

質問の情報量に着目した回答の抽象度決定法

1 R-9

藤本 和 則 湯 川 高 志

NTT コ ミ ュ ニ ケ ー シ ョ ン 科 学 研 究 所

1 はじめに

質問を受けたとき、人間は適切な抽象度で回答する。例えば、「ワンワンとほえる動物は?」と聞かれたとき、「コリー」や「肉食動物」ではなく「犬」と答える。このように人間は、複数の抽象度の答から“適切な抽象度”の答を選び出して回答する。

本稿では、意味的な階層をもつ確率モデル(以下、階層的確率モデルと呼ぶ)に基づく分類問題において、人間のように適切な抽象度の答を選び出して回答する方法を提案する。

2 答の抽象度を決定する問題

例として、図1に示すように、質問者が動物の特徴をもとに、その名称を調べる場合について考える。

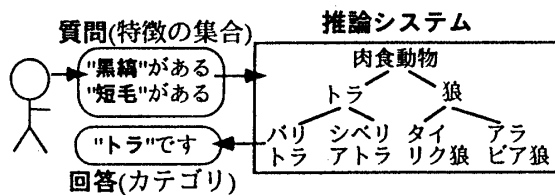


図1 質問と回答

図1で動物がバリトラであった場合、推論システムの回答する答としては、「トラです」、「バリトラです」などが考えられる。これらの答は、いずれも正しいが、推論システムとしては、これら複数の抽象度の答から、質問者にとって最も適切な答を選んで提示できることが望まれる。

3 適切な抽象度

ここでは、我々が“質問者にとって適切な抽象度”をどう捉えたかについて説明する。以下では、同一の親カテゴリをもつカテゴリの集合を子集合、子集合に属するカテゴリを子カテゴリと呼ぶ。まず、図2に示すよ

うに、ある子集合が{ウマ,ロバ,ヒツジ}で「ウマ」の子集合が{シマウマ,黒ウマ}である階層構造を例に考える。質問者が、動物の縞に着目して「白い縞がある」と質問したときは、「白い縞があるか否か」に基づいて分類されている{シマウマ,黒ウマ}の「シマウマ」が適切な回答と考えられる。これに対して、動物の走る速さに着目して「走るのが速い」と質問したときは、「走るのが速いか否か」に基づいて分類されている{ウマ,ロバ,ヒツジ}の「ウマ」が適切な回答と考えられる。このように、我々は“質問者が着目した分類を最も実現する抽象度が回答に最も適切な抽象度である”と捉えた。

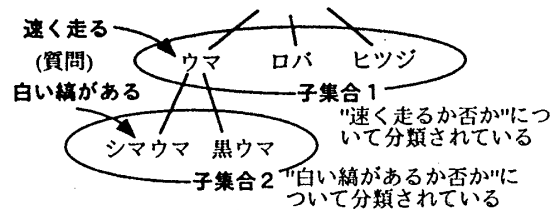


図2 適切な抽象度

4 抽象度の決定法

ここでは、各子集合での分類が質問者の着目した特徴に基づいているか否かの程度を表す分類度を算出し、それに基づいて適切な抽象度の答を決定する方法を示す(階層的確率モデルとしては、“カテゴリについての意味的な包含関係を表す木構造”と“特徴から各カテゴリの事後確率値を導く確率値データ”が与えられるとす(図3参照))。

4.1 分類度の計算法

まず、子集合の分類度の算出法について説明する。ある特徴 F に基づいて的確に分類されている子集合では、特徴 F をもつか否かで各子カテゴリを区別できる。したがって、特徴 F が質問として与えられると、子カテゴリの事後確率値のばらつきは大きくなる。これに対して、的確に分類されていない子集合では、特徴 F が質問として与えられても、子カテゴリの事後確率値のばらつきは小さいままである。したがって、分類度

の算出は、事後確率値のばらつきが大きいほど大きな値を与えるような計算式によって実現できる。

ここでは、以上の考えに基づいた分類度の計算式として、質問のもつ情報量に着目した計算式を示す。まず、質問として一つの特徴 F が与えられたときの子集合 $C(\{C_1, \dots, C_n\})$ の分類度 $D_e(C, F) \rightarrow [0, 1]$ を次式で定義する ($P_S = \sum_{i=1}^n P(C_i|F)$),

$$H(n, p_i) = -\sum_{i=1}^n p_i \log_2 p_i: \text{エントロピを与える関数}.$$

$$D_e(C, F) = \frac{H(n, \frac{1}{n}) - H(n, \frac{P(C_i|F)}{P_S})}{H(n, \frac{1}{n})} \quad (1)$$

