

条件付層別差分モデルによる需要予測の高精度化

～ ベイジアンネットによる連続値推定補正手法 ～

Improvement of Demand Prediction using Conditional Sliced Residual Model

本村陽一*1
Yoichi Motomura竹中毅*1
Takeshi Takenaka石垣司*1*2
Tsukasa Ishigaki*1 産業技術総合研究所サービス工学研究センター
National Institute of Advanced Industrial Science and Technology
Center for Service Research*2 東北大学大学院経済学研究科
Tohoku University
Graduate school of Economics and management

A conditional sliced residual model consisted from Bayesian networks are introduced in this paper. The models can be utilized to improve demand prediction in service engineering. We evaluate the improvement performance with real POS data.

1. はじめに

サービス現場の生産性を向上するためにあらかじめサービスの需要を予測することで、提供するリソースを最適化する。ここで、サービスの質を高めながら余剰コストを削減するために需要予測技術の重要性が大きい。本稿では、来店人数などの需要に対する予測外れの理由に注目した構造を複数のベイジアンネットモデルでモデル化して構成した差分モデルを任意のベースモデルと組みあわせることにより、これまでベイジアンネット単独では推定精度を上げることが難しかった連続値を予測する問題に対して予測精度を向上させる方法を提案する。

2. 条件付層別差分モデル

需要予測の方法には様々なものがあり、多くの研究もあるが、その多くが誤差分布には正規分布などの対称性のある分布を仮定している。しかし、実際の人の来店行動を見ると来店する理由(例えば休日)と来店しない理由(例えば降雨量)は異なることが多く、結果として、来店人数の上ブレと下ブレは非対称である場合がある。そこで、この上ブレと下ブレの原因となる構造に注目し、その差分を表すモデルを考え、データから学習する。またブレの構造が異なる要因から生じているにも関わらず、そのデータを混合し、単一の分布としてモデル化することは予測精度の悪化を招く。そこで、上ブレ、下ブレの因果的構造を独立に反映するために、任意の予測(z)に対する差分を外れの大きさを5段階(-2σ,+1σ,0,-1σ,+2σ)で層別し、そのうち上ブレ下ブレ計4状態についてベイジアンネットワークの構造学習によって、説明変数を探索し、各段階のブレ幅が生じる条件付確率としてモデル化する。σは真の値(y)と任意の予測値(z)との標本分散である。5段階の複数の異なる差分モデルとして構築した4つのベイジアンネットP(+2σ|x), P(+1σ|x), P(-σ|x), P(-2σ|x)を最終的に以下の式(1)のように統合して差分モデルdiff(x)を構成する。これをここでは条件付層別差分モデルと呼ぶ。

$$y - z = \text{diff}(x) \cong \sum_{k=1,2} \xi(x) k P(\xi(x) | k \sigma | x) \quad (1)$$

ここで例えばxを日、予測対象を来店人数とし、ξ(x)は上ブレ(y-z>σ)の時+1, 下ブレ()の時-1 それ以外で0をとる符号関数とする。その日に上ブレするか下ブレによりξ(x)が+1,-1,0の

いずれかをとることで、条件付層別差分モデルの値は上ブレか下ブレのいずれか片側についての期待値をとることになる。また各ベイジアンネットの親ノードにはxに依存するその日の属性(例えば平日かどうか、降雨量の大小など)が代入される。

この各ベイジアンネットを構築する際はその期間中の来店人数yとベースモデルの予測人数z、その日xの属性から構成されるデータを用いて、ベイジアンネットのモデル構築法[1]に基づきグラフ構造と条件付確率表を決定する。符号関数ξ(x)も同様に、上ブレ(y>z+σ)、下ブレ(y<z-σ)、ベースモデルの予測範囲内(z-σ<y<z+σ)の3状態を目的変数としたベイジアンネットとして学習する。

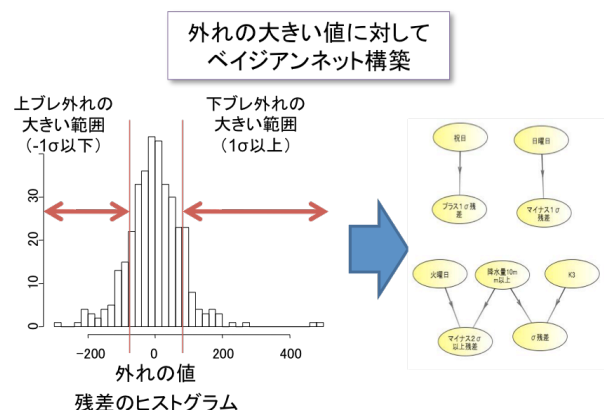


図1 条件付層別差分モデルで用いるベイジアンネット

3. 需要予測の高精度化

任意の予測方法による推定値をzとして、さらに条件付層別差分モデルによる補正をかけた推定値 \tilde{y} は以下ようになる。

$$\tilde{y} = z + \text{diff}(x) \quad (2)$$

を行うことができる。また、顧客セグメント毎にデータを層別して統合した混合モデル[3]に対しても、各顧客セグメントi毎のモデルの推定値に対して同様にセグメントi毎の条件付層別差分モデルを構築して、

$$\tilde{y} = \sum_{i \in \text{segment}} (z_i + \text{diff}_i(x)) \quad (3)$$

として予測を行うことができる。

4. 実験評価

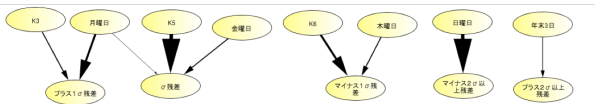
以上で提案した条件付層別差分モデルによる需要予測を実際のサービス現場で得られた大規模データに適用することで実験評価を行う。

