

学生の出欠時間を活用した学生の友人関係分析

Analysis of friendship relation among students using attendance records

下村幸作^{1*} 中野智文² 犬塚信博² 松尾啓志^{2 3}

Kosaku Shimomura¹, Tomohumi Nakano², Nobuhiro Inuzuka², and Hiroshi Matsuo^{2 3}

¹ 名古屋工業大学 工学部電気情報工学科

² 名古屋工業大学 大学院工学研究科情報工学専攻

³ 名古屋工業大学 情報基盤センター

¹ Department of Electrical and Computer Engineering, Faculty of Engineering, Nagoya
institute of Technology

² Department of Science and Engineering, Graduate School of Engineering, Nagoya institute
of Technology

³ Information Technology Center, Nagoya institute of Technology

Abstract: We show a method to predict friendship relation among students using attendance time records to lectures. The record system consists of ID-cards of students and card readers which are equipped at lecture rooms and records times when a student attends or leaves a room. The distribution of difference of times recorded among two students depends if they are friends. We show a method to have the probability that a pair of students are friends using their time records and the distribution. We may use the probability to analyze group structure of a class and also to observe characteristics of classes.

1 はじめに

本学は充実した「学びの場」を構築するため、学生に対してきめ細やかな教育支援を検討している。そのために必要な学生のセグメンテーションを正確におこなうには、学生の性格や学習意欲などの定性情報と筆記試験の成績や出席率などの定量情報の両方が必要である。しかし、学生の定性情報はアンケートをおこなうなど収集するのに手間と時間のコストがかかる。また、収集したい定性情報が心理的に答えにくいものであれば正しく答えてくれる可能性が低くなるなどの問題を含んでいる。そこで、情報の収集が容易で客観性が高い定量情報から定性情報を導くことができれば、低コストで客観性がある定性情報を得ることができる。

これまでも電子化された教育支援システムによって学生の行動、性格等を把握しようとする研究がおこなわれている。佐藤ら [2] は履修に関するデータから学生の修学指導を早期におこなう手法を報告している。中山ら [3] は web を中心とした教育システムを活用し学生の特徴を把握する方法を検討している。高橋ら [4]

は携帯電話を利用した出欠管理システムを用いて授業支援を与えている。

本研究では、学生間の対人関係の近さという定性情報を定量情報である学生間の出席を記録した時刻(打刻時刻)から導く手法を提案する。学生間の対人関係の近さの分析するために、学生間の打刻時刻の差(打刻差時間)に着目した。一般に、対人関係が近いほど、その二人は共に行動する機会が多くなり、打刻差時間は短くなると仮定される。この仮定が正しいかどうかを、対人関係の近さを尋ねたアンケートと学生の出席を記録した時間のデータと照らし合わせて検証する。仮定が正しいことを示し、この特性を基に対人関係の近さを数値化する手法、さらにこれを用いた学生の友人関係分析の方法を提案する。

2 出欠管理システムについて

本研究で利用する打刻時刻は、名古屋工業大学が平成19年度に導入した出欠管理システム [1] から得る。出欠管理システムは、学生が講義に出席した記録を管理することを目的としたもので、学生証であるICカードから学生IDと時刻を読みとり記録するICカードリー

*連絡先: 所属機関 名古屋工業大学
住所 愛知県名古屋市昭和区御器所町
E-mail inuzuka@nitech.ac.jp

ダ(各教室に設置)と、その記録したデータを管理・蓄積するサーバから構成されている．システムを利用する学生は、IC カードリーダーに講義の開始時と終了時に学生証を提示し出席の記録をする．

本研究で使用する出欠管理システムにおいて打刻データは、打刻情報テーブルにあり、その属性は、打刻した人物を表す学生 ID、打刻した場所を表す IC カードリーダーの ID、打刻した時刻を表す打刻時刻の 3 属性である．

打刻情報テーブルの打刻件数は、平成 18 年 4 月 1 日から平成 18 年 11 月 16 日までの 1313207 件である．このデータに含まれている学生人数は 5214 人であり、IC カードリーダーの数は 125 個である．