式(1)は、事後確率値の和が1になるよう正規化した値 ($P(C_i|F)/P_S$) に基づいて、特徴のもつ情報量を計算し、それを情報量の最大値 ($H(n, \frac{1}{n})$) で除した式である。

また、質問として複数の特徴 $F_1, \dots, F_m$ が与えられたときの、子集合 C の分類度 $D(C, F_1, \dots, F_m) \rightarrow [0, m]$ を次式で定義する。

$$D(C, F_1, \dots, F_m) = \sum_{k=1}^m D_e(C, F_k) \quad (2)$$

式(2)は、各特徴についての分類度を質問として与えられた特徴全てについて加えた式である。

以上の計算式に基づいて、複数の特徴が質問として与えられたとき、各子集合の分類度を算出することができる。

4.2 抽象度決定手続き

ここでは、以上の分類度に基づいて、適切な抽象度のカテゴリを一つ決定する手続きを示す。質問が与えられたら、まず、式(1)と式(2)に基づいて、各子集合の分類度を計算する。次に、最大の分類度をとる子集合を取り出す。そして、取り出した子集合に属すカテゴリのうち、最大の事後確率値をとるカテゴリを取り出す。以上の手続きによって、質問者の着目した分類を最も実現する抽象度において最も確からしいカテゴリ、すなわち、回答に最も適切な答を決定できる。

5 適用例

ここでは、動物の分類問題を取り上げ、本決定法の適用例を示す。いま、図3に示す階層的確率モデルについて、質問 {黒縞} と質問 {短毛} がそれぞれ与えられたとする。このときの各カテゴリの事後確率値と、各子集合の分類度の計算結果を表1に示す。

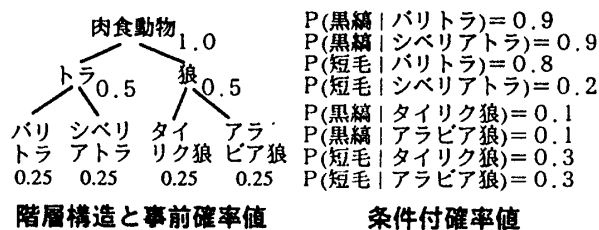


図3 階層的確率モデル

表1 計算の結果

(a) 事後確率値

カテゴリ	質問	黒縞	短毛
トラ		0.90	0.63
バリトラ		0.45	0.50
シベリアトラ		0.45	0.13
狼		0.10	0.38
タイリク狼		0.05	0.19
アラビア狼		0.05	0.19

(b) 分類度

子集合	質問	黒縞	短毛
{トラ, 狼}		0.53	0.05
{バリトラ, シベリアトラ}		0.00	0.28
{タイリク狼, アラビア狼}		0.00	0.00

表1に基づいて、質問が {黒縞} であるときは、まず、最大の分類度(0.53)をとる子集合 {トラ, 狼} に着目して、そのうち、最大の事後確率値(0.9)をとる「トラ」を適切な答として決定できる。質問が {短毛} であるときは、まず、最大の分類度(0.28)をとる子集合 {バリトラ, シベリアトラ} に着目して、そのうち、最大の事後確率値(0.5)をとる「バリトラ」を適切な答として決定できる。また、質問が複数の特徴 {黒縞, 短毛} であるときは、式(2)から得られる分類度を用いて、最大の分類度(0.58)をとる子集合 {トラ, 狼} に着目し、そのうち、最大の事後確率値(0.94)をとる「トラ」を適切な答として決定できる。

6 おわりに

本稿では、質問のもつ情報量に着目して、適切な抽象度の答を決定する方法を提案した。適切な抽象度を決定する方法としては、Fisherらの提案する方法[1]があるが、分類の数を多くすると必ず抽象度の高い答えが選ばれてしまうという問題点がある。本提案の決定法は、事後確率値そのものではなく、そのばらつきに着目した決定法なので、分類の数に依存せずに適切な抽象度の答を選び出すことができる。これら従来手法との関連については、稿を改めて報告する予定である。

参考文献

- [1] Fisher, D., "A Computational Account of Basic Level and Typicality Effects", Proc. of AAAI-88, pp233-238, 1988.