

ベイジアンネットワークを用いた 食事メニュー推奨システムの提案

長谷川 亮・寺田 吉壺・宿久 洋

本論文では食MAP データを用いて、消費者向けに毎日の食卓のメニューを推薦するシステムの枠組みについて提案する。ここで、推薦するメニューを決定する手法としてベイジアンネットワークモデルを用いた。提案するシステムでは意識調査アンケートを元に分類した利用者のグループそれぞれについてベイジアンネットワークモデルを構築し、毎日の利用者の意向に応じて適切なメニューを出力する。さらに、ベイジアンネットワークモデルについて交差検証法を行い、モデルの精度を算出した。

1. はじめに

一家の台所を預かる主婦にとって毎日の献立を考えることは、往々にして夫や子供が予想する以上に大変な作業である。冷蔵庫の食材、栄養バランスに加えて、家計状況、家族の好物なども考慮して、なおかつ飽きがこないように毎日の献立を構成しなければならない。このことから、その日の献立のヒントを提示することは主婦にとって大きな助けになると考えられる。実際、主婦への献立の提案は、TVの料理番組やグルメ雑誌、スーパーの広告などで行われているが、それらはどれも提示する側が献立を決定して伝えているものである。こういった献立の提案は、個々の主婦の事情に合わせて献立を推薦しているわけではないので、必ずしも主婦の要望に合致するとは限らない。例えば、鶏肉を使ったメニューが番組や広告で推薦されていたとしても、夫が鶏肉嫌いである家庭の主婦にとっては、そのメニューは参考にならないであろう。

本論文では献立の推薦について、食MAPデータ(齋藤, 2003)を元にした新しいアプローチを試みる。食MAPデータを用いた研究は、現在までに開発元の株式会社NTTデータライフスケープマーケティングを中心に数多く行われている(寺本, 2008; 矢野他, 2010)。しかしその多くは、食卓におけるメニュー調査による知見の発見や、それを利用した食品メーカーや小売店向けの販売戦略

の提示を主眼に置いた研究がほとんどである。そこで本論文では販売者向けではなく、消費者である主婦に対して献立の推薦を行うシステムを考える。

献立の推薦には、献立の構成そのもの(主食、副菜、汁物などのメニューの組み合わせ)の推薦を行う場合と、献立の構成要素であるメニューを推薦する場合の2種類がある。このうち、本論文ではメニューを推薦するシステムを考える。具体的には、食MAPデータの情報を生かし、食MAPデータに格納されている多くのメニューの中から主婦の意向(その日の主婦の気分や方針)に沿ったメニューを提示するシステムを提案する。例えば、「低カロリーで、簡単に作れる肉料理」を作りたいと考えている主婦に対して、その意向に沿ったメニューの例を提示することを目標とする。

ここで主婦の意向から推薦するメニューを導き出す手法として、ベイジアンネットワークを用いる(繁樹他, 2006)。ベイジアンネットワークは事象の因果関係をグラフ構造で表現する確率モデルの1つである。その強みとして、得られたデータから事象の条件付き確率とグラフ構造を学習し、確率推論を行えることが挙げられる。本論文ではベイジアンネットワークの確率推論を利用して、主婦の気分(以下、食MAPデータにおける呼び名にならない、食卓動機と呼ぶ)や大まかなメニュー方針(肉料理、魚料理など)を入力とし、具体的なメニューを出力とする食事メニュー推薦システム

表1 食MAPデータの構成

食卓に関するデータ	食卓メニューデータ (図1) 食卓イベントデータ (使用しない) 食卓動機イベントデータ (表2)
モニター家庭の属性に関するデータ	モニター情報データ (使用しない) モニター意識調査データ (表4)

表2 食卓動機イベントデータ

食卓動機	食卓動機 1. 主人の好きな料理を出した 食卓動機 2. 子供の好きな料理を出した : 食卓動機17. 家族みんなで楽しめる食卓にした
食卓イベント	食卓イベント 1. 客を夕食に招いた 食卓イベント 2. 昼食、お茶の時間に客を招いた : 食卓イベント10. 主人が家で夕食を食べなかった

の提案を行う。

2. 利用データ

本論文で利用したデータは、平成20年度データ解析コンペティションにおいて提供された食MAPデータである。食MAPとは株式会社NTTデータライフスケープマーケティングが開発した食卓マーケティング情報システムである。今回提供されたデータは表1のように、大別して食卓に関するデータとモニター家庭の属性に関するデータに分類される。

食卓に関するデータは東京首都圏30km圏内に在住の194のモニター家庭について、2006年1月1日から同年12月31日までの1年間、毎食ごとに記録されたものである。食卓データは3種類のデータからなり、朝昼夕の各食事機会ごとに、メニューとメニューに使われた食材の名称を記録した「食卓メニューデータ」と祝日の情報や食卓参加人数などを記録した「食卓イベントデータ」、その日の主婦の意識と家族の参加状況などを記録した「食卓動機イベントデータ」で構成される。

