

# ゲームマイニングの野望II ～MMOGのデータマイニング

立命館大学情報学科  
知能エンターテインメント研究室  
ラック・ターウォンマツ  
03/09/04

## 目次

- **第一部**
  - データマイニングの概要
  - MMOGへの応用例
- **第二部**
  - MBR法
  - K-means法
- **第三部**
  - 研究室の紹介
  - 研究成果

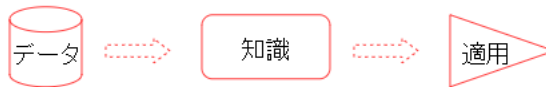
# 第一部

## なぜデータマイニング？

- **MMOGはeBusiness**
  - 商売の対象はインタラクティブ・コンテンツ
- **eBusinessの戦略**
  - 加入者(プレイヤー)獲得
  - 加入者保持
  - 購買(プレイ)の促進
- **プレイヤーを知る必要がある**
  - ツールはデータマイニング

# データマイニングとは

データマイニングとは、「大規模データ」から法則性や傾向などの「知識」を抽出する手法である。



データマイニングには、  
1. 知識の抽出と、  
2. 抽出された知識の適用  
という2つのフェーズが考えられる。

# MMOGへの応用例

- **実施済み**
  - バランスの可視化
  - 改造屋の特定
- **他の候補**
  - ボット、荒らしの発見
  - イベント参加者の選定
  - ホットな話題の抽出

## バランスの可視化(1/2) (Nexon's Dark Agesの例, REF#1)

- 仮説:
  - 職業別, レベルごとの平均Experience points per hour (EPH)からバランスが分かる
- データ:
  - 各プレイヤーに対して, 週ごとのデータから次のようなレコードを作成

ID	職業	レベル	獲得経験値	プレイ時間	EPH	EPH/AVG
1	fighter	15	320	2.0	160	1.6
2	wizard	15	160	4.0	40	0.4
3	fighter	19	300	3.0	100	1.0

-  $EPH = \text{獲得経験値} / \text{プレイ時間}$ ,  $AVG = EPH$ の平均

03/09/04

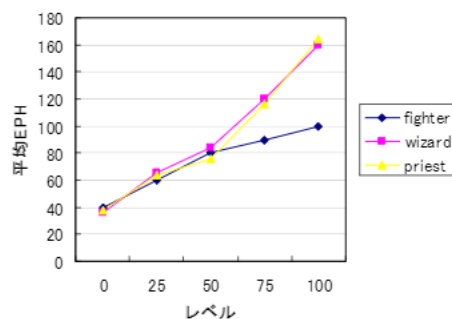
立命館大学, ラック・ターウォンマット

7

## バランスの可視化(2/2)

- 工程:
  - 職業別のレベルごとの平均EPHを図化

- 結果:



- 適用:
  - 中レベルのfighterに新しい攻撃スキルを追加

03/09/04

立命館大学, ラック・ターウォンマット

8

## 改造屋の特定 (REF#1)

- 仮説:
  - 不自然な高いEPHを持つプレイヤーは改造屋の可能性がある
- データ:
  - バランスの可視化の形式と同じ
- 工程:
  - 表におけるEPH/AVGの高いプレイヤーに着目
- 結果:

ID	職業	レベル	獲得経験値	プレイ時間	EPH	EPH/AVG
1	fighter	15	320	2.0	160	1.6
2	wizard	15	160	4.0	40	0.4
3	fighter	19	300	3.0	100	1.0

- 適用:
  - 該当プレイヤーをゲームマスターにより最終確認し、その後は運営方針に従う

03/09/04

立命館大学, ラック・ターウォンマツ

9

## 特殊プレイヤーの特定(1/2)

- 仮説:
  - ボットの行動, 荒らしやの行動にはパターンがある
- データ:
  - 学習用ゲームマスターにより過去に特定したボット, 荒らしや, 一般プレイヤーに対して, 各アクションの実行頻度から次のような項目からなるレコードを作成

