

統計学を応用する研究室

- 見えないものを見通せるようにする
- 未来を予測する
- ノイズの多い中から機微を捉える



今年度の卒業研究テーマ：

直接は観測できないものの予測

直接は調査できないものの調査

トレンドモデルの状態空間表現より、カルマンフィルタを利用して使用量の予測を行う。

予測する使用量のデータには明らかに増加のトレンドがあるため、定常性を仮定する時系列モデルをそのまま適用することはできない。そのため、トレンドモデルの状態空間を使って、カルマンフィルタにより予測を行う。

トレンドモデル
 $y_n = \mu_n + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma_\varepsilon^2), \quad n = 1, \dots, t$
 $\mu_n = \mu_{n-1} + \beta_{n-1}$
 $\beta_n = \beta_{n-1} + \zeta_n, \quad \zeta_n \sim N(0, \sigma_\zeta^2)$

状態空間モデル
 $x_n = Fx_{n-1} + Gv_n$
 $y_n = Hx_n + w_n$

カルマンフィルタ

1期先予測
 $x_{n|n-1} = Fx_{n-1|n-1}$
 $V_{n|n-1} = FV_{n-1|n-1}F' + GQ_nG'$

フィルタ
 $K_n = V_{n|n-1}H'(HV_{n|n-1}H' + R_n)^{-1}$
 $x_{n|n} = x_{n|n-1} + K_n(y_n - Hx_{n|n-1})$
 $V_{n|n} = (I - K_nH)V_{n|n-1}$

フィルタと予測

5期先予測と実測値の比較

アクセシブルを用いた調査の質評価に関する研究

目的
アクセシブルを用いた調査が従来調査の代替となる可能性を示す。

群1(z=1): アクセシブル(から)の集団
群2(z=0): 無作為抽出による標本集団

	群1(無回答)	群2(有回答)
群1の結果変数の値	アクセシブルからの調査結果 (Y_1)	欠測
群2の結果変数の値	欠測	無作為抽出集団からの調査結果 (Y_2)

共変量
違いを生んでいる共変量を見つけ

共変量の分布が異なる
→「傾向スコア」を用いて二群の分布を共通にする

1. 傾向スコアをロジスティック回帰モデルで推定

$$e(X) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)]}$$

2. IPW推定量で結果変数を調整

$$\hat{E}(Y_1) = \frac{1}{N} \sum_{i=1}^N z_i \times \frac{Y_{1i}}{e_i}$$

$$\hat{E}(Y_2) = \frac{1}{N} \sum_{i=1}^N (1 - z_i) \times \frac{Y_{2i}}{1 - e_i}$$

3. 因果効果の計算

$$\hat{d}_{12} = \hat{E}(Y_1) - \hat{E}(Y_2)$$

評価

ニーズの把握、ブランド評価、商品やサービスの評価、商品やサービスの使用実態などなど...

目録母集団
WEB調査の対象範囲

バイアス1: 偏りのある抽出
バイアス2: サンプル調査
バイアス3-4: 無回答偏差、調査モード偏差

調査母集団
回答者

アクセシブルを用いた調査

ネットのつぶやきからの予測

ジェスチャーでデータマイニング

KeyGraphを利用したソーシャルネットワークの解析

KeyGraphとは、文章を解析し、キーワード抽出によってそこから文章の構成を判別するアルゴリズム。
各単語の出現頻度、互いのつながりを計算し、解析した文章の中から重要な単語を見つけ出し、ネットワーク図によって視覚化する。

本研究ではKeyGraphにより、Twitterのようなマイクロブログを解析する。KeyGraphでのマイクロブログの解析はまだ前例がなく、またKeyGraphには時間経過の概念がなく、KeyGraphに時間による解析の機能を付加し、マイクロブログにおける時間経過による話題の発展性を観測したい。

遠隔操作ウイルス犯罪予告事件のKeyGraphによる解析

グラフの読み方
 水色のノードと実線がつながっている単語が解析した文章において、よく出てくる典型的な文章のつながり。
 そこから点線がつながっている部分は、そのような典型的な文章が現れたときに、その文章からその単語へ話題が進展するという可能性を示している。
 特に赤色のノードは水色ノードからの話題の発展性が高く、注目すべき単語である。

エヴァンゲリオンQの感想の傾向をKeyGraphによって解析

画面上のキューブ(多次元データベース)に対してのジェスチャーのみでOLAP分析操作を行えるようにする。

期間ディメンション選択後、うでを広げる動作
ドリルダウン

全製品	2000年1月...	2002年12月
東日本	18...	36
西日本	10...	20
海外	8...	16

製品ディメンション選択後、うでを振り下ろす動作
スライシング

製品A	2000年	2001年	2002年
東日本	80	150	200
西日本	20	40	100
海外	0	10	100

地域ディメンション、製品ディメンション選択後、うでをクロスさせる動作
ダイシング

全製品	2000年	2001年	2002年
東日本	120	240	360
西日本	50	100	200
海外	10	40	160

全地域	2000年	2001年	2002年
製品A	100	200	300
製品B	40	100	250
製品C	40	80	170

面談随時：アポイントメントはメールにて (mailto:watalu@inf.uec.ac.jp)

学生部屋の訪問歓迎：月・火・水の午後は誰かが構ってくれます