「食卓動機イベントデータ」は表2のように、調理の際にこだわったかどうかや家族の好物を出したかどうかなどの食卓動機に関する17の質問項目と、来客の有無やその日の忙しさなど、食卓イベントに関する10の質問項目からなる。

「食卓メニューデータ」には976種別のメニューと1931種別の食材が8桁のコードで格納され、そのレコード数は194家庭を合計して約180万である。メニューコードと食材コードは、図1のように左

2桁が大分類を表し、左5桁が中分類を表している。大分類は米類、魚料理などの59の 카테고리からなり、中分類はさらに細かく井もの、焼き魚など300の 카테고리からなる。

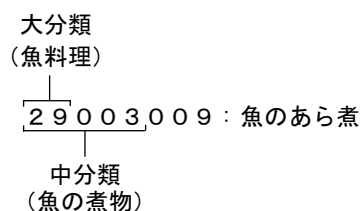


図1 メニューコードの大分類、中分類の例

これに加えて、農林水産省の食事バランスガイド(農林水産省, 2010)に従い、各メニューを主食、主菜、副菜など8つのカテゴリーに分類した(表3)。さらに、メニューの名称を参考に、メニューのスタイルを和風、洋風、中華風、その他(エスニック

表3 食事バランスガイド分類基準

主食	ご飯、麺類などを主原料とするメニュー
主菜	肉や魚を主原料とするメニュー
副菜	野菜や海藻などを主原料とするメニュー
牛乳・乳製品	牛乳、チーズなどを主原料とするメニュー
果物	果物など主原料とするメニュー
水・お茶	水分の摂取を目的とするメニュー
菓子・嗜好飲料	菓子や酒類などからなるメニュー
その他	上記以外

ク風など)の4つに分類する基準を設けた。

モニター家庭の属性に関するデータは、家族構成や年収などの家庭の属性データを記録した「モニター情報データ」と、主婦の意識調査アンケートの結果を記録した「モニター意識調査データ」の2つからなる。「モニター意識調査データ」に記録されている意識調査アンケートは表4のように、食生活全般に対する意識や食品の購入で気をつけていること、メニュー作りで心がけていることなどの食生活意識に関する69の質問項目と、商品の購入頻度や購入店舗などの商品の購入形態に関する77の質問項目、普段の食事時刻や休日に関する32の質問項目からなり合計178の質問項目で構成されている。「食卓イベントデータ」、「モニター情報データ」については、本論文では使用しない。

3. モニターの分類

ここでは、ベイジアンネットワークを構築する前準備として「モニター意識調査データ」を元にモニター家庭(以下、モニターと呼ぶ)の分類を行う。まず、分類の目的について述べ、次に食生活意識に関する質問項目の分類を行う手順について述べる。次に分類した質問項目を用いてモニターをグループ分けする流れについて述べる。

3.1. 分類の目的

ここで、料理に対する姿勢はモニター全体について共通ではなく、モニターはいくつかのグループ(セグメント)に分類できると仮定する。マーケティングの分野においても、消費者の嗜好の多様化や効率的経営の観点から、消費者の分類は重要な課題となっている。マーケティングにおいて消費者の分類に用いられる代表的な要因として、「地域的要因」、「デモグラフィック要因」、「心理的要因」、「消費行動要因」などが挙げられる(Hooley and Saunders, 1993;片平, 1987)。ここで、モニターの分類には家計状況や出身地などによる分類も可能だが、食卓におけるメニューの選択には、年収や職業などのデモグラフィック要因や出身地などの地域的要因よりも、個人の性格や考えなどの心理的要因の方がメニュー選択に直接に関連すると仮定し、心理的要因を用いて分類を行う。

また、性格などの心理的な特性の異なるグループ間では、同じ食卓動機に対する反応も異なると

表4 モニター意識調査データ

食生活意識	食生活全般に対する意識.01. 料理を作るのは好きだ
	⋮
	食品の購入について.01. 生鮮食品の特売があるとよく買う
商品の購入形態	⋮
	メニューづくりで心がけている事.01. 栄養バランス
	⋮
食事時刻と休日	商品購入頻度-ア. 購入頻度(鮮魚)
	⋮
	商品購入の店舗形態.01. 鮮魚-スーパー
食事時刻と休日	⋮
	一番よく利用する店舗.01. 一番よく利用する店舗(鮮魚)
	⋮
食事時刻と休日	食卓状況.01. 平日の朝食状況
	⋮
	食事の支度平均時間.01. 平日の朝食支度時間
食事時刻と休日	⋮
	本人の休日.01. 本人の休日(月)
	⋮