3 友人と判別する手法の提案

友人関係が有する打刻差時間の特徴を示し、その特徴をもとに友人であるか否かを判別する手法を提案する．

3.1 打刻差時間における友人出現の特徴

学生のペアが友人関係にあるとき打刻差時間が短いことを検証するため、予備実験として出欠管理システムのデータベースにある打刻情報から打刻差時間を求め、アンケートにおいて友人関係にあると答えた学生ペアと友人関係にない学生ペアに分けて打刻差時間の度数分布を調べた．アンケート調査の項目等は 4 節で述べる．打刻差時間は学生の各ペアが同一の IC カードリーダーにそれぞれ記録したときの打刻時刻の差であるため、学生ペア中の双方向からみたプラスとマイナスの 2 種類の打刻差時間が算出される．例えば、学生 A が 9 時 15 分 30 秒に打刻し、次に学生 B が 9 時 15 分 35 秒に学生 A と同じ IC カードリーダーに打刻した場合、学生 A からみた学生 B との打刻差時間は - 5 秒となり、学生 B からみた学生 A との打刻差時間は + 5 秒となる．算出した打刻差時間の対人関係別度数分布を対人関係別に分けたグラフを図 1 に示す．

図 1 より、友人関係にある学生のペアは、友人関係以外の学生のペアよりも打刻差時間が短い範囲において打刻回数が多くなる．打刻差時間における友人の割合を図 2 に示す．横軸は対数である．この図から打刻差時間によって割合が大きく異なることがわかる．

3.2 友人の判別式と友人スコア

前節の友人ペアの特徴を生かし、ある学生ペアが友人関係にあるか否かを判別するための手法の提案する．ある学生 2 名に対し f をそのペアが友人であるという事象とし、 T はそのペアの打刻差時間データからなる

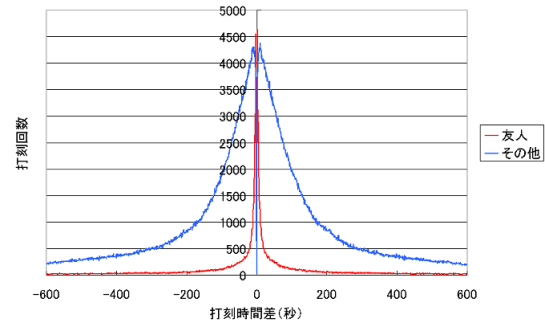


図 1: 打刻差時間ごとの対人関係別の度数分布

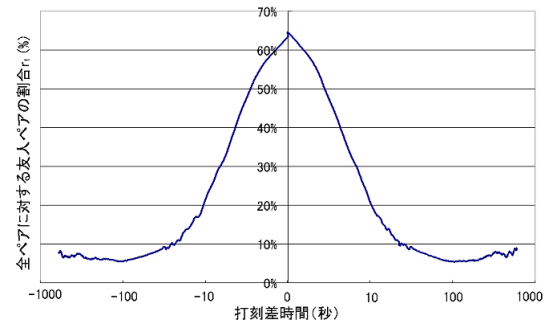


図 2: 打刻差時間ごとの友人ペアの割合 (r_t)

集合 $\{t_1, t_2, \dots, t_n\}$ とする．このとき T に対する f の事後確率 $p(f|T)$ はベイズの定理を用いて以下のとおり得られる．

$$p(f|T) = \frac{p(f) \cdot p(T|f)}{p(T)}$$

ここで、打刻差データ T 中の各要素 t について発生確率 $p(t)$ は、各々独立と仮定する．また事象 f に対する条件付き確率 $p(t|f)$ も各要素で独立とする．すると、次のとおり変形できる．

$$p(f|T) = p(f) \prod_{t \in T} \frac{p(t|f)}{p(t)} \quad (1)$$

次に図 2 で示した、打刻差時間ごとの友人ペアの割合について考える．図 2 の値は打刻差時間 t で打刻するすべての打刻対の内、友人であるものがおこなった打刻対の割合であった．この値を r_t としたときこれは次のとおり表せる．

$$\begin{aligned} r_t &= \frac{\text{友人ペアに限定した打刻差 } t \text{ の総データ数}}{\text{打刻差 } t \text{ の総データ数}} \\ &= \frac{X_f \cdot m_f \cdot p(t|f)}{X \cdot m \cdot p(t)} \end{aligned}$$

