

1 はじめに

本学は充実した「学びの場」を構築するために、早期学習指導システムの導入を検討している。そのシステムには、学生のようなデータとそのデータを加工し意味のある情報にするデータマイニングの手法が必要である。本研究では、学生が IC カードを使用して講義の出欠を記録している打刻情報のデータから、学生間の対人関係の親密さを分析する手法を提案する。

2 打刻差時間に現れる友人の特徴

学生ペアが同一の機器に出席の記録したときの打刻時間の差 t を算出した。これを友人関係の有無を学生に尋ねたアンケートの結果を用いて友人ペアとそれ以外ペアの二つに分類した。これにより、各打刻時間差における友人ペアの割合を得た。これを r_t とし、その結果を図 1 に示す。打刻時間差が短いとき、友人ペアの割合が大きいことがわかる。

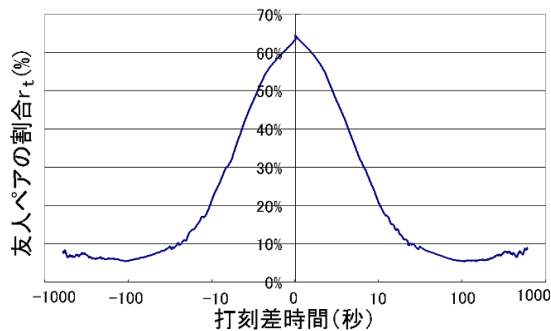


図 1: 各打刻時間差における友人ペアの割合 (r_t)

3 友人と判別するための手法の提案

前節の友人ペアの特徴を生かし友人と判別するための手法の提案する。 f をある学生のペアが友人であるという事象とし、そのペアの打刻差データの集合 T に対する事後確率としてベイズの定理を用いて以下のとおり得る。

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

また、友人ペアの割合 r_t は次のとおり表せる。

$$\begin{aligned} r_t &= \frac{\text{友人ペアに限定した打刻差 } t \text{ の総データ数}}{\text{打刻差 } t \text{ の総データ数}} \\ &= \frac{X_f \cdot m_f \cdot p(t|f)}{X \cdot m \cdot p(f)} \end{aligned} \quad (2)$$

ここで X は学生ペアの総数、 X_f は友人ペアのみの総数、 m は全ての学生ペアの平均打刻回数、 m_f は友人ペアのみの平均打刻回数である。この式 (2) から $p(t|f)$ を求め、式 (1) に代入し $p(f|T)$ が得られる。こ

れにロジット関数を使用し、友人の目安とする。

$$\text{logit } p(f|T) = \log p(f|T) - \log p(\bar{f}|T) \quad (3)$$

この式より求めた値を友人スコアと呼び、これが正の値ならば友人であると判別する。

4 提案手法の適用の結果と考察

前節で述べた提案手法を実際のデータに適用した。具体的には、いくつかの講義を単位としたアンケートによって受講生間の友人関係の有無に関するデータを得た。これと打刻情報のデータから各パラメータ m , m_f , $p(f)$, $p(\bar{f})$, r_t を算出し、式 (3) によって友人スコアを求めた。アンケートで友人であると答えたペアが提案手法によっても友人と正しく判断された割合を示す再現率は 67.8% となり、提案手法によって友人と判断したペアのうちアンケートでも正しく友人であると答えているペアの割合を示す適合率は 56.4% となった。友人と非友人が正しく判別した全体の精度は 96.0% となった。ただし、打刻時間差は ± 10 分以内に限定した。

アンケートにおいて答えるものの判断基準が一定して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

友人スコアを活用して、親しい学生同士集まりを作成する階層的クラスタリングや学年・学科ごとの友人の数の特徴を分析が可能となった。その例として、図 2 に学年ごとの友人数の分布を示す。

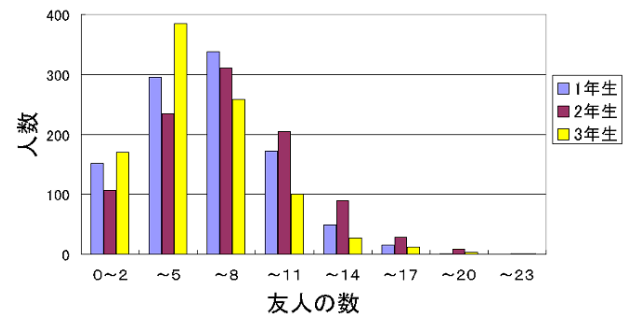


図 2: 学年ごとの友人数の分布

6 まとめと今後の課題

対人関係が親密な友人とは、共に行動する機会が多いため、打刻時間差が短い特徴があった。それにより、出欠管理システムの打刻情報を分析することで学生の対人関係の親密さを予測する友人スコアを導くことができた。また今後の課題は、友人スコアの価値を高めることである。そのためには、大学が有している学生のような情報を友人スコアと関係させることや友人スコア自体の精度の向上が必要である。