考えられる。例えば、日常的に料理に手間を惜しまないグループに属するモニターが、今日は少し手を抜きたいと思って作ったメニューと、日常的に料理は面倒だと感じているグループに属するモニターが、手を抜きたいと思って作ったメニューは当然異なってくるだろう。このことから食MAPデータのうち、心理的要因に関連する質問項目からモニターの性格や考えを推測し、それに応じて分類を行うことを考える。以下ではまず、質問項目の分類について述べる。

3.2. 質問項目の分類

モニターの分類には、前節で述べた「モニター意識調査データ」のうち、「心理的要因」に関係する、食生活意識に関する69の質問項目を用いる。ただし、このままでは質問項目が多すぎるため、まずは質問項目の分類を行い、質問項目同士の解釈可能な群を作成する。

最初に、質問項目間の非類似度行列を作成する。

異なる2つの質問項目についてクロス表を作成し、質問項目間の関連度指標を求める。本論文では、関連度指標の中で代表的なものとしてクラメールの連関係数Vを用いた。クラメールの連関係数は2つの質問項目の関連度を0から1までの値で表す指標であり、1に近いほど2つの質問項目の関連度が高いことを示している。ここで、質問項目*i*と質問項目*j*の連関係数 V_{ij} を用いて、非類似度行列 $D = (d_{ij})$ を作成する。ここで、

$$d_{ij} = \begin{cases} \frac{1}{V_{ij}} & (i \neq j) \\ 0 & (i = j) \end{cases} \quad (i, j = 1, 2, \dots, 69) \quad (1)$$

とした。この非類似度行列Dを用いて、Ward法(齋藤・宿久, 2006)による階層的クラスタリングを行う。クラスタリングの結果、質問項目は大きく以下の5つの群に分類された。

- (1) 新しい料理作りに関する意識や料理への積極性など、料理好きに関する質問項目群
- (2) 栄養バランスや食品の安全性など、健康指向に関する質問項目群
- (3) 料理に対して面倒くさいと感じる意識など、料理への億劫感に関する質問項目群
- (4) 外食に関する考えなど、外食に関する質問項目群
- (5) 食品の特売やチラシの参考度など、節約志向に関する質問項目群

3.3. モニターの分類と特徴付け

次に、質問項目の分類結果を踏まえてモニターの分類を行う。今回提供された食MAPデータには、個別の食材のPOSデータは付随していない。このため同じ食材であっても、高級なものか、特売などの安売りのものかが判断できない。また、本論文では家庭で調理されるメニューに注目している。このため、(4)、(5)については考慮しない。このことから、上記の質問項目群のうち、(1)、(2)、(3)を用いることにする。以下、簡単のためこれらの質問項目群をそれぞれ、「料理好き」、「健康指向」、「億劫感」と呼ぶことにする。これら3つの質問項目群をモニターの性格を特徴付ける基準と考える。その上で、3つの基準についてモニターがどの程度肯定的か、否定的かを判別する。「とてもそう思う」、「ある程度そう思う」、というように、ある1つの基準についてモニターを段階的に分割することが考えられるが、本論文では解釈

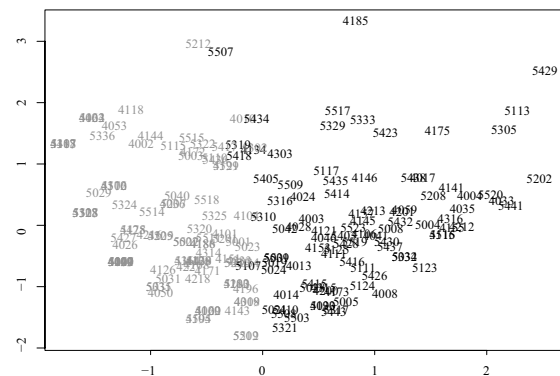


図2 「料理好き」に関するモニターの2分割

の容易さから「肯定的」、「否定的」の2値にモニターを分割する。

モニターの分割のため、3つの質問項目群それぞれについて質問項目群に含まれる質問項目を用いて、*k*-means法(齋藤・宿久, 2006)による非階層的クラスタリングを行い、モニターを2群に分ける。さらに質問項目群について数量化III類(駒澤, 1982)を行い、その結果と*k*-means法のクラスタリング結果をあわせて、3つの質問項目群それぞれについて2分割されたモニターが、該当する質問項目について肯定的か、否定的かの特徴付けを行う。図2は「料理好き」に関するモニター(ID化されている)の分類結果である。数量化III類で散布図に布置したモニターについて、「料理好き」について*k*-means法により2群に分けた結果を色分けして表している。

これにより、モニターは「料理好き」、「健康指向」、「億劫感」のそれぞれについて、肯定的モニターと否定的モニターの2つに分類される。例えば、「料理が好き (YES)」かつ「健康志向である (YES)」かつ「億劫感は感じない (NO)」グループ、というように性格を表す3つの属性についてYES、NOで答える形で各モニターの特徴付けが行われる。その組み合わせの数は $2^3 = 8$ 個となるので、モニターは8つのグループに分けられることになる。これら8つのグループをそれぞれtype 1、type2、...、type8と呼ぶことにする。表5には8つのtypeと「料理好き」、「健康指向」、「億劫感」に関する分類、及び分類されたモニター数と質問傾向の回答傾向からその特徴を記した。

