

財務指標を基にした企業のファジイクラスタリング とその応用

岸川, 善紀

<https://doi.org/10.15017/3000279>

出版情報：経済論究. 110, pp.19-28, 2001-07-31. 九州大学大学院経済学会
バージョン：
権利関係：

財務指標を基にした企業のファジイクラスタリングとその応用

岸 川 善 紀

1 はじめに

経済経営分野のデータは複雑で多岐にわたり、その量は膨大なものとなり、そのための処理の時間が増大する可能性がある。従って、実際の経営判断に利用するためにはその膨大な経済データをその情報を劣化させずに計算機等を利用して自動的に、使いやすい形にまとめる必要がある。

オブジェクトの類似性に注目して似たようなものをまとめあげ、いくつかのグループに分類する手法のことをクラスタリングと呼ぶ。ひとつのクラスタは類似のオブジェクトの固まりであって他のオブジェクトとは有意に異なるオブジェクトの固まりのことを言う。クラスタリングは外的基準を持たないデータの分類法である。よって、明確な評価を与えられていない生の観測データを分類するのに有効であり、計算機による自動的な処理にも対応させることができる。このように、経済の観測データをクラスタリングする場合、観測データをいくつかの特徴的な集団に分類することで、その類似データの検索を容易にしたり、経営データの場合、意思決定の参考としたりできる。

クラスタリングにはそのクラスタに対する所属、非所属が明確に分類されるハードクラスタリングがあるが、必ずしもその同一クラスタ内すべてのデータの類似性が高いとはいえない場合もある。また、クラスタを元に個体検索を行う場合、クラスタ内部の個体情報が失われているためクラスタレベルの粗い検索しか行われなくなる。このことは経営情報を用いて判断をおこなうシステムに利用する場合においては、その情報損失のために誤った情報を提供してしまう可能性がある。そのためハードクラスタリングの代わりにファジイ集合を導入したファジイクラスタリングの手法を用いることが考えられる。ファジイクラスタリングは一件の観測データに対し各クラスタへの所属度をメンバシップ値のファジイ集合の形で与えるので、メンバシップ値の情報を保存しておくことで個体レベルの情報を維持することができる。ファジイクラスタリングの手法としてファジイ c 平均法が代表的な手法であるため、本論文ではこれを適用する。また、応用として財務諸表データをもとにファジイクラスタリングを適応し、その分類結果について考察してみる。以下、2ではクラスタリングとその概念について説明する。3ではクラスタリングにファジイを適用する方法について説明を行ない、4では企業の財務指標を基にファジイクラスタリングを適用し、その結果を分析する。

2 クラスタリング

2.1 判別問題とクラスタリング

経済経営分野において、与えられた変数をもとにして、これを特徴とするサンプルをグループ化する必要性が生じる。最近のリスク管理における企業の発行する債券の格付けの問題を、財務指標を入力変数とする判別問題と考えることができる。あるいは、株価のこれまでの観測データに基づく特徴量を入力変数として、将来の上昇下降を予測する問題にも適用可能である。この場合、株価のレベルをいくつかのグループ（カテゴリ）に分割することにより離散化する必要がある。

このようなグループ化の問題では、主として2つの方法が適用され、これらは外的基準を与えるかどうかで区別される。その方法の1つは外的基準を与える方法であり、多変量解析における判別分析などがこれに相当する。経験的なデータ、観測などをもとにして、あらかじめサンプルが属するグループ（カテゴリ）が与えられており、これらの分類に適合するような判別関数を構成する問題である。これにより、分類が未知であるサンプルを入力変数をもとに分類することができ、企業の倒産予測、格付けの予測を行うことができる。しかし、この場合、過去のデータを収集することにほかに、外的基準を準備する必要があるなどの問題がある。

これに対する2番目の方法は、外的基準を用いない方法であり、クラスタリングが代表的な方法である。サンプルを特徴づける入力変数をもとにして、与えられたグループ数に分割する関数（正確にはグループの境界）を推定する方法である。このクラスタリングにより、最初からグループが分かっていない分類問題を解くことができる。新しい事業をはじめるときの顧客の分類や、企業業績のグループ化に有効である。

