和多田 淳三（大阪工業大学）

Data Analysis and Its Fuzzy Approach

Junzo Watada (Osaka Institute of Technology)

データ解析とファジィについて述べるために、その関連文献を概括しておく。

1965年にZadehによってファジィ集合の概念が提案されたあと、早い時期にクラスタリング手法にファジィ集合の概念が導入された。また、1971年にはZadehがファジィ集合に関する類似性の概念を提案している。Ruspiniによるファジィ分割法[44-46]によるクラスタリング手法、BallとHallの開発したISODATAアルゴリズムにファジィの概念を導入したDunn[14-17]やBezdek[3-5,7]のクラスタリング手法、その他各種の方法が提案されている、例えば動的計画法を適用する方法[18]、菅野の測度を用いる方法[38]、類似性概念とクラスタリングの関係[21,30,32]、さらにクラスタリングの正当性の議論をRoubens[42-43]およびWindham[76,77]が行っている。階層的クラスタリングについては大隅等が議論を行っている[41]。和多田らによる発見的な方法による階層的クラスタリング手法はZadehの類似性の概念を現実のデータに適用するための有効な方法論を提供している[62-63,65]。これらのファジィクラスタリングについては文献[8,13,16,41]に詳しい。

1980年の田中らのファジィ線形回帰モデルの提案の後各種回帰手法が提案されており、また種々の分野で応用された結果が報告されている[51-57,71]。さらにファジィ線形回帰モデルを基にした各種データ解析手法も提案されている。例えば、和多田らのファジィ時系列分析[29,66,74]、林らのファジィGMDH手法[28]などである。

Saatyの開発したAHPモデル[47,49]にファジィ理論を導入した方法が1984年にBuckley[9]や市橋[30]らによって提案されている。

日本で数量化理論の名前で取り扱われている質的なデータの数量化手法である回帰分析、グループ判別法、パターン分類、多次元尺度構成法にファ

ジィ確率[78]の概念を導入した手法[24-27]が和多田ら[60-61,68,72,73]によって提案され、各種領域で分析手法として用いられている[60,73]。

一方ファジィ測度に基づくファジィ積分をもちいた方法もデータ解析手法として実的な問題に用いられている[40]。

ファジィデータに関する判別分析については和多田ら[64]が取り扱っている。

以上の手法を整理すると表1のように分類することができる。

分類に関する手法をまとめると表2になる。事前に分類が与えられていない場合（分類基準のない場合）と、分類が与えられておりその分類の基になる意味的な構造を分析する場合（分類基準のある場合）とに、大きく分けることができる。

分類基準のある場合については、上記したようにRuspiniによるファジィ分割の手法、BezdekによるファジィISODATA、YehとBangによるファジィグラフの手法などのファジィクラスタリング手法や、ファジィ階層的クラスタリング手法、ファジィ数量化分析3の名前で知られている類質的データのパターン分類手法を挙げることができる。さらに、分類基準のある場合については、ファジィ数量化分析2類の名前で知られているファジィ群の判別分析法を挙げることができる。

ここでは、パターン分類の観点から3手法に限定して説明する。初めに座標が与えられたときにその空間的な距離から指定された数のクラスターにファジィクラスタリングするための手法であるファジィk-means法を、次に質的な一対比較等から得られるデータの階層的クラスタリングについて説明する。最後にファジィパターン分類について説明する。

表1 ファジィデータ解析手法

手法名	ファジィ理論の方法
ファジィクラスタリング	ファジィ分割 ISODATA ファジィグラフ
ファジィ階層的クラスタリング	ファジィ類似度、ファジィ関係
ファジィ回帰分析	ファジィ数
ファジィ時系列分析	ファジィ数
ファジィGMDH	ファジィ数
ファジィ判別分析	ファジィ数
質的データの回帰分析	ファジィ確率・ファジィ統計
質的データのファジィ判別分析	ファジィ確率・ファジィ統計
質的データのパターン分類	ファジィ確率・ファジィ統計
質的データの多次元尺度構成法	ファジィ確率・ファジィ総計
ファジィAHP	ファジィ数
ファジィ測度分析	ファジィ積分

表2 ファジィ分類手法

分類基準	手法名	目的
無し	(ファジィ分割) ファジィクラスタリング (ファジィ) (グラフ理論)	分類 分類 分類
	ファジィ階層的クラスタリング	分類
	質的データのパターン分類	意味付け、 多属性の数量化
有り	質的データのファジィ判別	意味付け、 多属性の数量化
	ファジィ判別分析	意味付け