表5 モニターの8つのtypeへの分類

type	料理好き	健康指向	億劫感	モニター数	特徴
1	YES	YES	YES	23	料理好きで健康やカロリーへの関心も高いが、できれば簡単なメニューを増やして手間を省きたいと考えている
2	YES	YES	NO	47	料理に対して積極的で、栄養バランスにも配慮する
3	YES	NO	YES	16	料理そのものは好きでグルメ雑誌などへの関心も高いが、忙しく手間をかけられず、偏ったメニューになりがち
4	YES	NO	NO	12	グルメ志向で料理を楽しんで作るが、あまり健康食品などは使用しない
5	NO	YES	YES	20	料理は好きではなく、家で作るよりも外食に出かけるタイプだが、偏食は極力避ける傾向にある
6	NO	YES	NO	26	料理することへの積極性が高く、栄養バランスにも気を配るが、自分は料理が下手であると思っている
7	NO	NO	YES	38	料理に対して消極的で、健康にも大きな関心はない
8	NO	NO	NO	12	下ごしらえや片付けに億劫感を覚えることはないが、料理そのものや栄養バランスには疎い傾向がある

4. ベイジアンネットワークモデルの構築

ここでは、推薦メニューの決定のために用いるベイジアンネットワークモデルの構築について述べる。ネットワークモデルの構築には「食卓メニューデータ」と「食卓動機イベントデータ」を用い、前節で分類したtypeごとに、そのtypeに属するモニターのデータを用いてネットワークモデルを構築する。本論文では、その日の利用者の意向に沿ったメニューの推薦を目標としているため、ネットワークモデルの構造は個々のメニューを子ノード、利用者の意向を親ノードとする。親ノードとしてモデルに組み込む変数は、「食卓動機イベントデータ」の質問項目のうち、食卓動機に関する質問項目から選択する。加えて、「食卓メニューデータ」からメニューに関するいくつかの分類基準を親ノードとして選択して組み込む。メニューの分類基準としては、図1に記したようなメニューコードの大分類、中分類を利用する他、「主食」、「副菜」などの分類や、「和食」、「洋食」などの分類を用いることが考えられる。親ノードと子ノードのリンク構造は探索的に求められることが多いが、本論文では計算量の観点から、親ノード間のリンクを考慮しない単結合なモデルを仮定している。

ネットワークモデルの構造を図で表すと、図3のようになる。図3の上部は「食卓動機イベントデータ」からのノードを表し、図3の下部は「食

卓メニューデータ」からのノードを表している。図中の矢印はノード間の因果関係を親から子への有向リンクで表している。

ここで、親ノードとして採用した食卓動機に関する n 個の質問項目を (X_1, X_2, \dots, X^n) と表わし、メニューに関する m 個の分類基準を $(X^{n+1}, X^{n+2}, \dots, X^{n+m})$ と表わす。これらの親ノードを確率変数と考えると、全ての親ノードの状態は式(2)のように確率変数ベクトル X として表すことができる。

$$X = (X_1, X_2, \dots, X_n, X_{n+1}, X_{n+2}, \dots, X_{n+m}) \quad (2)$$

また、メニュー名称を値域とする確率変数を Y とする。ここで、親ノードの状態値(実現値)ベクトル $x = (x_1, x_2, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+m})$ が与えられたとき、あるメニュー y_i ($i = 1, 2, \dots, K$) が食卓に出現する確率は、条件付き確率の形で

