) 曜日, 6) 時限, が記載されている。

SAMS 端末から出席確認ソフトウェアを立ち上げると、まず講義情報ファイルの入った USB メモリを SAMS 端末に挿入するように指示が出される。このときソフトウェアでは USB メモリの中から講義情報ファイルを読み取り、もし教員 ID が一つだった場合には教員名の選択画面をスキップする²。続いて出席確認を行いたい講義名（と曜日・時限）を選択すると、IC 学生証による出席確認画面となる。出席確認画面では、学生が学生証を近づけたときに 2 秒ほど学籍番号を大きく表示する仕様となっている。

50 名ほどの授業に対して、出席確認にかかる時間は 3～5 分ほどと短い。受講生の多い講義では、複数台の SAMS 端末を利用している。著者は授業開始 5 分前から出席確認を開始し、授業開始とともにそれ以降の入室者を遅刻として扱っている。

なお出席確認後、出席確認ソフトウェアのメニューで「出席ファイル」の出力を選択すると、以下の項目が各行に記載されたファイルが USB メモリに出力される。

- 1) 学籍番号
- 2) 端末番号
- 3) 教員 ID
- 4) 講義 ID
- 5) 講義名
- 6) 学生証読み取り日時
- 7) 曜日
- 8) 学生証再発行回数

学生証再発行回数を記入している理由は、再発行によって学生証を複数枚所持した不正を抑止するためである。なお学生が複数回出席確認を行った場合、SAMS 端末では同一の学生かどうかのチェックを行わず、複数行に同一学生の情報が記載され

る仕様としている。また「出席ファイル」は「出席確認」を終了したときの時刻をファイル名に入れておくことにより、複数回の出席確認にも対応している。さらに、「出席ファイル」の USB メモリへの出力時に、規定のフォルダへコピーを作成することにより、万一の情報損失に備える仕様としている。「出席ファイル」の各行の先頭を学籍番号としている理由は後述する。

2.3 出席登録・確認 WEB システム

湘南工科大学では学生情報、教務情報、履修申告・確認、成績登録・確認等の学務システムに、新日鉄住金システムソリューションズ(株)が提供する学校事務システム「キャンパススクエア」を利用している。「キャンパススクエア」には出席確認・登録を行う機能はもともと存在しているが、WEB 上で手で一人ひとり出欠を登録する必要があった。そこで前節で説明した各講義の「出席ファイル」を一度に「キャンパススクエア」にて読み込めるようカスタマイズの依頼を行った。具体的に追加した主な機能は以下である。

- 1) SAMS 端末の出力する「出席ファイル」の読み込み
- 2) 出席情報登録時に「出席ファイル」中の教員 ID 及び講義 ID による登録チェック³
- 3) 出席情報登録時に「出席ファイル」中の学籍番号と、履修者情報とのチェック⁴
- 4) 「出席ファイル」の形式は、SAMS 端末からの出力の他に、各行に「学籍番号」のみからなるファイルの読み込みをサポート
- 5) SAMS 端末で利用する「講義情報ファイル」の出力機能

上の 4) の機能により、各教員が SAMS 端末を利用していない場合でも、かなり容易に出席情報を入力することが可能である。また 3) の機能により、履修申告に失敗している学生の抽出なども容易となった。

3 単位修得不十分学生の推定

湘南工科大学では 2013 年より、1 年次及び 2 年次の教科において SAMS による出欠登録を基本的に必須とした⁵。このため 1～2 年生の多くの科目において 1 年分の出欠情報が SAMS に登録

³登録しようとする講義と出席ファイルに齟齬がある場合には、画面に赤字で注意が出力される。ただし「強制登録」ボタンをクリックすることにより、登録を行うことも可能。

⁴履修していない学生がいた場合、画面に赤字で注意が出力される。

⁵2013 年度は 3, 4 年生も数多く受講する教科や共通教養科目については必須ではなかったが、2014 年度からは完全に全教科必須である。

¹各教員に 1 台ずつ配布はしていない。またコスト及びセキュリティ面を考慮し、教室に備え付ける形ともしない。

²複数名いた場合にはまず教員名を選択する。

されている。この情報を有効に活用することにより、単位未修得者を早期に推定することが可能となるかもしれない。本節では学習理論を用いることにより、出欠情報から単位を十分に取得できそうかどうかを自動的に判定する方法について検討を行う。

まず SAMS にどの程度の出欠情報が登録されているかを明らかにするために、1, 2 年生における各学生当たりの履修科目数並びに SAMS 登録科目数の平均と分散を表 1 に示す。