使用したデータは関西地区の食料品スーパーのある一店舗に対する購買履歴(POS)データから購入人数を抽出したものにコーザルデータを追加したデータ[2]で、i) 訓練データとして2009年9月1日～2010年8月31日の12カ月分、ii) テストデータとして2010年9月1日～2010年9月30日の1カ月分を使用した。予測精度の評価のためにはサービス現場でよく用いられている前年同月同曜日との比較を行うこととして、2009年9月2日～2010年10月1日のデータを比較対象に用いた。

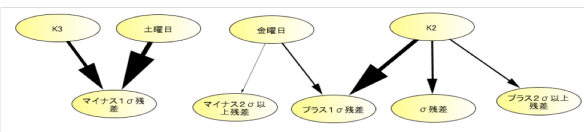
4.1 条件付層別差分モデルの構築

ベースモデルとしては線形モデル[2]、及びそれらをライフスタイルに基づく顧客セグメント毎に分割して統合する方法[3]を用い、それに対する差分モデルを先の方法に従って構築した。全体、及び顧客セグメント毎に構築したベイジアンネットワークモデルは以下ようになる。下ブレ上ブレの構造がそれぞれ大きく異なっており、条件付層別差分モデルの有効性が示唆される。

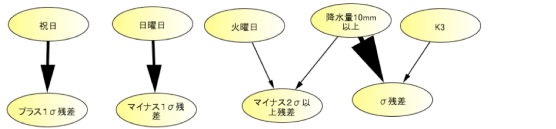
全体モデル



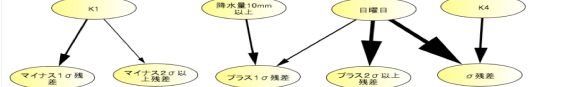
アクティブ消費派来店人数



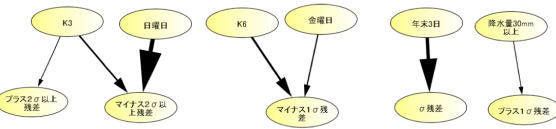
家庭生活充実派来店人数



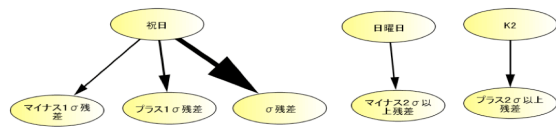
堅実生活派来店人数



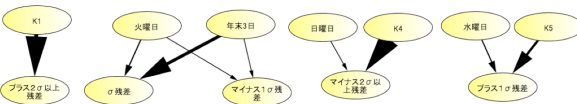
こだわり消費派来店人数



パパッと消費派来店人数



節約消費派来店人数



4.2 来店人数の予測精度の評価

1ヶ月分のテストデータを用いて、比較対象である前年同月同曜日(A)、線形モデルによる来店人数予測(B)、先に構築した全体の条件付層別差分モデルによる補正(C)、顧客セグメント毎の線形モデル(D)、顧客セグメント毎の条件付層別差分モデルによる補正(E)を評価した所、正負を考慮した残差の合計、及び残差の絶対値の合計の結果は表1ようになった。

予測手法	残差の合計(人)	絶対残差の合計(人)
(A)前年同月同曜日	-2835	3285
(B)線形モデル予測	-762	1599
(C)条件付層別差分	-555	1549
(D)顧客セグメント毎予測	-516	1542
(E)セグメント毎条件付層別差分	-141	1501

表1 来店人数予測精度の評価

この結果、提案した条件付層別差分モデルによる補正を全体モデルに適用したものの(C)は比較対象である(A)と補正前の(B)のいずれに対して、絶対誤差でそれぞれ 1736 人、50 人分減少しており、予測精度が向上している。また顧客セグメント毎に層別した場合でも、補正前の(D)と比べて条件付層別差分モデルによる補正により 41 人分誤差が減少している。

5. おわりに

本稿では、予測の外れ方の構造に注目して、ベイジアンネットワークを複数用いた条件付層別差分モデルを提案し、これまでベイジアンネットワーク単独では取り扱いが難しかった連続量に対する予測問題としてスーパーの来店人数予測において予測精度の向上を確認した。ここで提案したモデルは任意の予測モデルと組みあわせた補正効果により精度を向上させることができ、また需要予測に限らず幅広い問題にも適用できる手法になっている。

また、サービス工学への応用として見た場合には、単に需要を予測するだけでなく、本研究で示したような予測が外れる構造に注目し、ベイジアンネットワークとして得られる説明変数を解釈することで、下ブレ(上ブレ)がありえる時、その際リスク、及びその際のコストを意志決定者に示すことができ、実用上強力な手段になる。商材の仕入れや様々な施策などのコストは来店人数に比例して設定され、これが過剰である場合も不足である場合のいずれにおいても損失が発生することから、需要予測の精度の向上や、外れる場合の要因、外れ方が把握できることで、実際のサービス現場における損失を低減できると期待できる。

謝辞

本研究は経済産業省委託費事業「サービスと IT の融合による新市場創出事業(サービス工学研究事業)」の支援を受けた。

参考文献

- [1] 本村陽一, 岩崎弘利: ベイジアンネットワーク技術, 東京電機大学出版局, 2006.
- [2] 竹中毅, 石垣司, 本村陽一: “生活者行動に着目したサービス需要予測技術の検討,” 人工知能学会全国大会, 2011.
- [3] 石垣司, 竹中毅, 本村陽一: “潜在クラスモデルによる流通量販店舗の来店人数予測の精度改善の評価,” 人工知能学会全国大会, 2011.