ここで X は学生ペアの数、 X_f は友人ペアの数、 m は 1 組の学生ペアから発生する打刻データの平均件数、 m_f

は友人ペアに限定したとき 1 組の学生ペアから発生する打刻データの平均件数である．すると次式が得られる．

$$p(t|f) = \frac{X \cdot m \cdot p(t) \cdot r_t}{X_f \cdot m_f}$$

$X_f = X \cdot p(f)$ であるため

$$p(t|f) = \frac{m \cdot p(t) \cdot r_t}{p(f) \cdot m_f}$$

となる．これを式 (1) に代入する．

$$\begin{aligned} p(f|T) &= p(f) \prod_{t \in T} \frac{m \cdot r_t}{m_f \cdot p(f)} \\ &= p(f)^{(n-1)} \left(\frac{m}{m_f} \right)^n \prod r_{tf} \end{aligned}$$

ここで n は T に含まれるデータ数である．また，友人ではない確率 $p(\bar{f}|T)$ も同様に

$$p(\bar{f}|T) = p(\bar{f}) \prod_{t \in T} \frac{p(t|\bar{f})}{p(t)} \quad (2)$$

である．また各打刻差時間での友人以外ペアの割合 $1-r_t$ は次のとおり表せる．

$$\begin{aligned} 1 - r_t &= \frac{\text{友人でないペアの打刻差 } t \text{ の総データ数}}{\text{打刻差 } t \text{ の総データ数}} \\ &= \frac{X_o \cdot m_o \cdot p(t|\bar{f})}{X \cdot m \cdot p(t)} \quad (3) \end{aligned}$$

ここで X_o は友人以外ペアのみの数， m_o は友人ペア以外に限定したとき 1 組の学生ペアから発生する打刻データの平均件数である．

式 (3) と $X_o = X \cdot p(\bar{f})$ より $P(t|\bar{f})$ を求め，式 (2) に代入して次を得る．

$$p(\bar{f}|T) = p(\bar{f})^{(n-1)} \left(\frac{m}{m_o} \right)^n \prod (1 - r_{tf})$$

$p(f|T)$ にロジット関数を使用し，友人の目安とする．

$$\begin{aligned} \text{logit } p(f|T) &= \log \left(\frac{p(f|T)}{1 - p(f|T)} \right) \\ &= \log(p(f|T)) - \log(p(\bar{f}|T)) \quad (4) \end{aligned}$$

この式は，友人であるか否かの目安とすることができる．これより求めた値を友人スコアと呼ぶことにし，その値が正であれば友人であると判別する．

4 友人スコアの算出と実験及び結果

本節では，3 節で求めた提案手法に実際のデータを適用して友人スコアを算出をする．

4.1 友人スコアの算出

前節で導出した友人スコアの算出の流れを次に示す．(図 3 参照)

1. 出欠管理システムのデータベース内のすべての打刻データについてペアをつくり打刻差時間データを得る．ただし，友人を同定するのに必要な範囲として 10 分以内の時間差のみを扱う．
2. 打刻差時間データの件数を学生ペア数 (= 学生数 $\times$ (学生数 - 1)) で割り m を推定する．同様に友人ペアに限定した打刻差データ件数を友人のペア数で割り m_f を，非友人ペアのデータ件数を非友人ペア数で割り m_o と推定する．
3. 各打刻差時間に着目し，1 秒ごとの時間差における打刻回数を数える．その数をアンケートから友人ペアとそれ以外のペアに分け，各打刻差時間ごとの友人ペアの割合 r_t を推定する．
4. アンケートにより友人ペアの数を全ペアの数で割り，友人ペアの割合を得る，これによってペアが友人である事前確率 $p(f)$ を推定する．
5. 上記で算出したパラメータと各学生ペアの打刻差時間データを式 (4) に代入し，友人スコアを算出する．