表 1: 学生一人当たりの半期の履修科目数及び SAMS 登録科目数

	履修科目数	SAMS への登録科目数
平均	11.44	7.65
分散	2.04	3.43

なお SAMS 登録科目数とは、各学生の履修科目の中で、出席情報が SAMS に登録されている科目の数を意味する。今回は授業の 1 ～ 5 回目の出欠情報すべてが登録されている授業を SAMS 登録科目としてカウントした⁶。

教職科目を除いて各学生が履修可能な科目単位数の上限は半期で 24 単位である。従って表 1 から履修上限近くまで履修していることが見て取れ、そのため分散がかなり小さい。一方 SAMS 登録科目数は平均で 7.65 であるから、約 67 % の科目で 1～5 回目の出席情報が登録されていることになる⁷。

さて単位修得不十分学生の推定を行うために、学習理論におけるいくつかの自動分類手法を用いる。このとき 1 つのデータは 1 学生の半期の SAMS 登録科目の出欠情報及び全履修科目の単位修得情報とする⁸。

ここである学生の i 週目の出席率を x_i と表し⁹、この学生の特徴ベクトルを $\mathbf{x} = (x_1, x_2, \dots, x_d)$ と表す。従って第 d 週までの出席率を予測に用いることを意味する。また目的変数となるクラスを y と書き、全履修科目の中で単位修得率が α 以上の場合には $y = 1$ とし、 α 未満の場合には $y = 0$ とする。ただし実験では $\alpha = 0.7$ と設定する¹⁰。す

⁶ここで 1～5 回目を利用した理由は、これらのデータを実験に利用するためである。

⁷なお 1～5 回目の授業の中で 1 回でも登録がなかった科目はカウントしていないため、実際にはもう少し多くの科目で登録が行われている。

⁸SAMS 登録科目のみの単位修得情報についても実験を行ったが、結果に大きな違いは見られなかった。

⁹すなわち「 $x_i = i$ 週目の出席科目数/SAMS 登録科目数」である。

¹⁰24 単位履修している学生の単位修得率が 0.7 だった場合、平均的な取得単位数は 16.8 であり、3 年間でぎりぎり留年しない 100 単位を取得することになる。従ってここでは

なわち、本実験では単位修得率が 0.7 未満の学生を単位修得不十分学生と定義する。

表 2 に $d = 3$ 、すなわち 1～3 週の出席率を用いた場合の自動分類法による分類結果を示す。ただし学習データ数は 1,000 とし、残りを評価用データとして用いることをランダムに 10 回行った結果である。表 2 において、TM はテンプレートマッチングを、 k -NN は k 近傍法を、Fisher はフィッシャー判別を、SVM はサポートベクトルマシンを、RF はランダムフォレストをそれぞれ表している。ただし TM、 k -NN の距離関数にはユークリッド距離を、SVM のカーネルにはガウスカーネルを、CART 及び RF の不純度には GINI 関数をそれぞれ用いた。なお各手法において誤分類率の最も低いパラメータの結果を示している。ここで自動分類法により推定したクラスを $\hat{y}$ で表記し、誤分類率を $P_E = P(\hat{y} \neq y)$ と表記する。また $P(\hat{y}|y)$ は真のクラスが y だったにもかかわらず $\hat{y}$ と推定した割合を表すものとし、 $P_{E0} = P(\hat{y} = 1|y = 0)$ 、 $P_{E1} = P(\hat{y} = 0|y = 1)$ と書く。

表 2: $d = 3$ としたときの誤分類率

	P_E	P_{E0}	P_{E1}
TM	0.198	0.436	0.099
k -NN	0.203	0.543	0.060
Fisher	0.198	0.435	0.099
線形 SVM	0.225	0.639	0.051
SVM	0.226	0.643	0.050
CART	0.208	0.439	0.112
RF	0.198	0.535	0.057

さて表 2 の結果を見ると、誤分類率は 0.2 前後と一見なかなか良い結果のように見えるが、 P_{E0} の誤分類率がかなり大きい。すなわち $y = 0$ のクラスにもかかわらず、 $\hat{y} = 1$ と推定してしまうことがとても多い。これは単位修得不十分学生の多くを見逃してしまうことを意味しており、望ましい結果ではない。この原因はクラスの偏りにあり、多くの学生のクラスは $y = 1$ であるため、自動分類器は学生を $\hat{y} = 1$ と推定しようとしてしまうためである。これは 0-1 損失が適切ではない好例である。

そこでクラスの偏りを考慮するために、学習データに重みづけを行う。すなわちクラス $y = 0$ のデータには $y = 1$ のデータと比べて w 倍の重みを与えることにより、 $y = 0$ のデータをなるべく正しく推定するように各アルゴリズムを修正する¹¹。こ

$\alpha = 0.7$ とした。

¹¹最も簡単な方法は、学習データとして選択された $y = 0$ の各データを w 個ずつコピーして学習データとして用いてしまうことである。ただしこれと同じ効果を学習アルゴリズム

ここでは k -NN, CART, RF に対してこのような修正を行った結果を表 3 に示す. ただし k -NN では $k = 50$ に設定し, CART, RF については決定木のノード数を 31 に設定した. 表中 $P(\hat{y} = 0)$ は単位修得不十分学生と推定した割合である.