References

- [1] J.C. Bezdek and J.C. Dunn: Optimal fuzzy partitions: a heuristic for estimating the parameters in a mixture of normal distributions, *IEEE Transaction on Computer*, 24, 835-838, 1975.
- [2] J.C. Bezdek and J.D. Harris: Fuzzy partitions and relations: An axiomatic basis for clustering, *Fuzzy Sets and Systems*, 1, 111-127, 1978.
- [3] J.C. Bezdek: Numerical taxonomy with fuzzy sets, *Journal of Mathematical Biology*, 1, 57-71, 1974.
- [4] J.C. Bezdek: Cluster validity with fuzzy sets, *Journal of Cybernetics*, 3, 3, 58-73, 1974.
- [5] J.C. Bezdek: A physical interpretation of fuzzy ISODATA, *IEEE Transaction on Systems, Man, and Cybernetics*, May, 1976.
- [6] J.C. Bezdek: Prototype classification and feature selection with fuzzy sets, *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-7, 2, 1977.
- [7] J.C. Bezdek: A convergence theorem for the fuzzy ISODATA clustering algorithms, *IEEE Tran. on Pattern. Anal. Machine Intell.*, 2, 1-8, 1980.
- [8] J.C. Bezdek: *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [9] J.J. Buckley: Fuzzy Hierarchical Analysis, *Fuzzy Sets and Systems*, 17, 233-247, 1985.
- [10] M.R. Civanlar and H.J. Trussell: Constructing membership functions using statical data, *Fuzzy Sets and Systems*, 18, 1-13, 1986.
- [11] C.H. Coombs: *A Theory of Data*, Wiley, New York, 1964.