「ID」, 「アクション1頻度」, . . . , 「アクションn頻度」, 「プレイヤータイプ」

- 特定用
- 特定対象のプレイヤーに対して, 学習用と同じ形式のレコードを作成
- 
- ただし, 「プレイヤータイプ」は空に

03/09/04

立命館大学, ラック・ターウォンマツ

10

## 特殊プレイヤーの特定(2/2)

- 工程:
  1. 学習用データを用いて分類器(第2部のMBRなど)を学習
  2. 特定用データに対して学習済み分類器を適用
- 結果:

特定用の各レコードにプレイヤータイプが決定
- 適用:
  - 該当プレイヤーをゲームマスターにより最終確認し、その後は運営方針に従う

## イベント参加者の選定(1/2)

- 仮説:
  - イベントの成功はプレイヤータイプの割合(レシビ)が決め手
  - タイプは職業, レベル, 経験値, 行動パターンで定義
- データ:
  - クラスタリング用  
成功したイベントにフル参加したプレイヤーに対して, 次のような項目からなるレコードを作成  
  
「ID」, 「職業」, 「レベル」, 「獲得経験値」, 「各アクションの実行頻度」
  - 分類器学習用  
クラスタリング用の各レコードに後述の工程2の「タイプラベル」を追加
  - 参加者選定用  
選定対象のプレイヤーに対して, 分類器学習用と同じ形式のレコードを作成  
ただし, 「タイプラベル」は空に

## イベント参加者の選定(2/2)

- **工程:**
  1. クラスタ分析(第2部のK-means法など)によりクラスタリング用データを分割
  2. 分割ごとにタイプラベルを用意
  3. 分類器学習用データを用いて分類器を学習
  4. 参加者選定用データに対して学習済み分類器を適用
- **結果:**
  - 参加者選定用の各レコードにタイプラベルが決定
- **適用:**
  - イベントの成功レシピの基に選定されたプレイヤーに個別にイベントの開催情報を案内

03/09/04

立命館大学, ラック・ターウォンマツ

13

## ホットな話題の抽出(1/2)

- **仮説:**
  - チャットの会話は複数の連続な話題が混合
  - 話題は一連の単語で構成・抽象化
  - ホットな話題は時間と共に移り変わる
- **データ:**
  - チャットのストリームから次のような単語文書行列を作成

	文書1	文書2	文書3	...
バランス	1	0	0	...
fighter	1	0	0	...
スキル	1	0	1	...
イベント	0	1	0	...
参加	0	1	0	...
⋮	⋮	⋮	⋮	⋮

03/09/04

立命館大学, ラック・ターウォンマツ

14

## ホットな話題の抽出 (2/2)

- 工程:

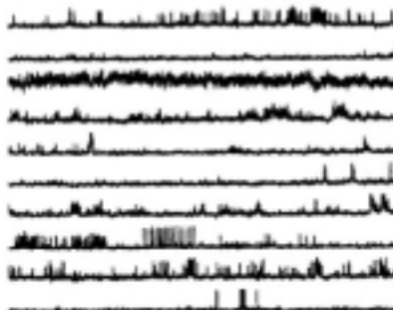
1. 単語文書行列に対して潜在的意味解析を適用
2. 1の結果に対して時間相関対応の独立成分分析手法(Complexity Pursuit法又は Thawonmas&Cichocki法-REF#2)を適用
3. 分離した各成分に関与した単語から話題を抽象化
4. 調査対象の話題を選定
5. 選定した話題と関連したチャット文を解析

- 適用:

- ギルドがもっとも好まれるパーティミッションの提供やバランスの改善などの解析結果に応じる

- 結果:

