

4320

ファジィ集合による概念形成モデルの一提案
A Proposal of Conceptual Clustering Model Using Fuzzy Sets

阪南大学 林 勲, ○阪南大学 津田 エミ子,
(株)フライト 横井 真也, 阪南大学 市川 隆男

I.Hayashi, Hannan Univ., E.Tsuda, Hannan Univ., S.Yokoi, Flight Co.Ltd., T.Ichikawa, Hannan Univ.

Abstract. A new conceptual clustering model is discussed here to treat subjective data. The new model is formulated using fuzzy sets, which is called here fuzzy conceptual clustering model. In this paper, a similarity measure is defined to scale the relationship between subjective data using membership values. We also discuss four kinds of operators to assign the data to the conceptual tree. The fuzzy conceptual clustering model is a powerful method to construct the conceptual tree.

1. はじめに

属性値を含むデータが与えられたとき、各データ間や属性間の類似度を計算し、階層的分類木を構成する概念クラスタリング¹⁾がある。その一手法として、COBWEB^{2),3)}は有用である。COBWEBでは、データ間の類似指標を定義し、各データをクラスに所属させる4つの操作手順を定義して、分類木を構成する。しかし、人間の主観的データを取り扱う構造が定義されていないので、人間の主観的、かつあいまいなデータの分類には適していない。

本論文では、ファジィ集合を用いて人間の主観量と属性間の重みを取り扱う新たな概念クラスタリング手法を提案する。ここでは、このモデルをファジィ概念形成モデルと呼ぶ。ファジィ概念形成モデルでは、メンバーシップ値を用いて各データの類似度を計算し、分類木の構築過程において、この類似指標が最大となるように木構造を操作する。操作は既存クラスへの付加、新クラスの作成、クラスの統合、クラスの解体の4つの操作の中から選択される。本手法を簡単な分類の数値例に適用し、ファジィ概念形成モデルの有用性について検討する。

2. 類似指標

いま、 N 個のデータ $d_i, i=1,2,\dots,N$ の属性 $x_j, j=1,2,\dots,n$ に対して、それぞれ、 m_j 個のファジィ集合 $A_{jk}, k=1,2,\dots,m_j$ が与えられているとする。このデータ d_i をクラス $C_s, s=1,2,\dots,L$ に分類することを考える。データ間の類似指標を次のように定義する。

$$I = \sum_{i=1}^n \mu(x_j) \left(\frac{m_j}{k=1} \prod_{s=1}^L \mu(x_j|C_s) + \frac{L}{s=1} \prod_{k=1}^{m_j} \mu(C_s|x_j) \right) / 2 \quad (1)$$

$$\mu(x_j|C_s) = 1 - \sum_{i \in C_s} |\mu_{A_{ik}}(x_j) - \mu_{A_{ik}}(x_j)| / N_s \quad (2)$$

$$\mu(C_s|x_j) = \left| \sum_{i \in C_s} \mu_{A_{ik}}(x_j) / N_s - \sum_{s=1}^L \left(\sum_{i \in C_s} \mu_{A_{ik}}(x_j) / N_s \right) / L \right| \quad (3)$$

ここで、 $\mu(x_j|C_s)$ はクラス C_s での各データのメンバーシップ値 $\mu_{A_{jk}}(x_j)$ に対するクラス内類似度であり、 $\mu(C_s|x_j)$ は各データの属性 x_j のメンバーシップ値 $\mu_{A_{jk}}(x_j)$ に対するクラス間類似度である。また、 $\mu(x_j)$ は属性 x_j に対する重みであり、 N_s は C_s に含まれるデータ数を表す。なお、 $a \cdot b = a + b - ab$ である。 $\mu(x_j|C_s)$ の値が大きい場合にはクラス内の類似度が高いことを意味し、 $\mu(C_s|x_j)$ の値が大きい場合には、クラス間の非類似度が高いことを意味する。したがって、評価値 I が大きいほど、データがクラスに属する度合いが高いことを示している。

