

謝辞

本論文は著者が筑波大学大学院 修士課程 理工学研究科での修士論文の研究から引き続き、筑波大学大学院 博士課程 工学研究科在学中に行った研究成果をまとめたものです。この期間中、大変多くの方々からご指導、ご協力いただきました。心から感謝申し上げます。

筑波大学 第三学群 工学システム学類の卒業研究に始まって6年間、あたたかいご指導を頂きました筑波大学 機能工学系 宮本 定明 教授に深く感謝致します。卒業研究を始める前、クラスタリングという言葉さえも知らなかった私を、博士論文を仕上げるまでにご指導いただきました。宮本先生からは、論文の投稿、学会への参加、研究室でのディスカッションを通じて、論理的な思考と論理的な説明について学びました。また、論文作成において、筑波大学 機能工学系 安信 誠二 教授、同 鬼沢 武久 教授、電子・情報工学系 稲垣 敏之 教授、社会工学系 イリチュ(佐藤) 美佳 助教授には貴重なご意見をいただきました。ここに深く御礼申し上げます。

本研究における数値実験で用いたデータは、筑波大学 社会工学系 竹村 和久 助教授に提供して頂きました。感謝致します。

本研究について、ゼミなどを通じてたくさんの貴重なご意見を頂きました筑波大学 機能工学系 宮本研究室 の皆様、卒業生、修了生の皆様に感謝致します。特に、北陸先端科学技術 大学院大学の馬屋原 一孝 先生には、宮本研究室におられたときに、多くの助言と共に、研究の指導をして頂きました。機能工学系 長谷川 学 助教授には、同じ研究室でお世話になりました。電子・情報工学系 河辺 徹 先生には、国際学会への参加の際にとってもお世話になりました。ここに深く御礼申し上げます。

最後になりましたが、大学院の博士課程への進学を勧め、その研究生活を暖かく見守っていただきました父 治門、母 みや子に感謝致します。

に L_1 距離に基づくアルゴリズムを開発した。このアルゴリズムは、区分線形なクラスター中心に関する偏導関数を探索するアルゴリズムである。アルゴリズムの計算量は、従来のファジィc-平均法とオーダーが同じであり、計算量の面で劣るわけではない。ユークリッド距離の二乗に基づく場合は、非線形な偏導関数の探索が必要で、複雑であるため扱わなかった。

5.2 展望と今後の課題

実データに基づく手法の検証

開発した手法では、不確定性を含むデータに対して最長距離を用いた場合、最短距離を用いた場合の2種類の結果が得られる。また、各々の成分に含まれる不確定性を何らかの値で代表することにより、不確定性を含まないデータとして扱った場合の従来のクラスタリング手法を行うと、さらにもう1つの結果が得られる。これら3つの結果を比較することができる。すでに分類がわかっている不確定性を含むデータが与えられた場合には、クラスタリングを行った結果の、誤分類の数を従来の手法と本手法で比較することができる。また、データの傾向が知られているような不確定性を含むデータに対して、従来の手法と本手法におけるクラスタリング結果が得られた場合に、これらの結果におけるデータの傾向を比較することができる。このような比較により本手法の性質が明らかになっていく。このためには、実データを用いた多くの数値実験が必要となる。

他の手法への応用

ファジィc-平均法では、メンバシップに対する制約として一般的に確率制約が用いられ、本稿でもこの制約を用いた。これ以外にファジィc-平均法の制約条件を可能性制約に変えた可能性クラスタリング [38] において、不確定性を含むデータを扱うことが考えられる。このとき、ここで扱った区間データとクラスター中心あるいは、ファジィデータとクラスター中心の距離の定義はそのまま利用できる。また、非階層クラスタリングだけでなく階層クラスタリングにおいて区間データやファジィデータを扱う手法を考えることができる。