本論文では、このクラスタリングにファジイ理論を適用する。

2.2 クラスタリング

クラスタリングとは、外的基準を用いず、ただそのデータ空間上での類似性をもって同種類のものとする集団を形成するデータの分類手法である。クラスタとは「ぶどうの房」を意味し、類似のオブジェクトが固まった集団となっている状態をあらわしている。すなわちクラスタリングという分類手法はデータセットについてクラスタと呼ばれる類似のデータを一まとめにした集団を作り上げる手法である。

あるデータセットに多量の観測データが格納されているとき、一件の観測データは、特徴パターンを説明するために多次元ベクトルの形で表現されている。よって、一件の観測データはそのデータ空間上に一つの点として配置されている。クラスタリングによって観測データがいくつかのクラスタに分類されたとき、データ空間上ではいくつかの点が凝集した固まりと見ることができる。空間上でのクラスタリングの状態を図1のボロノイ図に示す。同一のクラスタに所属するデータには同じ番号をラベル付けされることとなる。

クラスタリングでは観測データをクラスタに分類するとき、それぞれのクラスタリングの手法に

従って観測データからその特徴パターンを抽出し、共通するパターンを近いもの、同一のものとして統合しラベル付けを行い、いくつかの集団にグループ化するのである。一旦クラスタリングされたデータセットの情報はフィードバックされ、さらに最適な分類となるようにクラスタリングが更新される。

基本的なクラスタリングの各段階の概念図を図2に示す。

2.3 クラスタリングの手法

クラスタリングの手法には類似性をどのように見出すかによって多様なアプローチが考えられる。データが多次元ベクトルの形で表現されたデータ空間上でのクラスタリングにおいては、超平面を境界線として設定して各クラスを分割する手法や、データ間の距離をもとに類似性を考える手法など

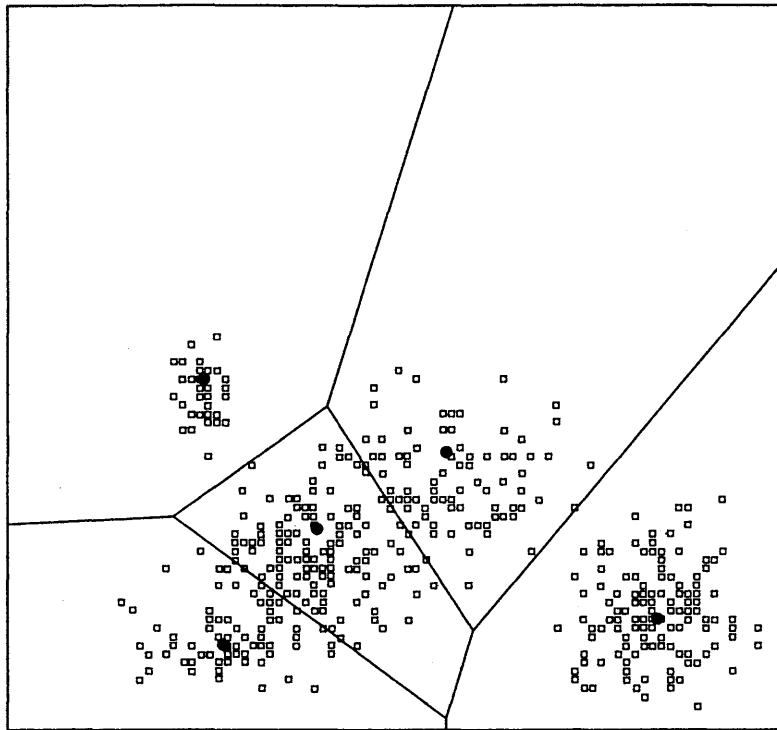


図1：座標空間におけるクラスタリングのボロノイ図

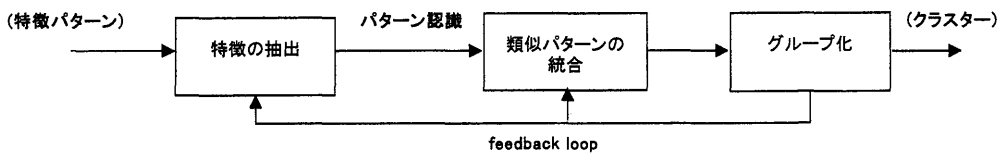


図2：クラスタリングの各段階

がある。データ間の距離を類似性の基準としたクラスタリングは、二乗誤差を評価関数にとるものが代表的な手法である。これは、ひとつのクラスタにその代表値を設定し、それを中心とした二乗誤差を基準に凝集したクラスタ集団を形成する手法である。また、多くのデータ分布において、各クラスタの境界はファジイであるためハード（クリスプ）クラスタリングよりもファジイクラスタリングの方が安定して動作する。