表 3: 重みづけを行った時の誤分類率 ($d = 3$)

	w	P_E	P_{E0}	P_{E1}	$P(\hat{y} = 0)$
k -NN	2	0.220	0.374	0.155	0.291
k -NN	3	0.258	0.273	0.251	0.384
CART	2	0.231	0.335	0.187	0.328
CART	3	0.248	0.284	0.233	0.376
RF	2	0.218	0.364	0.155	0.299
RF	3	0.238	0.297	0.212	0.359

重みづけを行う場合にはパラメータによって結果が大きく異なり, また評価尺度が一つではないため最適性を追求することが難しい. ただしターゲットとした P_{E0} は表 1 の結果と比較して大幅に低くなっている. 一方誤分類率 P_E と P_{E1} が上昇してしまっていることも見て取れる. ここで $P(\hat{y} = 0)$ は単位修得不十分学生を推定したときに, 具体的な対策を施す必要のある学生の割合を意味しており, 重要である¹².

さて次に特徴ベクトルの属性数をもう少し増やし, $d = 5$ としたときの結果を表 4 に示す. $w = 1$ のときは重みづけしていない通常のアプローチと等価である.

表 4: $d = 5$ としたときの誤分類率

	w	P_E	P_{E0}	P_{E1}	$P(\hat{y} = 0)$
k -NN	1	0.178	0.479	0.052	0.187
k -NN	2	0.195	0.346	0.131	0.281
k -NN	3	0.231	0.263	0.217	0.363
CART	1	0.187	0.457	0.073	0.212
CART	2	0.206	0.351	0.145	0.293
CART	3	0.248	0.259	0.244	0.391
RF	1	0.181	0.454	0.066	0.208
RF	2	0.196	0.336	0.138	0.294
RF	3	0.217	0.268	0.196	0.354

表 4 の結果を表 2, 3 と比較すると, 推定精度が向上していることが見て取れる. これは 5 週目までの出欠情報を考慮したためであり, 自然である. また CART の結果があまり良くならず, ランダムフォレストによる推定が若干優れていること

ム中に組み込むことも比較的容易である.

¹²例えば個人的に注意喚起を行う, 面談に呼び出すなどの対策を検討する必要がある.

も見て取れる. ランダムフォレストは属性空間全体を利用し, また属性間の交互作用も考慮して予測を行っている効果が表れているものと思われる. ただしこれらの推定法を用いた場合, 単位修得不十分となる学生に対して誤分類率 P_{E0} を 40% 以下とするためには, 全体の約 30% 近くの学生に対して対策を施す必要がある.

4 まとめ

本稿では学生の授業出席データを収集・管理するシステム SAMS の概要を示した. また実際に得られた 1 年分の出欠情報と単位修得情報を用いて, 自動分類法を用いた単位修得不十分学生の推定法を検討し, 評価を行った. 今回は一番簡易的な週ごとの出席率のみを用いて推定を行ったが, 学生の所属する学科や担当教員, 過去の出席状況や単位修得状況など, 他の要因も特徴ベクトルとして多数利用することにより, 分類精度をさらに向上させる方法を検討する必要がある. また今回は 1,2 年生の前後期分である延べ約 2,000 件のデータを用いたが, 今後は每期全学生の出欠情報と単位修得情報がデータベースに蓄積される. これらをさらに有効活用することにより, 分類精度の向上が期待できる.

謝辞

本研究の一部は JSPS 科研費基盤研究 (C) No.25330045 及び JSPS 科研費基盤研究 (B) No.26282090 の助成を受けたものです.

参考文献

- [1] 後閑裕志, 大谷 真, “出席状況把握システムの解析”, 第 76 回情報処理学会全国大会予稿集, CD-ROM, 2014.
- [2] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone: “Classification and Regression Trees”, Wadsworth & Brooks, Monterey, 1984.
- [3] V. Vapnik, *Statistical Learning Theory*, Wiley; New York, 1998.
- [4] J. C. Platt : “Fast Training of Support Vector Machines using Sequential Minimal Optimization”, *Advancis in Kernel Methods: Support Vector Leaming*, pp.185-208, MIT Press, 1998.
- [5] L.Breiman: “Random Forests”, *Machine Learning*, Vol.45, No.1, pp.5-32, 2001.
- [6] C.Bishop: *Pattern Recognition and Machine Learning*, Springer, 2008.