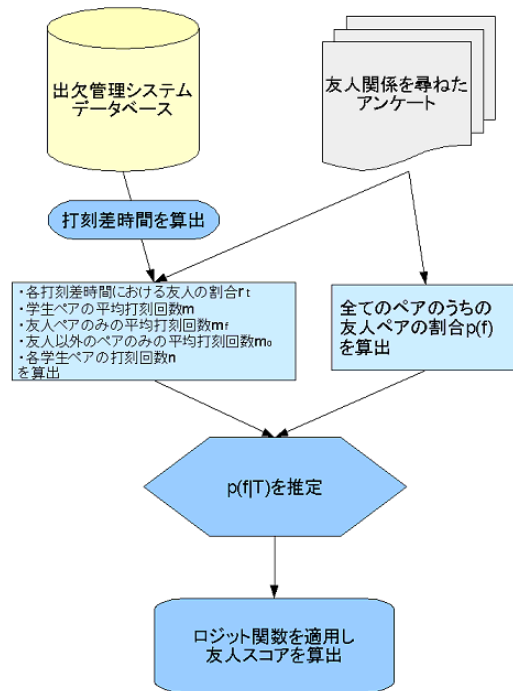


図 3: 友人スコアの算出までの流れ

4.2 対人関係の近さを尋ねたアンケート

ある講義の受講生に対して、他の受講生との対人関係の近さを尋ねた。アンケートの質問内容は、1:友達、2:話す、3:知っている、4:それ以外の4つのグループに対し、相手がどれに属するかを尋ねた。このアンケートを3つの講義で実施した。回答者のデータは表1に示し、アンケート結果のは表2に示す。なお、学生ペアをみたとき、友人関係の有無が双方向ごとにある。

表 1: アンケート回答者

	講義 K	講義 J	講義 P
受講者数	144	61	66
実施人数	90	58	44
無回答者数	21	14	8

表 2: アンケート結果

	講義 K	講義 J	講義 P
友人ペアの数	382	175	108
話すペアの数	380	125	162
知合いペアの数	948	162	145
それ以外ペアの数	8014	2178	1949

4.3 出欠管理システムの蓄積データ

出欠管理システムに蓄積されているデータの中で、今回の分析に使用するデータの種類の記録した人物を表す学生 ID、記録した時間を表す打刻時刻、記録をおこなった場所を表す IC カードリーダーの ID である。学年ごとの総打刻回数と一人当たりの打刻回数と利用者数を表3に示す。4年生は講義開講数が少なくデータ量が少ないため、分析はしない。

表 3: 出欠管理システムの利用状況

学年	総打刻回数	1人当たりの平均打刻回数	利用者数
1	400746	392	961
2	363950	370	956
3	256830	269	941
4	36808	39	632

4.4 友人関係に関する実験結果

前節で述べたパラメータを算出手順にしたがって各パラメータを算出した。ただし、打刻差時間が ± 30 秒以上はデータ量が少なく、友人の割合がバラツキが発生するため 3 1 秒間の移動平均を用いて平滑化をおこなった。また、打刻差時間が 0 秒のデータも同様にデータが少ないため ± 1 秒の友人の割合の平均を用いる。

これらにより、求めたパラメータを表4に示し、 r_t を図4に示す。

表 4: 各講義から算出した各パラメータ

	$p(f)$	m	m_f	m_o
講義 K	0.393	80.5	168	76.7
講義 J	0.663	162	330	151
講義 P	0.477	74.4	138	71.2
平均値	0.501	106	212	99.7

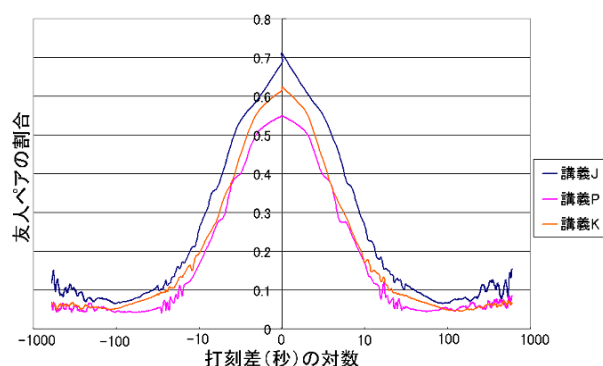


図 4: 打刻差時間ごとの対人関係別の度数分布 r_t

提案手法の結果とアンケートの結果の比較をおこなう。具体的には各講義で求めたパラメータを各講義の友人スコアに適用して求めた。友人スコアが正のペアとアンケートで友人であると回答したペアを比較して、友人の再現率と適合率を算出した。結果を表5に示す。また、3つの講義を使って交差検定をおこなった結果は、友人の再現率は67.8%となり、友人の適合率は56.4%となった。