クラスター中心に対する不確定性の仮定

クラスター中心は不確定性を含まないと仮定して議論を行ってきたが、仮にクラスター中心に区間あるいはファジィ数の不確定性を許容したとしても、最適化の結果として不確定性を含むクラスター中心は得られない。ファジィ c -平均法では、クラスター中心とデータとの間の距離を小さくすることにより、目的関数は減少する。クラスター中心に不確定性を仮定して、クラスター中心と個体との間の距離を最小にする最適化を行うと、最長距離を用いた場合は、不確定性を含まないクラスター中心が最適解となり、最短距離を用いた場合は、クラスター中心の不確定性を無限大にすることで、目的関数は0になるので、最適解が得られない。つまり、本論文の枠組みで、最長距離や最短距離を用いる限りにおいて、クラスター中心に不確定性を仮定する意味はない。ただし、最長距離あるいは、最短距離以外の距離を用いることや、ファジィ c -平均法以外の別の目的関数を用いることにより、ファジィなデータをファジィクラスタリングした結果、ファジィなメンバシップとファジィなクラスター中心が得られるといった手法があるかもしれない。

不確定性を含むデータに対するデータ解析は、あまり研究されていない領域で、研究の余地が多くある。ここで述べた手法以外にも様々な手法やアルゴリズムが考えられる。また、多様な結果が得られることで、結果の解釈についても様々な可能性がある。このように、理論的にも実用的にも発展の余地が大きく、今後のさらなる研究が望まれる。

謝辞

本論文は著者が筑波大学大学院 修士課程 理工学研究科での修士論文の研究から引き続き、筑波大学大学院 博士課程 工学研究科在学中に行った研究成果をまとめたものです。この期間中、大変多くの方々からご指導、ご協力いただきました。心から感謝申し上げます。

筑波大学 第三学群 工学システム学類の卒業研究に始まって6年間、あたたかいご指導を頂きました筑波大学 機能工学系 宮本 定明 教授に深く感謝致します。卒業研究を始める前、クラスタリングという言葉さえも知らなかった私を、博士論文を仕上げるまでにご指導いただきました。宮本先生からは、論文の投稿、学会への参加、研究室でのディスカッションを通じて、論理的な思考と論理的な説明について学びました。また、論文作成において、筑波大学 機能工学系 安信 誠二 教授、同 鬼沢 武久 教授、電子・情報工学系 稲垣 敏之 教授、社会工学系 イリチュ(佐藤) 美佳 助教授には貴重なご意見をいただきました。ここに深く御礼申し上げます。

本研究における数値実験で用いたデータは、筑波大学 社会工学系 竹村 和久 助教授に提供して頂きました。感謝致します。

本研究について、ゼミなどを通じてたくさんの貴重なご意見を頂きました筑波大学 機能工学系 宮本研究室 の皆様、卒業生、修了生の皆様に感謝致します。特に、北陸先端科学技術 大学院大学の馬屋原 一孝 先生には、宮本研究室におられたときに、多くの助言と共に、研究の指導をして頂きました。機能工学系 長谷川 学 助教授には、同じ研究室でお世話になりました。電子・情報工学系 河辺 徹 先生には、国際学会への参加の際にとってもお世話になりました。ここに深く御礼申し上げます。

最後になりましたが、大学院の博士課程への進学を勧め、その研究生活を暖かく見守っていただきました父 治門、母 みや子に感謝致します。

参考文献

- [1] M. S. Aldenderfer, and R. K. Blashfield, *Cluster Analysis (Sage University Papers Series: Quantitative Applications in the Social Sciences:44)*, Sage Publications, Beverly Hills, 1984.
- [2] M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, 1973.
- [3] E. Backer, *Cluster Analysis by Optimal Decomposition of Induced Fuzzy Sets*, Delftse Universitaire Pers, 1978.
- [4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.
- [5] J. C. Bezdek (ed.), *Analysis of Fuzzy Information, Volume I Mathematics and Logic*, CRC Press, Boca Raton, 1987.
- [6] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic Publishers, Boston, 1999.
- [7] J. P. Bigus, *Data Mining with Neural Networks*, McGraw-Hill, 1996.
- [8] L. Bobrowski, and J. C. Bezdek, *c*-Means Clustering with the l_1 and l_∞ Norms, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, No. 3, pp. 545-554, 1991.
- [9] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi, *Discovering Data Mining*, Prentice Hall PTR, New Jersey, 1998.