3 ファジイクラスタリング

3.1 ファジイ理論とクラスタリング

ファジイ理論は多岐にわたっているが、この中でもファジイ推論システムは、これまでさまざまな分野に応用されてきており、工学分野における制御をはじめ、経済分野においても判別問題や時系列の予測問題に適用されている。これまで、我々の研究においても、ファジイ推論システムを企業の債券の格付問題に適用することや、株価の予測問題に適用するなどの試みを行って有用な成果を得ている。

ファジイ推論システムを、機能としてとらえた場合に、1つは判別関数を構成する推論システムとしての機能がある。もう1つは、与えられたサンプルから、これを生成するシステム関数を推定、あるいは近似する機能である。ファジイ推論システムは、このようなことからuniversal approximatorとよばれることがある。これまでの研究で、カオス時系列から、これを生成するカオス力学系をファジイ推論システムにより推定する課題に対して、ファジイ推論システムを多段化してルール数を大幅に削減することに成功している。

本論文では、このようなファジイ推論システムの関数近似能力を、クラスタ構成のもとになる関数を近似することに用いる方法の1つであり、経済経営分野における応用を目的としている。

3.2 ファジイ集合

ある要素が、ある集合に所属している度合いを0と1の間の1つの数値として表すという考え方（クリスプ集合）がある。これに対して、ファジイ集合は例えば、完全に所属している場合に1、完全に属していない場合に0とし、そして属している度合いにしたがってその中間の値を与えようというのである。すなわち、その所属の度合いを0と1との間の任意の数値として認めたような集合がファジイ集合である。

クリスプ集合の場合はある要素がある集合に所属しているならば1、属していないならば0のいずれかの数値を取る。ファジイ集合は次のように定義する。 U を全体集合とし、 x を U の要素とする。このとき、ファジイ集合 A はメンバーシップ関数 $\mu_A(x)$ によって表現される。

$$\mu_A(x) : \rightarrow [0, 1]$$

この関数 $\mu_A(x)$ は x が集合 A に属する程度を示している。例えば、 μ_A の値が0のときは x は A にまったく属さず、逆に $\mu_A(x)$ の値が1の時は完全に A に属している。また、 $\mu_A(x)$ の値が0.5といった場合には、要素 x はそれなりに属することを示している。

ファジイ集合は連続量の場合、以下のような式で表現される。

$$A = \int_X \mu_A(x) / x \quad (1)$$

離散量の場合、以下のような式で表現される。

$$A = \sum_{i=1}^n \mu_A(x_i) / x_i \quad (2)$$

計算機においては連続量の場合も擬似的に離散数として取り扱うことになる。また、クラスタリングにおいてはいくつかの決まった数のグループに分類するため、離散量の形で計算される。

3.3 ハードクラスタリングとの違い

ハード（クリスプ）クラスタリングはひとつの観測データがひとつのクラスタに所属する確率が1または0で表現されるクラスタリングであるといえる。多くの場合、データ分布はそのクラスタの境界においてファジイなものとなり、クラスタ内部においてもすべての情報は類似であるといっても完全に同一ではないため固体そのものを示す特徴パターンは存在している。ハードクラスタリングによるラベル情報においては観測データの個体レベルでの情報が消失してしまう。同一クラスタ内での個体差の情報が失われてしまうため、同一クラスタなどの荒いレベルでの情報しか提供されなくなる。ファジイクラスタリングは各クラスタの所属度がメンバシップ関数のファジイ集合で与えられるため、その情報を記憶することで固体特有の特徴も維持することが可能である。

3.4 ファジイ c 平均法

ファジイ c 平均法はDunnが導入し、Bezdekが一般化したクラスタリングのアルゴリズムであり、繰り返し計算により最適値を求める反復アルゴリズムである。クラスタの中心と各データとの二乗誤差の和を用いたオブジェクト関数が与えられたとき、その関数 $J_m(U, v)$ を最小にするファジイ集合 U とクラスタの代表値 v を求める。