アンケートを実施した講義から算出した各パラメータの平均を使用してアンケート対象の受講生のペアの友人スコアを求め、その友人スコアをアンケートをもとに友人と友人以外の2種類に分類する。友人と回答した方の度数分布を図5に、友人以外と回答した方の度数分布を図6に示す。図5は、友人スコアが正の値であれば友人であると正しく判定しており、逆に友人スコアが負であれば誤って友人以外である判定してい

表 5: 友人の再現率と適合率 (%)

パラメータ の提供側	評価 の尺度	パラメータの使用側		
		講義 K	講義 J	講義 P
講義 K	再現率	75.3	90.3	77.8
	適合率	60.0	54.5	16.5
講義 J	再現率	63.4	86.9	66.7
	適合率	86.4	88.9	53.7
講義 P	再現率	55.3	76.0	65.7
	適合率	88.2	93.0	78.9

る．同様に図 6 は，友人スコアが負の値であれば友人以外であると正しく判定しており，逆に友人スコアが正の値であれば誤って友人であると判定している．

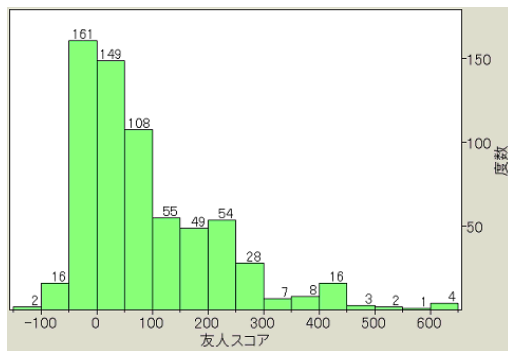


図 5: 友人のみの友人スコアの度数分布

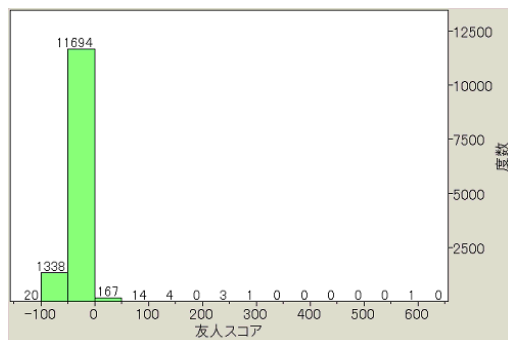


図 6: 友人以外の友人スコアの度数分布

アンケートにおいて答えるものの判断基準が一定してないことを考慮すると今回得られた再現率，適合率は満足できると考える．この値の正しい評価はアンケート自身の精度を含めて検討する必要がある．

5 友人スコアの応用例

5.1 組織ごとの友人の数の分析

友人スコアを用いて学年・学科などの組織の単位で友人の数の特徴を把握することができる．

表 4 の平均値のパラメータを使い友人スコアを求め，一人の学生の友人スコアの正のものを数えることでその学生の友人の数を求めた．これに関するヒストグラムを求めた．学年毎のヒストグラムを図 7 に示す．

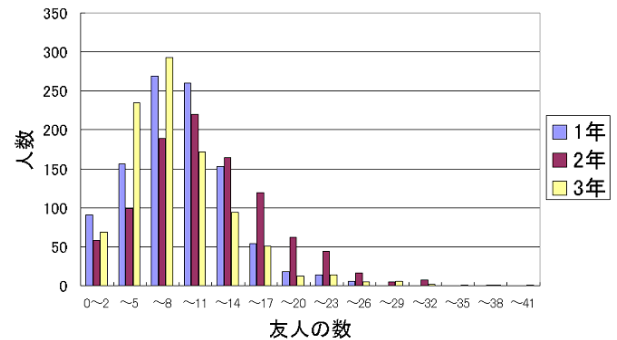


図 7: 学年ごとの友人の数の度数分布

同様の分析は学科毎にも可能であり，こうした単位での特性を見ることができる．学科ごとの友人の数の分布を図 8 に，学年・学科ごとの一人当たりの友人の数の平均を表 6 に示す．