$$P(Y = y_i \mid X = x) \quad (3)$$

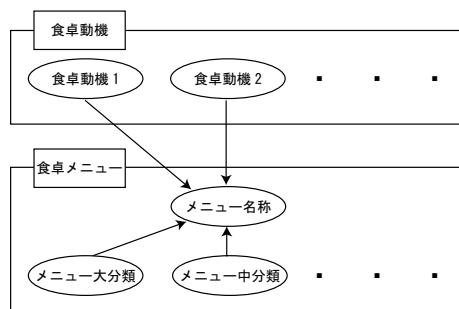


図3 構築するベイジアンネットワークモデルの構造

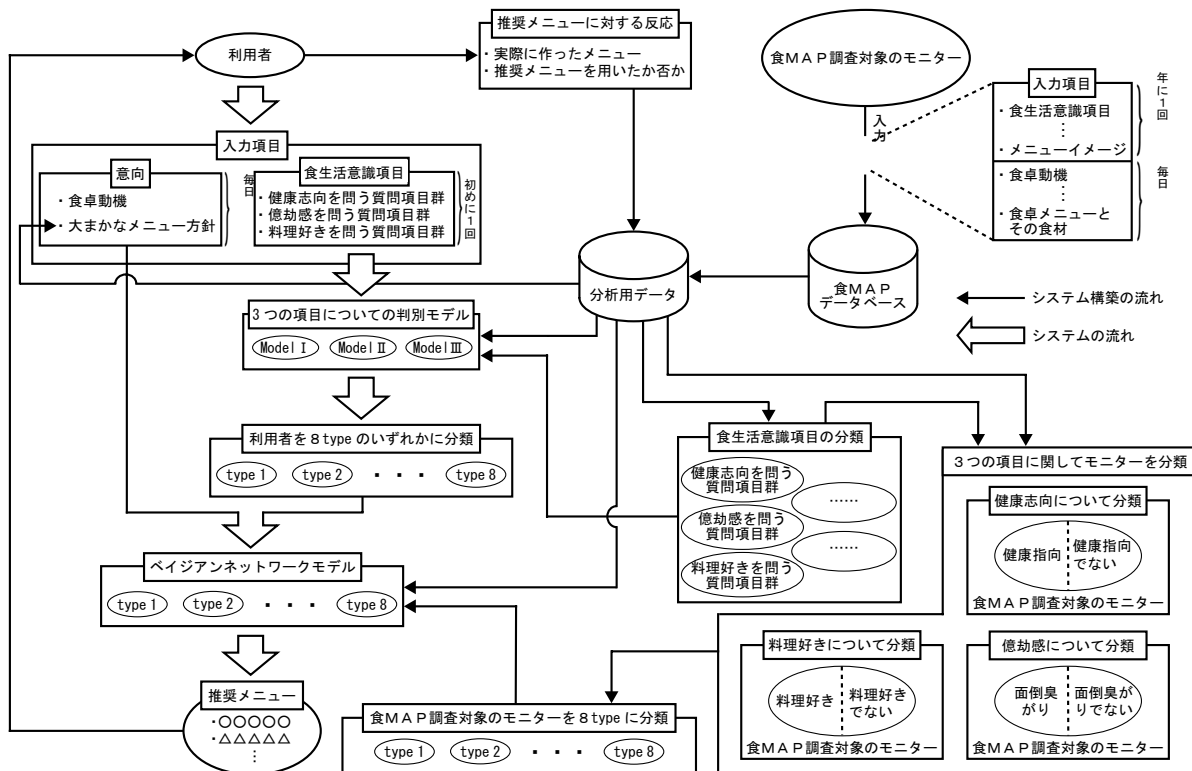


図4 システムの流れ

と表すことができる。ベイジアンネットワークモデルを構築する目的は、親ノードと子ノードの因果関係に従ってこの条件付き確率を求めることである。

ここで、ある親ノード X_j のとりうる状態値の数を N_j とすると、 X のとる全ての状態値の組み合わせ数は、 $L = \prod_{j=1}^{n+m} N_j$ 個となる。このときメニューのCPT(conditional probability table, 条件付き確率表)はメニューの数 K を行にとり、親ノードの状態値の数 L を列にとって K 行、 L 列の行列となる。計算量を考慮に入れなければ親ノードの種類や数は任意に定めることができる。しかし、親ノード数が増えるとCPTのサイズが増大し、条件付き確率の計算量が膨大になり、計算が困難になる可能性がある。特に今回は推薦するメニューの数 K が約1000種類にも及ぶためCPTのサイズが大きくなりがちであり、モデルに組み込む親ノードの選択とノードの数には注意しなければならない。

本論文では、ネットワークの構築とCPTの計算にBayoNet (Motomura, 2001) を用いた。BayoNetはSQLデータベースやCSV形式のデータからベイジアンネットワークをGUIベースで構築し、確率

推論を簡単に行う機能を備えたソフトウェアである。

5. 提案システム

ここでは、本論文で提案するシステム全体の構造と、利用者がシステムを利用する流れについて述べる。まず、提案するシステムの概要について、システムの処理の流れに沿って説明する。次に、本システムの特徴であるネットワークモデルの再学習について述べる。

5.1. システム概要

4節で構築したベイジアンネットワークモデルを用いて、利用者の意向からメニューを推薦するシステムを提案する。システムの内容は以下のように、前処理を行うStep0と、主となるStep1、Step2、Step3からなる。

Step0 メニュー分類基準の策定

利用者の大まかなメニューの方針に対応するメニューの分類基準を定める。

- Step1** 利用者の分類
システム利用者の食卓意識調査アンケートの結果から利用者を判別し、3節で作成した8つのtypeに分類する。
- Step2** メニューの決定
利用者のその日の意向を入力として、4節で構築したベイジアンネットワークモデルによる確率推論を行い、メニューの条件付き確率の値から推薦メニューを決定、提示を行う。
- Step3** 再学習
最終的に利用者が決定した献立のメニュー構成を新たなデータとして蓄積し、ネットワークモデルの学習を行う。