分離した各成分の時系列



## 第一部のまとめ

- 他の応用はまだある
- まず、仮説を立ててみよう
- 適用(とその評価)までやらないと意味がない

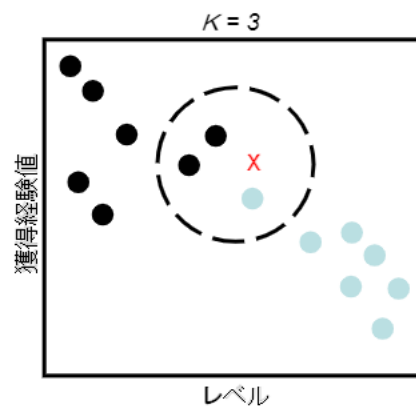


## 第二部

## 記憶ベース推論(MBR) (REF#3)

- 主な選択事項:

1. レコードの適切な集合の決定
2. レコードを表現するための最も効率的な方法の決定
3. 距離関数, 結合関数および近傍数 $K$ の決定



## MBRにおける距離の計算

### ■ 数値フィールドに対する一般的な距離関数:

- 距離の絶対値:  $|A - B|$
- 距離の2乗:  $(A - B)(A - B)$
- 標準化絶対値:  $|A - B| / \text{最大距離}$

### ■ 結合に関する一般的な結合関数:

- 合計:  $d_{\text{sum}}(A, B) = d_{\text{職業}}(A, B) + d_{\text{レベル}}(A, B) + d_{\text{経験値}}(A, B)$
- 標準化合計:  $d_{\text{norm}}(A, B) = d_{\text{sum}}(A, B) / \max(d_{\text{sum}})$
- ユークリッド距離:  $d_{\text{euclid}}(A, B) = \sqrt{d_{\text{職業}}(A, B)(A, B) + d_{\text{レベル}}(A, B)(A, B) + d_{\text{経験値}}(A, B)(A, B)}$

## MBRにおける結合関数

- $K$ 個の最近傍が答えに対して投票し, 多数決で決まる
- プレイヤーのミッション放棄のデータを考えよう

ID	職業	レベル	獲得経験値	放棄
1	fighter	27	190	no
2	wizard	51	640	yes
3	wizard	52	1050	yes
4	fighter	33	550	yes
5	wizard	45	450	no
6	fighter	45	1000	空

## MBRにおける多数決

プレイヤーID6がミッションを放棄するかどうかを判断

	近傍	近傍の放棄	K=1	K=2	K=3	K=4	K=5
dsum	4,3,5,2,1	Y,Y,N,Y,N	yes	yes	yes	yes	yes
deuclid	4,1,5,2,3	Y,N,N,Y,Y	yes	?	no	?	yes

信頼水準による放棄予測

	K=1	K=2	K=3	K=4	K=5
dsum	yes,100%	yes,100%	yes,67%	yes,75%	yes,60%
deuclid	yes,100%	yes,50%	no,67%	yes,50%	yes,60%

<注意点> Kの値が異なれば分類が変わる

## MBRの長所

- ・すぐに理解可能な結果
- ・あらゆるデータタイプに適応
- ・学習用のデータセットの維持が容易

## MBRの短所

- 分類の計算量が多い
- 大量な記憶領域が必要
- 結果が「距離関数」, 「結合関数」, 「近傍数」の選択に依存

## K-means法 (REF#3)

- 1967年、J.B.MacQueenによって発表
- 最も一般的なクラスタ分析の手法
  - レコードの類似性に注目してクラスタを形成する手法
  - 探索的な分析
- 多次元空間で適用可能

## K-means法のアルゴリズム

2次元平面の場合で

Step1: 形成したいクラス数を選択する(それがK-means法のKとなる)