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2, \quad 1 \leq m < \infty \quad (3)$$

$X = \{x_1, x_2, \dots, x_n\}$ は有限のデータセットである。 $x_k \in R^d$, $1 \leq k \leq n$ は d 次元のベクトルの形で与えられる k 番目のデータである。 $\|\cdot\|$ はユークリッド距離であり、類似性の尺度として用いることができる。 $v = (v_1, v_2, \dots, v_c)$ はクラスタの代表値のベクトルであり、 v_i は i 番目のクラスタの中心となる座標として表現される。 d 次元のベクトルの形で与えられ、分類するクラスタの数 c だけ存在する。ここで $m=1$ とした場合、上のオブジェクト関数は古典的な二乗誤差関数となり、 v_i はハードクラスタリングにおける c 平均法のクラスタの代表値となる。Bezdekは $m \geq 1$ の場合、 $x_k \neq v_i, \forall i, k$ の仮定のもとで、 $J_m(U, v)$ を最小にする最適値 (\bar{U}, \bar{v}) を求めるために必要な条件を以下のように与えている。条件1ではメンバシップ関数 U のファジイ値を求める。

$$\bar{u}_{ik} = 1/T, \quad \forall i, k \quad (4)$$

ただし、

$$T = \sum_{j=1}^c \left(\frac{\|x_k - \bar{v}_j\|}{\|x_k - \bar{v}_j\|} \right)^{2/(m-1)} \quad (5)$$

条件 2 は各クラスタの代表値を与える。

$$\bar{v}_i = \frac{\sum_{k=1}^n (\bar{u}_{ik})^m x_k}{\sum_{k=1}^n (\bar{u}_{ik})^m}, \quad \forall i \quad (6)$$

この二つの式を繰り返し計算することで $J_m(U, v)$ を最小化する最適な値 (\bar{U}, \bar{v}) を求めていく。ファジィ c 平均法のアルゴリズムは以下のようなステップで説明される。この概要を図 3 に示す。

step 1 : $p=0$ とする。クラスタ数 c を $2 \leq c \leq n$ の範囲内, m を $1 \leq m < \infty$ の範囲内で与える \bar{U}_0 を乱数によって与える (初期化)。 $p=1$ にする。

step 2 : 条件 2 を使って \bar{v}^p を \bar{U}_{p-1} を利用して算出する。

step 3 : step 2 で計算された \bar{v}^p と条件 1 の式を使って \bar{U}_p を更新する。

step 4 : 終了条件 $\|\bar{U}_p - \bar{U}_{p-1}\| \leq \varepsilon$ を満たした場合, 計算を終了する。それ以外の場合 $p \leftarrow p+1$ としてステップ 2 に戻る。

4 企業のクラスタリング

以下では、ファジィクラスタリングの応用例として企業を、その財務指標の類似度によりグループ化する問題を考察する。企業の特徴は、その業種が大きな影響を与えることは予想されるが、業種

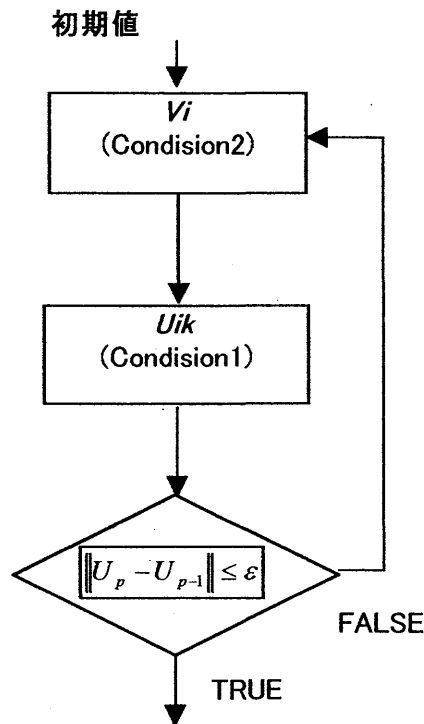


図 3 : ファジィ c 平均法のアルゴリズム

を越えて好調な企業やそうではない企業など、パフォーマンスに関する評価も重要である。企業の公表するデータの主要なものとして財務指標がある。以下では、財務指標をもとにして、企業をいくつかのクラスタに分割する問題にファジイクラスタリングを用いる。