システムの全体像は図4のようになる。図4の右半分は食MAPデータから8つのtypeを作成する流れを表し、左半分は新しい利用者が推薦メニューを得るまでの流れを表している。それぞれのStepを詳細に説明すると以下ようになる。

メニュー分類基準の策定

- Step0** 利用者の大まかなメニュー方針に対応する変数として、メニュー分類基準を定める。メニュー分類基準は、分析用データから導出可能なメニューの持つ属性（肉料理、洋食など）に従ってメニューを分割する形で定める。例えば、肉料理、野菜料理などの分類や、和食か洋食かなどが考えられる。

利用者の分類

- Step1.1** 新しい利用者に対して食生活意識に関するアンケートを行う。アンケート内容は「モニター意識調査データ」に含まれる食生活意識に関する69の質問項目である。
- Step1.2** Step1.1で得られたアンケート結果を元に、判別分析などを用いて利用者が料理好き、健康指向、億劫感のそれぞれに対して肯定的か、否定的かを判別する。
- Step1.3** Step1.2で得られた判別結果を元に、利用者を8つのtypeのいずれかに分類する。

メニューの決定

- Step2.1** 利用者が入力したその日の食卓動機や大まかなメニュー方針を元に、分析用デー

タから作成したベイジアンネットワークの親ノードの状態値を設定する。ここで親ノードは、「食卓動機イベントデータ」の食卓動機に関する17の質問項目と、大まかなメニュー方針として、メニューの大分類や中分類などのメニューに関する分類基準から選択する。親ノードの数や種類は、実際の利用者の傾向や計算量の問題を考慮した上で設定する。

- Step2.2** Step2.1で親ノードの状態値を設定した上で、利用者のtypeに対応したベイジアンネットワークで確率推論を行い、メニューの条件付き確率を算出する。
- Step2.3** Step2.2で得られた条件付き確率の値に従ってメニューを並べ替え、上位いくつかのメニューを利用者に提示する。

再学習（詳細は5.2節参照）

- Step 3.1** 利用者のその日の献立のメニュー構成データを得る。
- Step3.2** Step3.1で得られたメニュー構成データと、Step2.1で得られた入力項目を新たなデータとして分析用データに蓄積する。

5.2. ネットワークモデルの再学習

新しい利用者の入力項目と、最終的に選んだメニューの情報から、ベイジアンネットワークの再学習を行うことができる。具体的には、新しいデータの追加により、親ノードと子ノードの条件付き確率が更新される。利用者の入力項目を受けて推薦したメニューが、その日の献立に採用されたのであれば、そのメニューは入力項目について適切なメニューとして、条件付き確率が上向きに修正される。逆に、推薦したメニューが献立に選ばれなかった場合は、そのメニューは不適切として、条件付き確率が下向きに修正される。新しい利用者が現れる度にこのような処理が行われ、メニューの条件付き確率が学習される。

この点がシステムにベイジアンネットワークモデルを用いた意義である。システムの利用者の入力項目と選択メニューの履歴データが蓄積されていくほど、システムは利用者にとって望ましい結果を出力できるように再学習していくことが期待される。

6. モデルの精度

ここでは、交差検証法 (cross validation) を用いてモデルの精度を検証する。まず、検証に用いたモデルの構造とデータについて述べ、その後交差検証法の結果得られたモデルの精度について述べる。

6.1. 検証に用いたモデルの構造

本論文で提案したベイジアンネットワークの構造は食卓動機とメニューの分類基準の組み合わせによって無数に存在するが、ここではモデルの検証のため、typeごとに共通の構造を持つと仮定し、モデルの構造をある一つに固定した上で検証を行う。ここで検証用モデルの親ノードとして採用した食卓動機の質問項目は、表6に記す3つである。

表6 親ノードとして用いた質問項目

質問コード	質問項目
Q1_11	買い置きの材料を中心に食事を支度
Q1_13	家族全員が喜ぶ料理を出した
Q1_14	栄養バランスを考えた食卓にした

これらの食卓動機の質問項目に加えて、メニューの分類基準としてメニューの大分類と、和洋などのメニューのスタイルを親ノードとして組み込んだ。ここで用いるベイジアンネットワークモデルの構造を図示すると図5のようになる。

