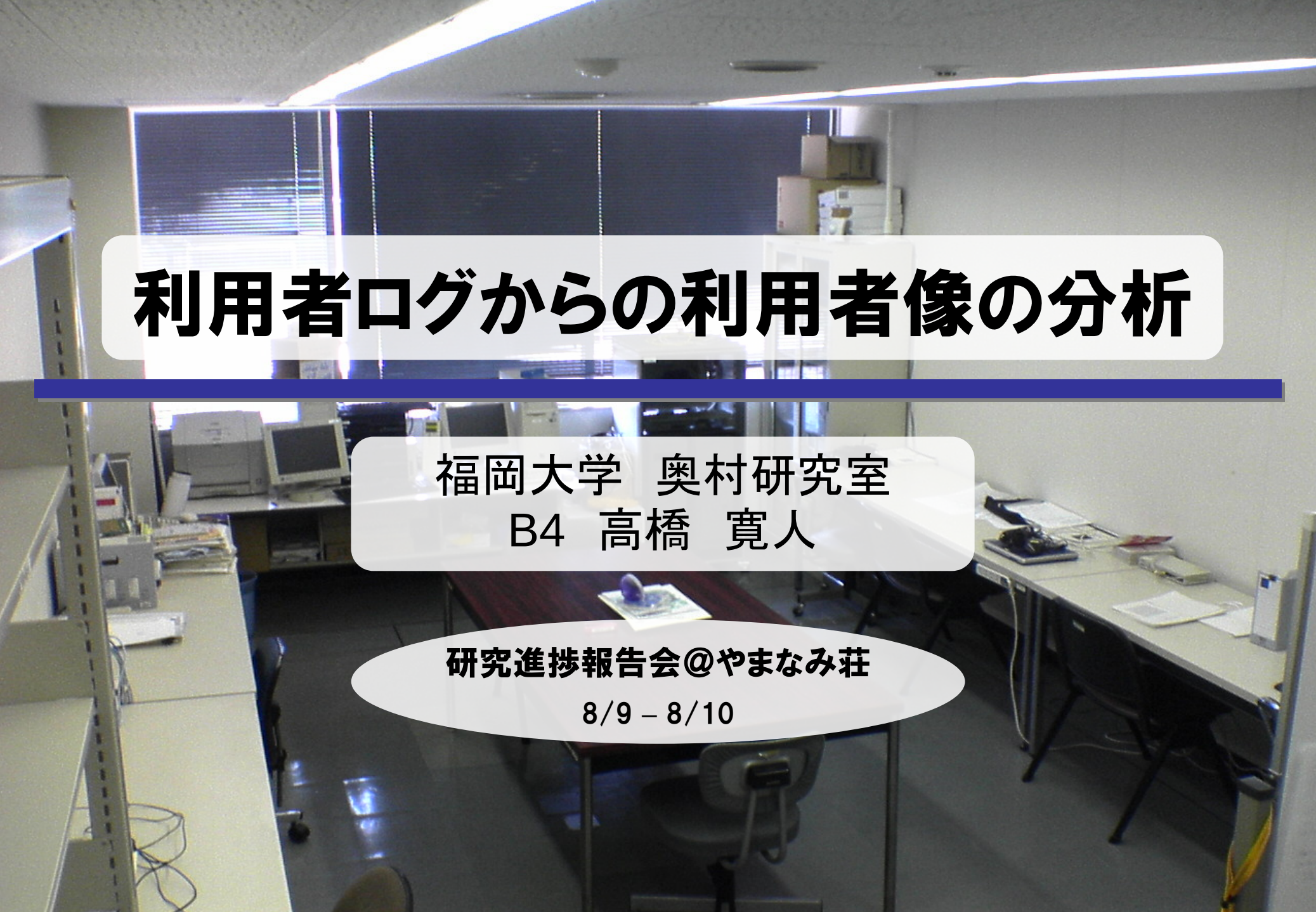


利用者ログからの利用者像の分析

福岡大学　奥村研究室
B4　高橋　寛人

研究進捗報告会@やまなみ荘

8/9 – 8/10

発表内容

- 研究の目的
- 使用する利用者ログの種類
- プリンタ利用ログの解析
- データマイニングの概要
- 今後の課題

研究の目的

- 福岡大学約2万人の学生が学内PCを利用することで残る様々なログを解析して、利用者像の実態や相関性を分析すること。
- さらにデータマイニングという技術を駆使して、大規模データに隠された価値ある情報を引き出すこと。

使用するログの種類

1. 各種ログインログ
   Windows,Linux,汎用UNIXサーバへのログインログ
   日時、対象ホスト、利用者名（学籍番号）
2. ActiveMailログ
   福岡大学webメールへのログインログ
   日時、接続元IPアドレス、利用者名（学籍番号）
3. 各種アプリケーション利用ログ
   Windows,Linux,汎用UNIXサーバ上のアプリケーションの起動ログ
   日時、対象ホスト、対象アプリ名、利用者名（学籍番号）
4. プリンタ利用ログ
   日時、対象プリンタ名、印刷枚数、利用者名（学籍番号）
5. 学内DHCP情報コンセント利用ログ
   日時、利用者名（学籍番号）
6. RAS（PPP接続）利用ログ
   日時、利用者名（学籍番号）
7. SSL-VPN接続利用ログ
   日時、利用者名（学籍番号）
8. 学生ポータルログインログ
   日時、接続元IPアドレス、利用者名（学籍番号）
9. 休講補講情報のアクセスログ
   日時、利用者名（学籍番号）
10. パスワード変更者ログ
    日時、利用者名（学籍番号）

プリンタの利用ログの解析

- 福岡大学内に設置された42台のプリンタの7か月分（2005年9月～2006年3月）の利用データから、利用者名・日時・印刷枚数を抜き出して、質の高いデータとしてまとめる。
- またそのデータを元に利用者像の実態を探る。

月別印刷枚数

	A3未満モノクロ（枚）	A3以上モノクロ（枚）	A3未満カラー（枚）	A3以上カラー（枚）	合計（枚）
2005年9月	88，932	337	2，269	191	91，729
2005年10月	155，899	1，090	3，326	149	160，464
2005年11月	148，009	590	2，501	72	151，172
2005年12月	186，392	1，276	4，751	81	192，500
2006年1月	162，003	2，118	4，516	80	168，717
2006年2月	51，138	261	2，716	39	54，154
2006年3月	61，482	536	4，408	215	66，641
合計（枚）	853，855	6，208	24，487	827	885，377