Step2: K個の各クラスを「シード」として平面上に自由に配置する

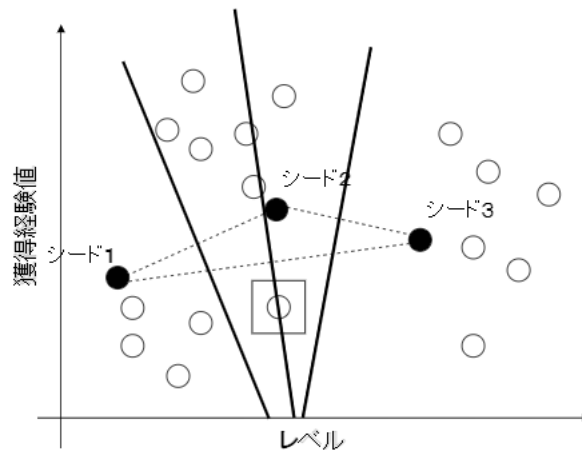
Step3: K個の中から2つのシードを選び、垂直二等分線を引く

Step4: 分割されたフィールドから、各シードに重心が最も近いレコードを割り当てる

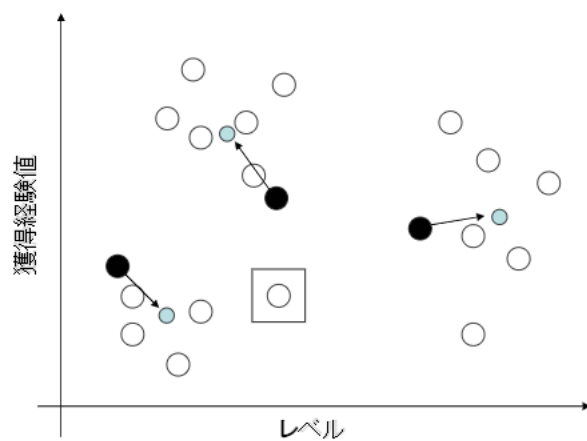
Step5: 割り当てられたレコードの重心を求め、各シードを移動させる

Step6: Step3~Step5を移動がなくなるまで繰り返す

## アルゴリズムのStep1からStep4まで



## アルゴリズムのStep5

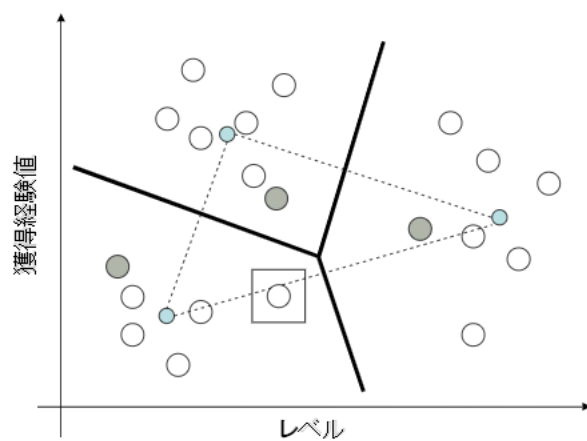


03/09/04

立命館大学, ラック・ターウォンマツ

27

## アルゴリズムのStep6



03/09/04

立命館大学, ラック・ターウォンマツ

28

## K-means 法の欠点

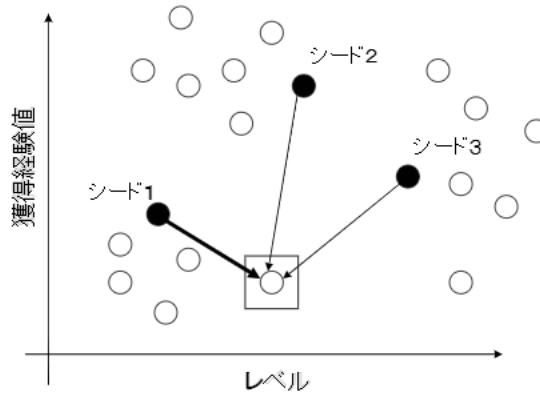
- クラスタが重複している場合、あまりよく機能しない
- 遠く離れたデータによって、クラスタの中心から簡単に引っ張られてしまう
- クラスタ内のレコードの重要性順位の概念がない