6.2. 検証に用いたデータの範囲

ここで検証に用いるデータは、「食卓メニューデータ」に格納されている朝昼夕の食事データのうち、一般的にモニターが最もメニュー構成に悩

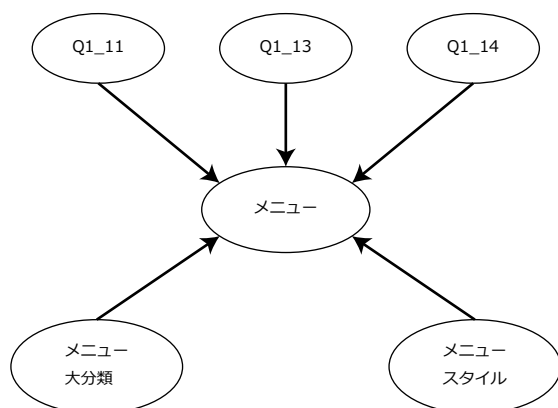


図5 検証に用いたベイジアンネットワークモデル

むだろうと考えられる夕食のデータを用いた。さらに、食卓の中でもメインデュッシュと考えられる「主菜」のデータに絞り、その中でも「手作り」かつ、メニューのF値(NTTデータライフスケープマーケティング, 2007) が2以上のものを検証用データとした。F値とは、そのメニューが平均いくつの食材を用いて作られているかを表す指標である。この結果、検証用データは30種別のメニュー大分類カテゴリと302種別のメニューより構成され、その総トランザクション数は56695となった。表6に記した3つの質問項目はYESかNOの値をとり、メニューのスタイルは(和風、洋風、中華風、その他)の4つの値をとる。また、本論文ではモニターのtype別に別々のベイジアンネットワークモデルを構築するため、上述した検証用データはさらに8つのtypeごとに分割されている。

6.3. 交差検証法の実施

次に、本論文で行った交差検証法の手順について述べる。交差検証法とは、検証用データを n 個に分割し、そのうちの一つの分割を評価用データ、残りの分割を学習用データとし、学習用データから構築したモデルを用いて評価用データとの適合度を計測するモデルの評価手法である。全データが一度は評価されるようにモデルの学習と評価を繰り返すことが特徴であり、モデルの妥当性評価に広く用いられている(元田他, 2006)。ここでは、6.1節、6.2節で述べたモデル構造と検証用データを用いて、モニターのtypeごとに交差検証法によるモデルの精度を検証する。

本論文では計算量の観点から検証用データを3つに分割した。そのうち2つを学習用データとし、残りの1つを評価用データとする。評価用データから、ある日のモニターの食卓動機とメニューの大分類、及びメニューのスタイルをモデルへの入力として親ノードの状態設定を行った。その上で学習用データから構築したベイジアンネットワークモデルで確率推論を行い、最も確率の値が高いメニューを出力とした。この出力されたメニューと評価用データ内で実際に食卓に並んだメニューを比較し、その一致率を求めた。実際にシステムを運用する場合は、出力するメニューを複数提示した上で利用者にメニューの選択を求める方が望ましいが、今回は一番に推薦されるメニューのみに着目し、実際に選択されたメニューとの一致率を用いてモデルの検証を行う。ここでは全データ

2011を3つに分割しているので、同様の手順を全データが一度は評価用となるように3度繰り返し、算出された一致率の平均をモデルの精度とした。表7にtypeごとの精度を記した。

表7 交差検証法の結果

type	精度	type	精度
1	0.5077	5	0.4941
2	0.5387	6	0.5393
3	0.5568	7	0.5456
4	0.5347	8	0.5611

表7から、検証用モデルの精度はどのtypeにおいても0.5前後の値となった。精度の比較のためにメニューの大分類を親ノードから除外したモデルで同様の交差検証法を行った場合の精度は表8のようになる。

表8 メニュー大分類を除いたモデルでの交差検証法の結果

type	精度	type	精度
1	0.1198	5	0.1402
2	0.1487	6	0.1656
3	0.1889	7	0.1590
4	0.1529	8	0.1561

表8から、メニューの大分類を除いたモデルの精度は平均0.1539となり、メニューの大分類を含めたモデルと比べて大きく精度が下がった。メニューの大分類を親ノードから外した場合に出力される推薦メニューは、和食ならば焼き魚、洋食ならばハンバーグ、中華ならば焼き餃子など、万人受けするような定番メニューが多く推薦された。このことから、全食卓に並んだメニューのうち、一割強のメニューは定番メニューで構成されていると考えられる。表7、表8の比較から、定番メニューから外れるようなメニューを推薦するためには、魚料理、肉料理などのメニューの大分類基準を親ノードに採用し、利用者からの入力で細かい条件を設定する必要がある。多数の親ノードを設定すればモデルの精度は上がっていくと考えられるが、利用者の手間や4節で述べた計算量の問題から、バランスよく親ノードを選ぶ必要がある。