データマイニング概要

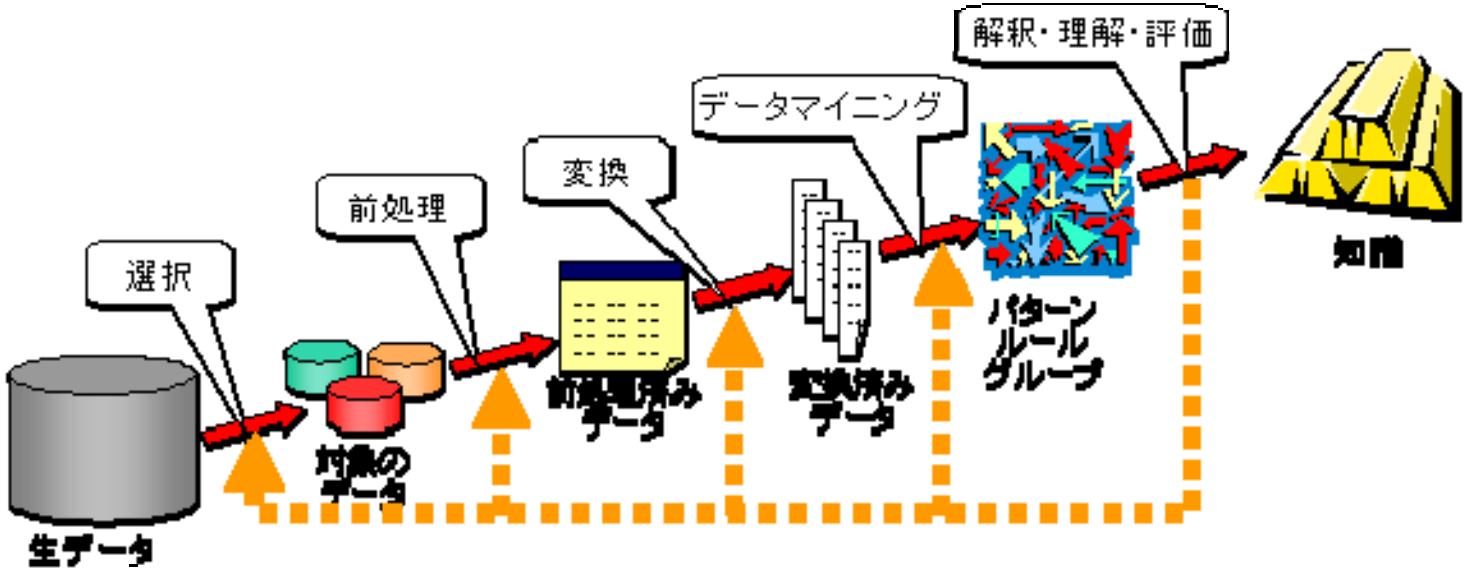
◆データマイニングとは？

- データの中に潜んでいる価値ある情報を掘り出す（mine:掘る）こと
- 大規模データに対応可能なデータ処理技術
- データを知識としてより戦略的に活用すること

技術的背景

- 安価で大容量の記憶装置が開発されネットワーク環境が整備された
- 高性能の演算処理装置や大容量のメモリが低価で利用できるようになった
- 解析技術の高度化

データマイニングのプロセス



知識発見の手法

分野	具体的な手法
人工知能	人工知能（AI）
遺伝アルゴリズム	遺伝アルゴリズム（GA）
決定木	ID3,C5.0，C4.5，エキスパートシステム
ニューロ	ニューラルネットワーク、ファジィ理論
パターン認識	k近傍法，k-NNルール、マハラノビスの汎距離、ベイズの決定境界
データ解析	相関分析、バスケット分析、主成分分析、クラスター分析、因子分析
確率論	確率論、各種データ分布
統計学	統計学、ビジュアリゼーション（可視化）、推定、検定、有意水準

適用分野と利用方法の事例

- **流通サービス業**　マーケットに関する情報を採取
  販売予測：天気や催し物、過去の実績といったデータから商品ごとの特性や購買数を分析したうえで、仕入れ量を決定する。
  販売分析：顧客（個客）の購買パターンや「紙おむつを買った顧客はビールを同時に買う」というような隠れた併売パターンを見つけ、収益性をより高める。
- **製造業**　コストや品質管理に関わる原因を分析
  品質管理：製造工程の不良原因を分析し、不良多発原因を排除する。
- **金融・保険業**　リスクや顧客特性の情報を採取
  リスク分析：過去の実績や契約者の属性データから、契約のリスクを分析し、上限値等を決定する。
  特性分析：顧客属性と契約商品の関係を分析し、顧客ごとに効果のある商品を紹介する。また、顧客離れを防止する。
- **医療・バイオ産業**　要因や効果に関する情報を採取
  要因分析：病気の原因を分析したり、薬の改善効果を調査する

今後の課題

- データマイニングを行う上で、各種ログの生データから適切なデータを抽出する。
- データマイニングの理解を深め、知識発見手法の検討をする。

終わりのコメントを入力
（ご清聴ありがとうございました等）