- [10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification 2nd Edition*, John Wiley & Sons, New York, 2001.
- [11] D. Dumitrescu, B. Lazzerini, and L. C. Jain, *Fuzzy Sets and Their Application to Clustering and Training*, CRC Press, Boca Raton, 2000.
- [12] J. C. Dunn, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *Journal of Cybernetics*, Vol. 3, pp. 32-57, 1974.
- [13] J. C. Dunn, Well-Separated Clusters and Optimal Fuzzy Partitions, *Journal of Cybernetics*, Vol. 4, pp. 95-104, 1974.
- [14] B. S. Duran, and P. L. Odell, *Cluster Analysis: A Survey*, Lecture Notes in Economics and Mathematical Systems, Vol. 100, Springer, Berlin, 1974.
- [15] Y. El-Sonbaty, and M. A. Ismail, Fuzzy Clustering for Symbolic Data, *IEEE Transactions on Fuzzy Systems*, Vol. 6, No. 2, pp. 195-204, 1998.
- [16] Y. Endo, and K. Horiuchi, On Clustering Algorithm for Fuzzy Data, *Proc. of NOLTA'97*, November 29-December 2, 1997, Honolulu, U.S.A., pp. 381-384, 1997.
- [17] B. S. Everitt, *Cluster Analysis 3rd Edition*, Edward Arnold, London, 1993.
- [18] B. S. Everitt, A. J. Gourlay, and R. E. Kendell, An Attempt at Validation of Traditional Psychiatric Syndromes by Cluster Analysis, *The British Journal of Psychiatry*, Vol. 119, pp. 399-412, 1971.
- [19] E. W. Forgy, Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications, *Biometrics* Vol. 21, pp. 768-769, (Abstract) 1965.
- [20] R. E. Frank, and P. E. Green, Numerical Taxonomy in Marketing Analysis: A Review Article, *Journal of Marketing Research*, Vol. 5, pp. 83-98, 1968.
- [21] M. Goodfellow, Numerical Taxonomy of Some Nocardioform Bacteria, *Journal of General Microbiology*, Vol. 69, pp. 33-80, 1971.

- [22] A. D. Gordon, *Classification 2nd Edition*, Chapman & Hall/CRC, Boca Raton, 1999.
- [23] K. C. Gowda, and E. Diday, Symbolic Clustering using a New Dissimilarity Measure, *Pattern Recognition*, Vol. 24, No. 6, pp. 567-578, 1991.
- [24] K. C. Gowda, and E. Diday, Symbolic Clustering using a New Similarity Measure, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 2, pp. 368-378, 1992.
- [25] K. C. Gowda, and T. V. Ravi, Divisive Clustering of Symbolic Objects using the Concepts of Both Similarity and Dissimilarity, *Pattern Recognition*, Vol. 28, No. 8, pp. 1277-1282, 1995.
- [26] P. E. Green, R. E. Frank, and P. J. Robinson, Cluster Analysis in Test Market Selection, *Management Science*, Vol. 13, No. 8, pp. B387-B400, 1967.
- [27] J. A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, New York, 1975.
- [28] R. J. Hathaway, J. C. Bezdek, and W. Pedrycz, A Parametric Model for Fusing Heterogeneous Fuzzy Data, *IEEE Transactions on Fuzzy Systems*, Vol. 4, No. 3, pp. 270-281, 1996.
- [29] F. G. J. Hayhoe, D. Quaglino, and R. Doll, *The Cytology and Cytochemistry of Acute Leukaemias*, Medical Research Council Special Report Series No. 304, Her Majesty's Stationery Office, London, 1964.
- [30] F. R. Hodson, Searching for Structure within Multivariate Archaeological Data, *World Archaeology*, Vol. 1, pp. 90-105, 1969.
- [31] F. R. Hodson, Cluster Analysis and Archaeology: Some New Developments and Applications, *World Archaeology*, Vol. 1, pp. 299-320, 1969.
- [32] F. R. Hodson, D. G. Kendall, and P. Tautu, (eds.), *Mathematics in the Archaeological and Historical Sciences*, Edinburgh University Press, Edinburgh, 1971.