4.1 財務データ分析

企業の財務分析に使われるデータは毎年有価証券報告書などの形で決算データとして公表され、これを企業の業績評価として利用することができる。経営財務分析においては貸借対照表や損益計算書などのデータを解析して行う。貸借対照表の数値をそのまま解析の対象として利用する分析を実数値分析と呼ぶが、この分析方法では企業の規模によってほとんどの結果が決まってしまう。クラスタリングを行った場合も単純に企業の規模の分析となってしまう可能性があるため企業の経営状況を分類したい場合この数値を利用することは好ましくない。

そのため、企業の規模で実数値を割った比率を解析の対象とする比率分析のデータを利用することとなる。財務指標データは利益率や資本比率などの指標で与えられる。財務分析に使われるデータについて、ほかにキャッシュフローのデータがある。キャッシュフローとは設備投資およびそれに伴う資金繰りの問題であり、簡易キャッシュフローは稼いだお金、設備投資額が長期的視点で使ったお金などを考慮するということになる。

4.2 実験

今回の実験では企業の財務データについてファジイクラスタリングを行い、その分類結果についての評価を行う。

企業の財務データは自動車、造船、電力、ガス、建設、電気機器、製紙、鉄鋼、通信関連などの業種から約100社の財務指標とキャッシュフロー関連の指標をデータセットとし、CD-ROMから2000年3月の決算データを入手し、それを利用する。観測データのパラメータとして以下の財務指標データを用いる。財務指標データの中には業種、企業によって存在しない指標もあるため、各企業に共通に存在するデータのみを用いている。

4.3 クラスタリングの評価

クラスタリングを行った結果より分析を行う。次の表はクラスタの数を変化させていった場合において、クラスタリングをクリस्प化（所属度が最も高いクラスタに振り分ける）した場合における分類の成功度である。あるクラスタに対するメンバシップ値が著しく高い場合に、その観測データはそのクラスタに所属するものとみなす。表の上の段の数字は、分類されるクラスタの数である。

次に企業毎のクラスタリング分類の凝集率を示す。メンバシップ値 $u_{ik} \geq 0.8$ のとき、そのクラスタに所属しているとみなしている。

業種によっては製紙のようにクラスタ数を増やしたことで分類がいくものもあるが、基本的にクラスタの数を増やすほど、その所属するクラスタがはっきりしない観測データが増えている。かといって、クラスタ数を少なくしても、分類上の意味があまり見い出せない。よって、クラスタの数は3か

表 1. 財務指標

回転率	固定資産回転率 (回), 有形固定資産回転率 (回), 有形固定資産回転率・取得価格基準 (回), 流動資産回転日数 (日), 売掛金回転日数 (日), 棚卸資産回転日数 (日)
設備投資費率	売上高減価償却費率 (B) (%), 売上高設備投資比率 (B) (%) 償却前内部留保率 (%)
利益率	使用総資本総利益率 (%) 使用総資本営業利益率 (%) 使用総資本経常利益率 (%), 使用総資本税引前利益率 (%) 使用総資本利益率 (%), 売上高総利益率 (%) 売上高営業利益率 (%), 売上高経常利益率 (%)
従業員一人当り	従業員 1 人当売上高 (10万円), 従業員 1 人当経常利益 (10万円) 従業員 1 人当当期利益 (10万円), 労働装備率 (10万円) 従業員 1 人当使用総資本 (10万円)
各種比率	支払準備率 (%), 当座比率 (%), 流動比率 (%) 正味運転資本額 (100万円), 株主資本比率 (A) (%) 株主資本比率 (B) (%), 負債比率 (%), 減価償却率 (%) 減価償却実施率 (%), 有形固定資産減価償却累計率 (%) 増収率 (前年比) (%)
キャッシュフロー	簡易版キャッシュフロー (B) (100万円) CF対負債比率 (%), CF対流動負債比率 (%) 売上高利益率 (%), 売上高税引利払前利益率, 使用総資本回転率 (回)