## 正規混合モデル

- K-means法に確率的概念を導入
- シードを正規分布の平均と見る
- 推奨ステップと最大化ステップ

## 推奨ステップ

- 各正規分布がそれぞれのデータの点に対して持つ“信頼度”を計算する



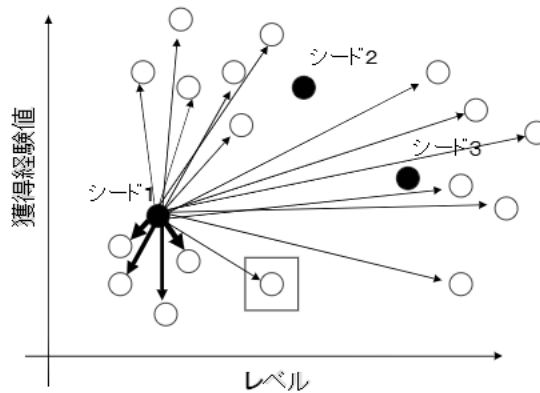
03/09/04

立命館大学, ラック・ターウォンマツ

31

## 最大化ステップ

- 各正規分布が“信頼度によってウエイト付け”され、データセット全体の重心に移動する



03/09/04

立命館大学, ラック・ターウォンマツ

32



## 第二部のまとめ

- タイプ分け済みのレコードを参考に解析対象のレコードを分類したい場合は、MBRなどの分類器を使用
- 決定木やニューラルネットワークといった分類器もある
- 解析対象のレコードを似通ったレコード同士(クラスタ)に分割したい場合は、K-means法などのクラスタ分析を使用
- Fuzzy c-meansや自己組織化マップ(SOM)といったクラスタ分析法もある

## 第三部

# 知能エンターテインメント研究室の紹介

- **MMOG研究:**
  - ゲームマイニング(行動解析, 話題抽出)
  - 感性工学
  - 教育用コンテンツ・システム開発(100~1000人規模)
- **Interactive Comedy Drama研究:**
  - アーキテクチャー, エージェント・プランニング
  - 笑い科学
  - UT2003エンジンを用いた学生実験の開発

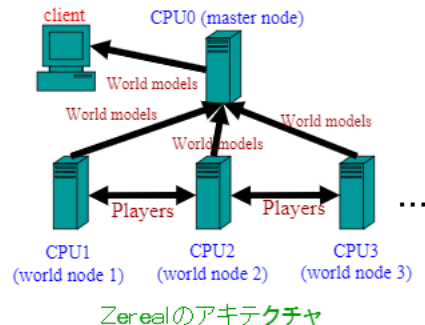
03/09/04

立命館大学, ラック・ターウォンマット

35

## 行動解析の研究成果

- **アプローチ:**
  - MMOGシミュレータ(Zereal, REF#4)からのデータで各手法を開発・検証してから実データへ
- **プレイヤーモデル(3つの性格):**
  - Killer
    - ・ モンスターへの攻撃を最優先  
→ 攻撃性
  - Markov Killer
    - ・ 状態遷移の確率にしたがって行動を選択  
→ 知的かつ攻撃性
  - Plan
    - ・ キーアイテム探しを最優先し,  
キーを見つけるとドアへ向かう。  
→ 探求性