3. 分類木の構築操作とアルゴリズム

ファジィ概念形成モデルでは、データが逐次的に入力された場合に、(1)式の類似指標が最大となる操作を次の4つの中から選択して、分類木を構築する。

- (操作1) 既存クラスへの付加
- (操作2) 新たな単一クラスの生成
- (操作3) 複数の既存クラスの統合
- (操作4) 既存クラスの解体

分類木は次のアルゴリズムにしたがって構築される。

- 1) データを逐次的に選択する。
- 2) 類似指標 I を計算する。

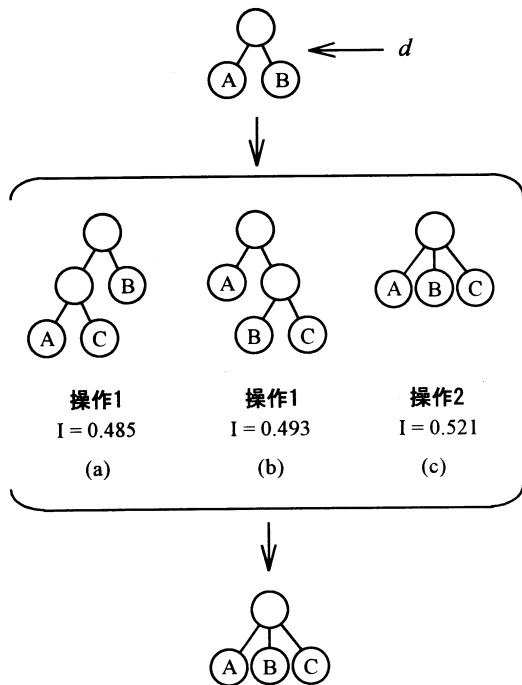


図1 ファジィ概念形成モデルの分類木の構築過程の例

3) 分類木の構築過程において、適用できる4つの操作の中で類似指標 I が最大な操作を1つ選択する。値が同値となる場合には、適用順序の最後の操作を採用する。

4) すべての与えられたデータを分類木にクラス化した時点で、アルゴリズムを終了する。

アルゴリズムの手順の例を図1に示す。いま、AとBの2つのクラスからなる分類木にデータ d を分類することを考える。

4つの操作の中で、この分類木に適用できる操作は操作1と操作2であるので、(a), (b), (c)の3つの分類木の候補が考えられる。類似指標 I を計算した結果、(c)の I の値が最も高くなったので、データ d は新たな単一クラスとして分類木に分類される。

4. 数値例

ファジィ概念形成モデルの有用性を検討するために、アルコール飲料の分類の事例を取り上げる。いま、表1のような、原料、製造法、アルコール度数、発酵・貯蔵期間、価格の5つの属性に対するアルコール飲料のデータが与えられているとする。これらの属性に対して、16種類のファジィ集合を用いてメンバーシップ値のデータを得た。

表1 アルコール飲料のデータ

種類	原料	製造法	度数	発酵貯蔵期間	価格
ウイスキー	麦, 穀物	蒸留法	40	10年	2,000円
ワイン	ブドウ	発酵法	10	2年	3,500
ビール	麦, 米	発酵法	5	2ヵ月	600
ブランデー	ブドウ, 果実, 穀物	蒸留法	60	6年	5,000
清酒	米, 麦, 穀物	発酵法	15	4週間	1,500

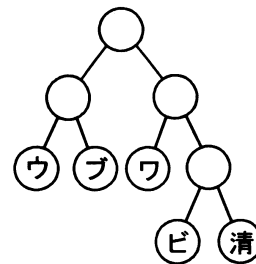


図2 結果の分類木

ファジィ概念形成モデルのアルゴリズムを適用し、得られた結果の分類木を図2に示す。表1のデータから、ビールと清酒とは比較的、属性値が類似しており、ワインとの属性値も類似しているといえる。この関係が結果の分類木に反映されている。また、ウイスキーとブランデーとは製造法やアルコール度数の観点から類似しているといえ、この点も分類木に反映されている。分類木における全データの概念構成は我々の直感とよく一致しているといえる。

5. おわりに

ここでは、ファジィ概念形成モデルを提案した。今後、類似指標の再検討や分類木の階層の差異をアルゴリズムに考慮する必要がある。

参考文献

- 1) R.S.Michalski, Y.Kodratoff: "Research in Machine Learning: Recent Progress, Classification of Methods, and Future Directions," Machine Learning III, pp.3-30, Morgan Kaufmann (1990)
- 2) D.H.Fisher: "Knowledge Acquisition via Incremental Conceptual Clustering," Doctoral Dissertation, Department of Information and Computer Science, University of California, Irvine (1987)
- 3) 榎木哲夫, 岩井壮介, 片井修: "概念クラスタリング手法による決定モデルの動的学習とリアクティブルールの獲得", 第18回知能システムシンポジウム, 第19回システムシンポジウム合同シンポジウム, pp. 199-203 (1993)