- [46] E.H. Ruspini: New experimental results in fuzzy clustering, *Information Sciences*, 6, 1973.
- [47] T.L. Saaty: A scaling method for priorities in hierarchical structures, *J. Math. Psychology*, 15, 234-281, 1977.
- [48] T.L. Saaty: Exploring the interface between hierarchies, multiple objectives and fuzzy sets, *Fuzzy Sets and Systems*, 1, 57-68, 1978.
- [49] T.L. Saaty: *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.
- [50] Michael Smithson: Fuzzy set theory and the social sciences: the scope for applications, *Fuzzy Sets and Systems*, 26, 1-21, 1988.
- [51] H. Tanaka, S. Uejima and K. Asai: Fuzzy linear regression model, Int. Congress on Applied Systems Research and Cybernetics, Acapulco, Mexico, 1980.
- [52] H. Tanaka, S. Uejima and K. Asai: Linear regression analysis with fuzzy model, *IEEE Trans. Systems Man Cybernetics*, 903-907, 1982.
- [53] H. Tanaka, T. Shimomura, J. Watada and K. Asai: Fuzzy linear regression analysis of the number of staff in local government, FIP-84, Kauai, Hawaii, July 1984.
- [54] 田中英夫、和多田淳三、林彪: ファジィ線形回帰分析の3つの定式化について, 計測自動制御学会論文集, 22, 10, 1051-1057, 1986.
- [55] H. Tanaka: Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems*, 24, 363-375, 1987.
- [56] H. Tanaka and J. Watada: Possibilistic linear systems and their application to linear regression model, *Fuzzy Sets and Systems*, 27, 1-15, 1988.
- [57] H. Tanaka, H. Hayashi and J. Watada: Possibilistic linear regression for fuzzy data, *European J. of Operational Research*, 40, 3, 389-396, 1989.
- [58] M. Tenenhaus and F.W. Young: An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and other Methods for Quantifying Categorical Multivariate Data, *Psychometrika*, 50, 1, 91-119, 1985.
- [59] E. Trauwaert: On the meaning of Dunn's partition coefficient for fuzzy clusters, *Fuzzy Sets and Systems*, 25, 217-242, 1988.
- [60] 和多田淳三、田中英夫、浅居喜代治: ファジィ数量化理論2類による購買要因の分析, 日本経営工学会誌, 9, 5, 385-392, 1981.
- [61] 和多田淳三、田中英夫、浅居喜代治: ファジィ数量化理論2類, 行動計量学, 9, 2, 24-32, 1982.
- [62] 和多田淳三、田中英夫、浅居喜代治: ヒューリスティックな手法による集団選考構造の同定, システムと制御, 26, 2, 24-32, 1982.
- [63] J. Watada, H. Tanaka and K. Asai: A heuristic method of hierarchical clustering for fuzzy intransitive relations, in: R.R. Yager ed. *Fuzzy Sets and Possibility Theory: Recent Developments*, Pergamon Press, 215-220, 1982.
- [64] 和多田淳三、田中英夫、浅居喜代治: ファジィ判別分析, 日本経営工学会誌, 34, 2, 126-131, 1983.
- [65] 和多田淳三、田中英夫、浅居喜代治: ファジィ集合による多属性代替案の評価選考構造の同定, システムの制御, 27, 6, 403-409, 1983.
- [66] 和多田淳三、田中英夫、浅居喜代治: ファジィ時系列モデルと予測問題への応用, 日本経営工学会, 34, 3, 180-186, 1983.
- [67] J. Watada: *Theory of fuzzy multivariate analysis and its applications*, Ph.D. Dissertation, University of Osaka Prefecture, 1983.
- [68] 和多田淳三、田中英夫、浅居喜代治: ファジィ数量化理論1類, 行動計量学, 11, 1, 66-73, 1984.
- [69] J. Watada, K.S. Fu, and J.T.P. Yao: Damage assessment using fuzzy multivariate analysis, *Technical Report no CE-STR-84-4, School of Civil Eng., Purdue University*, 1984.
- [70] J. Watada, H. Tanaka and K. Asai: Fuzzy discriminant analysis in fuzzy groups, *Fuzzy Sets and Systems*, 19, 261-271, 1986.
- [71] J. Watada, H. Tanaka and T. Shimomura: Identification of learning curve possibilistic concepts, In: W. Karwowski & A. Mital eds., *Applications of Fuzzy Sets Theory in Human Factors*, Elsevier Science Publisher, 143-168, 1986.
- [72] J. Watada and H. Tanaka: Fuzzy quantification methods, 2nd IFSA Congress, Tokyo 1987.
- [73] 和多田淳三: ファジィ数量化理論, 寺野、浅居、菅野編著、ファジィシステム入門, オーム社, 1987.
- [74] J. Watada, H. Tanaka and K. Asai: Analysis of time series data by possibilistic model, Int. Workshop on Fuzzy Systems Applications, Fukuoka, 20-24, 1988.
- [75] 和多田淳三: 多属性意思決定, 寺野、浅居、菅野編著、応用ファジィシステム入門, オーム社, 1989.
- [76] M.P. Windham: Cluster validity for fuzzy clustering algorithms, *Fuzzy Sets and Systems*, 3, 1-9, 1980.
- [77] M.P. Windham: Cluster validity for fuzzy c-means algorithms, *IEEE Transaction on Pattern Anal. Machine Intell.*, 4, 357-363, 1982.
- [78] L.A. Zadeh: Probability measures of fuzzy events, *J. of Mathematical Analysis and Applications*, 23, 421-427, 1968.
- [79] L.A. Zadeh: Similarity relations and fuzzy orderings, *Information Sciences*, 3, 177-200, 1971.