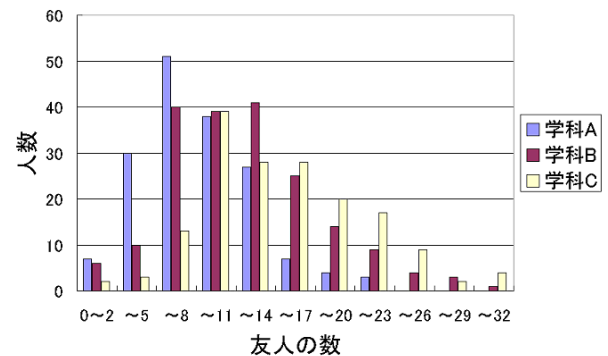


図 8: 学科ごとの友人の数の度数分布

図 7 からは，1 年生のなかで友人の数が少ない学生があり，まだ大学に馴染めていないことが想像できる．図 8 からは，学科 C の友人の数が多いため，共に行動している機会が多いと読みとれる．表 6 からは，学科間の差異が 1 年生よりも 2 年生で顕著になることが読みとれる．

表 6: 学科ごとの一人当たり友人の数の平均

学科	1 年生	2 年生
a	7.93	8.69
b	8.77	11.3
c	9.91	11.9
d	8.87	11.0
e	10.1	14.6
f	10.1	10.4
g	8.96	9.80

5.2 学生の友人グループの構造分析

お互いに親しい学生が集まるグループを把握することは、本プロジェクトにおいて重要である。ここではクラスタリングによってグループを分析する方法について検討する。

友人スコアを表 7 のように各学生の総当たりにした表を求めたとき、その各行は対応する学生の属性ベクトルとみることができ、このベクトル間距離は類似度を与えると考えられる。なお相手が自分自身となる場合は友人スコアを 0 とした。この友人スコアのクロス表に対して、探索的統計解析ソフトの JMP の階層的クラスタ分析をおこなった。学生間の距離を求める手法には Ward 法を使用した。Ward 法の距離の 2 乗 D_{KL} は、次の式である。

$$D_{KL} = \frac{\left| \overline{X_K} - \overline{X_L} \right|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

ここで、 $\overline{X_K}$ は K 番目のクラスターの平均ベクトルであり、 N_K は K 番目のクラスターに含まれるオブザベーションの数である。

表 7: 友人スコアのクロス表

	学生 A	学生 B	学生 C
学生 A	0	103	-1.52
学生 B	103	0	-10.5
学生 C	-1.52	-10.5	0

講義 J のクラスタリングをおこなったときのデンドログラムを図 9 に示す。

図 9 からは、親しい学生が集まるグループが読みとれるが、特定の人とあまり共に行動しない学生をどのグループに属しているかが判断がつきにくい。

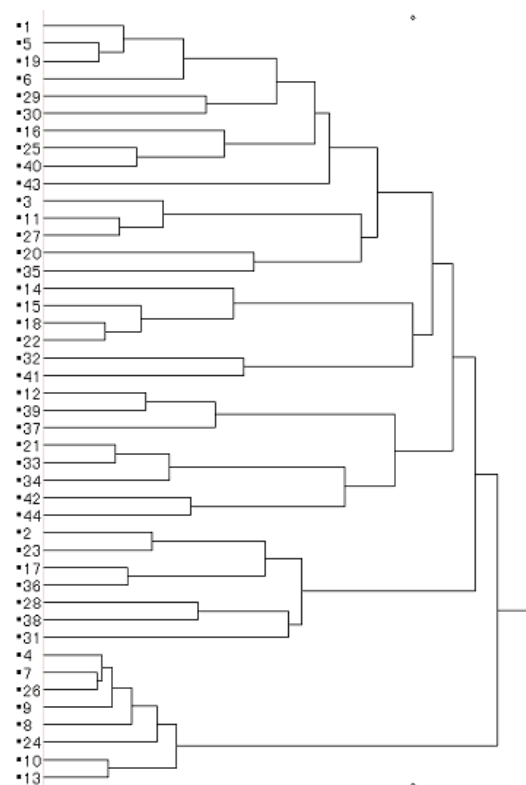


図 9: 講義 J のデンドログラム

5.3 学生間の距離の可視化

前節のクラスタリングでは、相手の学生の数だけ次元であるため直観的に学生同士がどの程度近いのか、遠いのが理解しにくい。そこで、学生を二次元上にプロットすることで学生間の距離を可視化し、学生同士の近さを把握する方法を検討する。