03/09/04

立命館大学, ラック・ターウォンマット

36

## ゲームのスクリーンショット



03/09/04

立命館大学, ラック・ターウォンマット

37

## 以下のようなログからプレイヤーのタイプを特定

```
20|2003-5-28: 12:12:24|1000008|Walk|(18,7)|(17,6)|MarkovKiller
20|2003-5-28: 12:12:24|1000009|PickFood|(18,12)|(19,12)|MarkovKiller
20|2003-5-28: 12:12:24|1000010|Walk|(24,8)|(23,9)|MarkovKiller
20|2003-5-28: 12:12:24|1000011|Walk|(16,5)|(16,4)|MarkovKiller
20|2003-5-28: 12:12:24|1000025|Removed|(29,10)|(.)|Killer
20|2003-5-28: 12:12:24|1000013|Attack|(30,10)|(29,10)|Monster
20|2003-5-28: 12:12:24|1000016|Attack|(39,9)|(39,10)|Monster
20|2003-5-28: 12:12:24|1000018|Walk|(27,10)|(28,10)|PlanAgent
21|2003-5-28: 12:12:24|1000007|Walk|(32,11)|(33,12)|MarkovKiller
21|2003-5-28: 12:12:24|1000008|Walk|(17,6)|(16,5)|MarkovKiller
21|2003-5-28: 12:12:24|1000009|Walk|(19,12)|(18,11)|MarkovKiller
21|2003-5-28: 12:12:24|1000010|Walk|(23,9)|(23,10)|MarkovKiller
21|2003-5-28: 12:12:24|1000013|Walk|(30,10)|(29,10)|Monster
21|2003-5-28: 12:12:24|1000014|Walk|(31,10)|(30,10)|Monster
21|2003-5-28: 12:12:24|1000016|Attack|(39,9)|(39,10)|Monster
21|2003-5-28: 12:12:24|1000026|Attack|(39,10)|(39,9)|Killer
```

03/09/04

立命館大学, ラック・ターウォンマット

38

## 実行したアクションの頻度vs獲得したアイテムの頻度

### アクションの頻度

Agents	Walk	Attack	PickFood	PickPotion	PickKey	LeaveWorld
Plan1	162	0	12	0	10	4
Killer1	12	47	0	4	4	0
Killer2	48	135	0	4	0	0
Markov Killer1	128	20	28	9	0	0
Markov Killer2	97	43	15	14	2	0
Plan2	118	0	0	0	58	5

### アイテムの頻度

Agents	Monster	Food	Potion	Key	Door
Plan1	0	12	0	10	4
Killer1	16	0	4	4	0
Killer2	30	0	4	0	0
Markov Killer1	5	28	9	0	0
Markov Killer2	10	15	14	2	0
Plan2	0	0	0	58	5

03/09/04

立命館大学, ラック・ターウォンマツ

39

## 実験結果

75体に対してleave-one-out実験法によるMBR(K=1)の認識率の表

入力	ノイズ ほぼなし	ノイズ あり	ノイズ やや多い
行動情報	97%	85%	69%
アイテム情報	96%	94%	79%

03/09/04

立命館大学, ラック・ターウォンマツ

40

## 第三部のまとめ

- エンタテインメントを研究対象
- 成果が出はじめている
- 共同研究相手を募集中

## 全体のまとめ

- データマイニングの活用 = より良いサービスの提供
  - 本公演をヒントにデータマイニングの更なる応用を目指そう
  - この後の議論の場は
    - <http://www.egroups.co.jp/group/gamemining>
- または
- [ruck@cs.ritsumeai.ac.jp](mailto:ruck@cs.ritsumeai.ac.jp)

# REF

- #1: David Kennerly, Better Game Design through Data Mining  
[http://www.gamasutra.com/features/20030815/kennerly\\_01.shtml](http://www.gamasutra.com/features/20030815/kennerly_01.shtml)
- #2: Ruck THAWONMAS, ゲームマイニングの野望: チャットルーム・掲示板から話題を特定する技術  
[http://www.ice.ritsumei.ac.jp/~ruck/PAP/RT\\_EC03.pdf](http://www.ice.ritsumei.ac.jp/~ruck/PAP/RT_EC03.pdf)
- #3: 『データマイニング手法』  
マイケルJ.Aベリー, ゴードン・リノフ著 SASインスティテュート ジャパン, 江原淳, 佐藤栄作共訳, ISBN4-303-73430-6 C3004, 海文堂出版
- #4: Zereal.  
<http://abiody.com/game/mining/software/zereal/>