

## 1 はじめに

本学は、教育支援の一つとして平成 19 年度に出欠管理システムを導入し、下村 [1] らは打刻データから友人関係が推測出来ることを示した。

このように定量的なデータからネットワークの構造を分析することがより正確に行われるようになれば、社会ネットワーク分析の分野に貢献できると言える。本研究では下村らが打刻データを単一のものとして捉え、友人関係の推定を行ったのに対し学生が行ったそれぞれの打刻データに、講義の出席時や退席時に行った打刻である、というような付属情報を加えることで友人関係推定の整合性を高めることができるかを検証する。

## 2 打刻差時間と友人関係の特徴

学生ペアが同じ日に同一の IC リーダに打刻した時の時間差  $t$  を算出した。学生に友人関係を尋ねたアンケートを用い、友人ペアと非友人ペアに分けた。これにより各打刻時間差  $t$  毎の友人ペアの割合を算出し図 1 に示す。打刻時間差が近い学生ペアほど友人関係にある割合が多いことが分かる。

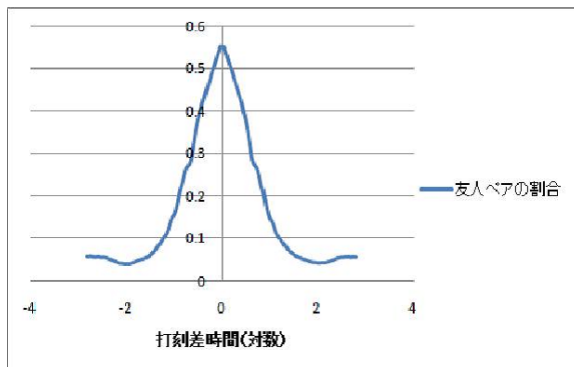


図 1: 各打刻時間差毎の友人ペアの割合

## 3 属性付きデータを用いた友人推定の手法

本研究では打刻時間差  $t$  に対し属性を付与し、友人関係を推測する。ここで付与する属性とは行われた打刻が講義の出席時に行われたものなのか、退席時に行われたものなのかといった打刻データに関する付属情報である。ある学生ペアが友人であるという事象を  $f$  とし、そのペアの属性付き打刻差のデータ列  $T_a$  に対する事後確率はベイズの定理を用いて以下となる。

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

属性付き打刻時間差毎の友人ペアの割合  $r_{t_a}$  や各種パラメータを用いて式 (1) を変形すると次式を得る。

$$p(f|T_a) = p(f)^{(n_a-1)} \prod_{t_a \in T_a} \frac{m_a \cdot r_{t_a}}{m_{af}} \quad (2)$$

ここで  $m_a$  は各属性付き打刻データにおけるすべての学生ペアの平均打刻回数で、 $m_{af}$  は各属性付き打刻データにおける友人ペアのみの平均打刻回数である。この式 (2) にロジット関数を使用して属性付き打刻データを用いた友人スコアを算出する。

## 4 属性付きデータの適用

下村らの研究では友人スコア算出において、学生間の打刻時間差に注目した。しかし友人関係を推定する上で打刻データを単一のものとして捉えるのではなく、それぞれのデータに意味があるものと考え、分析を行う。図 2 は打刻データ回数を講義の出・退席と時間割別に示したものである。各属性毎に打刻回数に違いがみられ、この違いが友人スコア算出に影響を及ぼすと推測する。下の表は属性付きデータと属性なしのデータを用いた友人スコア算出結果である。結果に二つの間での大きな変化は見られなかった。

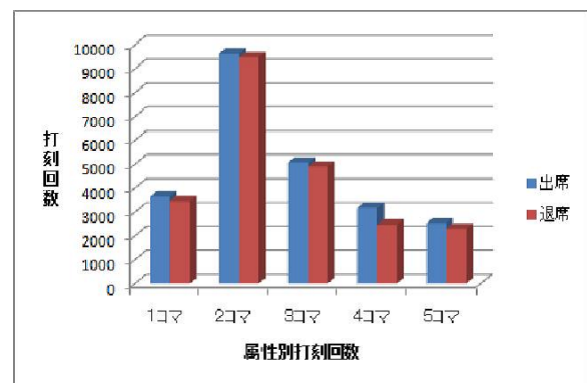


図 2: 属性別打刻回数

出・退席属性	適合率	再現率
なし	73.4%	65.2%
出席	64.4%	62.9%
退席	65.5%	65.0%
出・退席	55.9%	65.9%

## 5 まとめ

打刻データに新たな情報を加えることに注目し、これを友人関係の推定の算出式に組み込んだ。しかし、新たな情報を用いて行う友人推定では友人スコアの精度の大幅な変化は見られなかった。他の観点から打刻情報に手を加える、もしくは別の新たな情報を組み込むことで友人推定の精度向上を検討したい。

### 参考文献

- [1] 下村幸作、中野智文、犬塚信博：学生の出欠時間を活用した学生の友人関係分析、第 6 回学会データマイニングと統計数理研究 (SIG-DMSM) (2008)