表 2. 全体のクラスタリングの凝集率

u_{ik} の値	($c =$) 3	5	7	9
$u_{ik} \geq 0.7$	0.94	0.90	0.66	0.60
$u_{ik} \geq 0.8$	0.92	0.82	0.59	0.54
$u_{ik} \geq 0.9$	0.84	0.73	0.53	0.39

表 3. 各業種毎のクラスタリングの凝集率

業 種	($c =$) 3	5	7	9
自動車	0.86	0.86	0.21	0.21
建設	1.00	1.00	0.5	0.5
電力	0.7	0.7	0.7	0.6
電 器	0.75	0.25	0.25	0.5
ガ ス	1.00	1.00	0.82	0.54
製 紙	1.00	0.63	0.38	0.75
造 船	1.00	0.92	0.85	0.77
鉄 鋼	1.00	0.88	0.44	0.33
通 信	0.92	0.84	0.72	0.64

表4. 各企業のクラスタの散らばり

(クラスタ)	1	2	3	4	5	(未分類)
自動車(14)	0	1	11	1	0	(1)
建設(4)	0	0	4	0	0	(0)
電力(10)	4	0	3	0	1	(2)
電器(4)	0	2	1	0	0	(1)
ガス(11)	0	0	11	0	0	(0)
製紙(8)	0	0	5	0	0	(3)
造船(13)	0	0	12	0	0	(0)
鉄鋼(9)	0	1	8	0	0	(0)
通信(25)	0	0	23	0	0	(2)

ら5のあたりで行うのが望ましいようである。

以下はクラスタ数を5に固定した時、各業種の企業がどのクラスタに分類されているかを示す表である。クラスタの番号は分類時に任意でつけたラベルであり、何らかの外的基準によるものではない。

所属度最大のメンバシップ値が0.7以下のものは未分類としている。

大半の企業がクラスタ3に集中している。クラスタ3は標準的な企業の集団であると見ることができ、電力は比較的ばらつきがある。ガスや造船などの業種はクラスタ数を変化させてもそれに類する企業は同一クラスタに集まる傾向にある。

5 おわりに

今回は、ファジイc平均法のアルゴリズムによるファジイクラスタリングの手法を財務指標分析に適用し、その評価を行なった。また、クラスタリングによる分類法は、外的基準を使用していないため、いわば特別なガイドラインを持たない分類法となっており、データの预处理に有効であると思われる。

今後の研究としては、企業の各付データなどの外的な評価基準との比較を行ない、その関連性を確認し、ファジイクラスタリングを利用した個別データに対応したデータ検索などにつなげていきたい。

参考文献

- [1] 日経新聞社：日経財務データCD-ROM, (2001)
- [2] Bezdek, J. C, "Pattern Recognition with Fuzzy objective Function Algorithms", *Plenum Press*, New York (1981)
- [3] A. K. Jain, M. N. Morty, P. J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, vol.131, No. 3, pp264-323 (1999)
- [4] 井上光平, 浦浜喜一, "共起関係行列に基づくファジークラスタリングとデータ検索への応用", 電子情報通信学会

論文誌D-2, vol.183, pp.957-966 (2000)

- [5] K. Tan and S. Tokinaga: "Applying the multi-stage fuzzy inference system with the genetic algorithm", *technical Report of IEICE*, DSP97-161, pp.105-112 (1998).
- [6] K. Tan and S. Tokinaga: "The design of multi-stage fuzzy inference systems with smaller number of rules based upon the optimization of rules by using the GA", *Trans. IEICE*, vol.E82-A, no.9, pp.1865-1873 (1999).
- [7] Y. Kishikawa and S. Tokinaga: "Prediction of stock trend by using the wavelet transform and the multi-stage fuzzy inference system optimized by the GA", *Trans. IEICE*, vol.E83-A,no.2, pp.357-366, (2000).
- [8] Y. kishikawa and S. Tokinaga: "Approximation of multi-dimensional chaotic dynamics by using multi-stage fuzzy inference systems and the GA", *Trans.IEICE*, vol.E84-A, to apper, (2001).
- [9] A. Zadeh: "Fuzzy sets", *Information and Control*, 8, pp.338-353 (1996).
- [10] T. J. Procyk and E. H. Mandami: "A linguistic self-organizing process controller", *Automatica*, 15, pp.15-30 (1979).