クラスタリングをおこなった友人スコアのクロス表で列を変数として主成分分析をおこない、第一主成分と第二主成分を求めた。これを 2 軸として二次元のグラフとする。

講義 J に対して、可視化をおこなったものを図 10 に示す。なお、前節の階層的クラスタリングをおこない、8 つのグループごとに分け、学生を示すマークはグループごとに異なっている。

図 10 からは、学生グループ間の距離がある程度読みとれる。しかし情報量を第一主成分と第二主成分の寄与率の合計が約 27% であり、必ずしも適切な分析が得られなかった。

6 まとめと今後の課題

本研究では、IC カードとカードリーダーを用いて、学生が講義室に入室、退室した時刻を記録する出欠記録

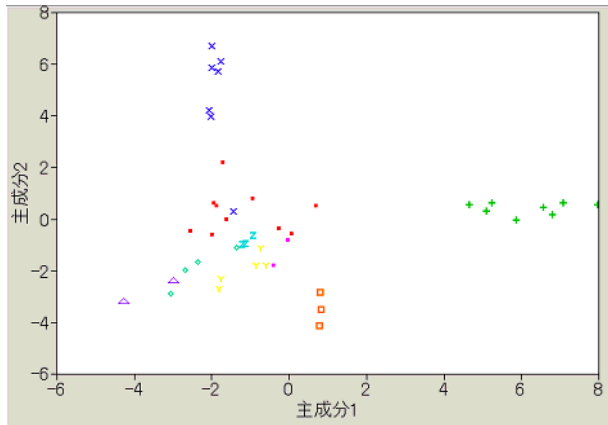


図 10: 講義 J の友人グループの可視化

システムを活用し、学生間の友人関係を推測することができることを示した。学生の 2 人の組の打刻差時間が、その組が友人であるかどうか大きく依存することに注目し、その組の打刻差データから友人である確率を算出した。この結果を用いて、1 つのクラスの中での友人グループの構成や、学科・学年などの違いによるクラスの特徴分析などに使用できることを例に挙げて示した。

今後の課題は、友人スコアの価値を高めることである。方法は、次の 2 点が考えられる。まず、友人スコアを成績情報やアンケートなどの学生情報と連係していくことで学生を多角的にとらえる方法である。次に、友人スコアの導出仮定において、友人とそれ以外の人の異なりが大きくなっている時間帯に区切って分析するなどによって、友人スコアの精度を向上させる方法である。

また、IC カードを活用した電子マネー、社員証、交通機関の乗車券など今後も世の中に普及していくことが予想される。これらにおいても、IC カードが記録した情報を分析するために本研究の方法と同様の分析が可能と考えられる。

謝辞

本研究は、文部科学省特別研究経費“充実した「学びの場」の構築 - 教員の教育力向上及び双方向型教育支援システムの整備 - ”の支援によっておこなわれた。

参考文献

- [1] 伊藤宏隆, 舟橋健司, 中野智文, 松尾啓志, 内匠 逸, 大貫徹者: コースマネジメントシステム Moodle と IC カード出欠システムとの連携, 平成 19 年度電気系学会東海支部連合大会講演論文集 O-376 (2007)
- [2] 佐藤和彦, 大川輝人: 事前対応型の修学指導支援システム, 電子情報通信学会技術研究報告, ET2007-34 (2007)
- [3] 中山実, 山本洋雄, Santiago Rowena: プレンディット学習の学習行動における学習者特性の影響, 電子情報通信学会技術研究報告, ET2007-40 (2007)
- [4] 高橋道祐, 吉田法雅, 小林洋介, 小林洋介, 熊澤弘之: 携帯端末を用いた出席登録・管理及び授業支援システムの開発, 電子情報通信学会技術研究報告, ET2005-86 (2006)