7. 今後の課題

本論文では、食MAPデータを用いた利用者向けのメニュー推薦システムの仕組みを提案した。ここでは、システムの課題点について述べる。

7.1. ネットワークモデルの構造と計算量の問題

4節ではベイジアンネットワークモデルの構築の流れについて述べた。本論文では、親ノード間のリンクが存在しない単結合なリンク構造を仮定したが、複結合 (multiply connected) なリンク構造のモデルを構築することも考えられる。ただし、親ノード間にリンクが存在する (ループが存在する) 複結合なモデルの場合、条件付き確率の計算が複雑になるため、5.2節の再学習において単結合なモデルよりも計算量が多くなる。4節で述べた通り、全ての状態値の組み合わせ数は各々の親ノードとのとる状態値の数の積となるため、あまりに子細な分類基準を親ノードとして採用すると、計算量が爆発的に増加してしまう。このため、親ノードの選択とネットワークの構造の選択には注意が必要である。

また、本論文では親ノードの候補は食卓動機に関する質問項目とメニューの分類基準のみとしたが、買い置きの食材データや前日の残り物についての情報があれば、それらをノードとして取り込むことでより満足度の高いモデルが構築できると考えられる。加えて、誕生日などの「食卓イベントデータ」をノードとして取り込むことで、ハレの日のメニューなどにも対応できるだろう。6節で述べたように、親ノードの数や種類によってモデルの精度は大きく変わり、推薦されるメニューも異なってくるため、実際にシステムを運用する場合には利用者の傾向や計算量を考慮した上で柔軟にノードを選択する必要がある。

7.2. マンネリ化の問題

満足度の高いシステム構築のための課題として、現在のシステムでは、利用者の短期的な嗜好の変化に対応できないことが挙げられる。現状では、同じ食卓動機、同じ大まかなメニュー方針の下では同じメニューしか推薦できない。このため、システムを継続的に利用する利用者には推薦されるメニューに対しての「飽き」が発生してくるだろうと考えられる。こういったマンネリ化を避けるためには、推薦メニューの中に意外性のある例外

メニューを含めるなど、利用者に刺激を与える仕組みの採用を検討せねばならない。

7.3. その他の問題

提案したシステムはメニューの名称を提示するに留まり、レシピまでは提示しない。このため、レシピが複雑なメニューの場合、利用者はその日の意向に応じたメニューの名称を知ることができても、結局推薦されたメニューを献立に採用できない可能性がある。メニューの材料データとレシピも同時に提示する仕組みをシステムに組み込めば、この問題を解決できると考えられる。

加えて、現在のシステム構造では利用者は食生活意識や食卓動機に関するアンケートに答えなければならない。より手軽に利用できるシステムの開発には、食MAPのアンケートそのままではなく、利用者が負担を感じないような質問項目の考案が必要になるだろう。

8. おわりに

本論文では食MAPデータの新しい応用方法として、主婦が毎日考えなければならない献立の構成に注目し、主婦の要望に応じたメニューを推薦するシステムについて提案した。提案したシステムでは、食MAPデータに基き利用者を8つのtypeに分類し、各typeとにベイジアンネットワークを用いてモデル化を行った。ベイジアンネットワークモデルを用いることで、利用者の性質と意向に応じたメニューの推薦が可能になると期待できる。また、システムが利用されればされるほど、ネットワークモデルの学習が進み、状況に応じて適切なメニューを推薦できるシステムに成長していくことも提案したシステムの利点である。しかし、システムとして解決すべき課題も多く、システムのパフォーマンス向上には異なるシステム同士で利用者の満足度を比較することが必要となるだろう。

参考文献

- Hooley, G. J. and Saunders, J.(1993). *Competitive Positioning: The key to market success*, Prentice-Hall International.
- Motomura, Y.(2001). BAYONET: Bayesian network on neural network, *Foundation of Real-World Intelligence*, pp.28-37.
- NTTデータライフスケープマーケティング(2007). 『食卓解体新書2007年度版』, 株式会社NTTデータライフスケープマーケティング.
- 片平秀貴(1987). 『マーケティング・サイエンス』, 東京大学出版会.
- 駒澤勉(1982). 『数量化理論とデータ処理 (林知己夫監修)』, 朝倉書店.
- 齋藤隆(2003). 『365日の食卓マーケティング』, 株式会社NTTデータライフスケープマーケティング.
- 齋藤堯幸・宿久洋(2006). 『関連性データの解析法. 多次元尺度構成法とクラスター分析法.』, 共立出版.
- 繁樹算男・植野真臣・本村陽一(2006). 『ベイジアンネットワーク概説』, 培風館.
- 寺本益英(2008). 『現代人の食卓と緑茶: 食MAPによる分析』『総研論集』, 19, pp.1.15.
- 農林水産省(2010). 『農林水産省/食事バランスガイドについて』, http://www.maff.go.jp/j/balance_guide/index.html.
- 元田浩・津本周作・山口高平・沼尾正行(2006). 『データマイニングの基礎』, オーム社.
- 矢野順子・後藤和宏・桑田修平・恒松直幸・生田目崇(2010). 『食卓メニューデータに基づくメーカー主導のタイミングを考慮した販売促進のための分析フレームワーク』『オペレーションズ・リサーチ』, 55 (2), pp.98.105.