- [12] D. Dimitrescu: Hierarchical pattern classification, *Fuzzy Sets and Systems*, 28, 145-162, 1988.
- [13] D. Dubois and H. Prade: *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.
- [14] J.C. Dunn: A fuzzy relative of the ISODATA process and its use in detecting compact and well-separated clusters, *Journal of Cybernetics*, 3, 32-57, 1974.
- [15] J.C. Dunn: A graph theoretic analysis of pattern classification via Tamura's fuzzy Relation, *IEEE Transaction on Systems, Man, and Cybernetics*, May, 1974.
- [16] J.C. Dunn: Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems, *Journal of Cybernetics*, 4, 32-57, 1974.
- [17] J.C. Dunn: Indices of partition fuzziness and the detection of clusters in large data sets, in: M. Gupta and G. Saridis, eds., *Fuzzy Automata and Decision Processes*, American Elsevier, New York, 1976.
- [18] A.O. Esogbue: Optimal clustering of fuzzy data via fuzzy dynamic programming, *Fuzzy Sets and Systems*, 18, 283-298, 1986.
- [19] I. Gitman and M.D. Levine: An algorithm for detecting unimodal fuzzy sets and its application as a clustering techniques, *IEEE Transaction on Computer*, C-19, 7, 1970.
- [20] I.R. Goodman: Fuzzy sets and equivalence classes of random sets, in: R.R. Yager, Ed., *Fuzzy Set and Possibility Theory: Recent Developments*, Pergamon, New York, 1982.
- [21] T. Gu and B. Dubuisson: Similarity of classes and fuzzy clustering, *Fuzzy Sets and Systems*, 34, 213-221, 1990.
- [22] D.E. Gustafson and W.C. Kessell: Fuzzy clustering with a fuzzy covariance matrix, in: M.M. Gupta, Ed., *Advances in Fuzzy Set Theory and Applications*, North-Holland, Amsterdam, 605-620, 1979.
- [23] J. Halpern: Set adjacency measures in fuzzy graphs, *Journal of Cybernetics*, 5, 4, 1975.
- [24] C. Hayashi: On the quantification of qualitative data from the mathematico-statistical point of view, *Annals of the institute of Stat. II*, 1950.
- [25] C. Hayashi: On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, *Annals of the institute of Stat. Math.*, 3, pp.69 ~ 98, 1952.
- [26] 林知巳夫: 数量化の方法, 東洋経済新報社 1972.
- [27] 林知巳夫, 樋口伊佐夫, 駒沢勉: 情報処理の統計, 産業図書, 1970.
- [28] I. Hayashi and H. Tanaka: The fuzzy GMDH algorithm by possibility models and its application, *Fuzzy Sets and Systems*, 36, 245-258, 1990.
- [29] B. Heshmaty and A. Kandel: Fuzzy linear regression and its applications to forecasting in uncertain environment, *Fuzzy Sets and Systems*, 15, 2, 159-191, 1985.
- [30] 市橋秀友: ファジィ AHP, 寺野、浅居、菅野編著、応用ファジィシステム入門, オーム社, 1989.
- [31] J. Jacas: Similarity relations-The calculation of minimal generating families, *Fuzzy Sets and Systems*, 35, 151-162, 1990.
- [32] R.A. Jarvis and E.A. Patrick: Clustering using a similarity measure based on shared near neighbors, *IEEE Transaction on Computer*, C-22, 11, 1973.
- [33] A. Kandel: *Fuzzy Technique in Pattern recognition*, John Wiley & Sons, New York, 1982.
- [34] M. Kochen and A.N. Badre: On the precision of adjectives which denote fuzzy sets, *Journal of Cybernetics*, 4, 49-59, 1974.
- [35] P.J.M. van Laarhoven and W. Pedrycz: A fuzzy extension of Saaty's priority theory, *Fuzzy Sets and Systems*, 11, 229-241, 1983.
- [36] L. Lebart, A. Morineau and K.M. Warwick: *Multivariate Descriptive Statistical Analysis*, John Wiley & Sons, 1984.
- [37] E.T. Lee: An application of fuzzy sets to the classification of geometric figures and chromosome images, *Information Sciences*, 10, 1976.
- [38] K. Leszczynski, P. Penczek and W. Grochulski: Sugeno's fuzzy measure and fuzzy clustering, *Fuzzy Sets and Systems*, 15, 147-158, 1985.
- [39] G. Libert and M. Roubens: Non-metric fuzzy clustering algorithms and their cluster validity. in: M. Gupta and E. Sanchez, Eds., *Approximate Reasoning in Decision Analysis*, North-Holland, Amsterdam, 417-425, 1982.
- [40] T. Onisawa, M. Sugeno, Y. Nishiwaki, H. Kawai and Y. Harima, Fuzzy measure analysis of public attitude towards the use of nuclear energy, *Fuzzy Sets and Systems*, 20, 259-289, 1986.
- [41] 大隅昇: ファジィクラスタリング, 数理科学, 191, MAY, 34-41, 1979.
- [42] M. Roubens: Pattern classification problems and fuzzy sets, *Fuzzy Sets and Systems*, 1, 239-253, 1978.
- [43] M. Roubens: Fuzzy clustering algorithms and their cluster validity, *European J. Oper. Res.*, 10, 294-301, 1982.
- [44] E.H. Ruspini: A new approach to clustering, *Information and Control*, 15, 1969.
- [45] E.H. Ruspini: Numerical methods for fuzzy clustering, *Information Sciences*, 2, 1970.