- [33] A. K. Jain, and R. C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, New Jersey, 1988.
- [34] K. Jajuga, L_1 -Norm Based Fuzzy Clustering, *Fuzzy Sets and Systems*, Vol. 39, pp. 43-50, 1991.
- [35] M. Jambu, and M.-O. Lebeaux, *Cluster Analysis and Data Analysis*, North-Holland, Amsterdam, 1983.
- [36] 樺山 忠雄, アナリテイカル・マーケティング, 創成社, 1996.
- [37] L. Kaufman, and P. J. Rousseeuw, *Finding Groups in Data, An Introduction to Cluster Analysis*, John Wiley & Sons, New York, 1990.
- [38] R. Krishnapuram, and J. M. Keller, A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, pp. 98-110, 1993.
- [39] K.-M. Lee, K.-M. Lee, J.-H. Lee, and H. Lee, Fuzzy Clustering for Fuzzy Data using Dissimilarity Measure, *Proc. of IFSA '99*, August 17-20, 1999, Taipei, Taiwan ROC, Vol. II, pp. 879-883, 1999.
- [40] J. B. MacQueen, Some Methods of Classification and Analysis of Multivariate Observations, *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281-297, 1967.
- [41] B. Mirkin, *Mathematical Classification and Clustering*, Kluwer Academic Publishers, Dordrecht, 1996.
- [42] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, Dordrecht, 1990.
- [43] S. Miyamoto, and Y. Agusta, An Efficient Algorithm for l_1 Fuzzy c -Means and Its Termination, *Control and Cybernetics*, Vol. 24, No. 4, pp. 421-436, 1995.
- [44] S. Miyamoto, and M. Mukaidono, Fuzzy c -Means as a Regularization and Maximum Entropy Approach, *Proc. of IFSA '97*, June 25-30, 1997, Prague, Czech, Vol. II, pp. 86-92, 1997.

- [45] S. Miyamoto, O. Takata, and K. Umayahara, Handling Missing Values in Fuzzy c -Means, *Proc. of the 3rd Asian Fuzzy Systems Symposium*, June 18-21, 1998, Masan, Korea, pp. 139-142, 1998.
- [46] 宮本 定明, クラスター分析入門, 森北出版, 1999.
- [47] 宮本 定明, 階層的クラスター分析の方法-I, システム/制御/情報, Vol. 33, No. 8, pp. 424-430, 1989.
- [48] 宮本 定明, 馬屋原 一孝, 向殿 政男, ファジィ c -平均法とエントロピー正則化法におけるファジィ分類関数, 日本ファジィ学会誌, Vol. 10, No. 3, pp. 548-557, 1998.
- [49] W. Pedrycz, J. C. Bezdek, R. J. Hathaway, and G. W. Rogers, Two Nonparametric Models for Fusing Heterogeneous Fuzzy Data, *IEEE Transactions on Fuzzy Systems*, Vol. 6, No. 3, pp. 411-425, 1998.
- [50] H. C. Romesburg, *Cluster Analysis for Researchers*, Robert E. Krieger, Florida, 1990.
- [51] E. H. Ruspini, A New Approach to Clustering, *Information and Control*, Vol. 15, pp. 22-32, 1969.
- [52] 坂和 正敏, ファジィ理論の基礎と応用, 森北出版, 1989.
- [53] M. Sato, and Y. Sato, Fuzzy Clustering Model for Fuzzy Data, *Proc. of the 4th IEEE Conf. on Fuzzy Systems*, March 20-24, 1995, Yokohama, Japan, pp. 2123-2128, 1995.
- [54] M. Sato, Y. Sato, and L. C. Jain, *Fuzzy Clustering Models and Applications*, Physica-Verlag, Heidelberg, 1997.
- [55] R. R. Sokal, and P. H. A. Sneath, *Principles of Numerical Taxonomy*, W. H. Freeman, San Francisco, 1963.
- [56] H. Späth, *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, Ellis Horwood, Chichester, 1980.

- [57] 高橋 渉, 非線型関数解析学, 近代科学社, 1988.
- [58] O. Takata, and S. Miyamoto, Optimal Clustering using Metaheuristics and Automatic Evaluation for Estimating Parameters in the Objective Function, *Proc. of the 4th Asian Fuzzy Systems Symposium*, May 31-June 3, 2000, Tsukuba, Japan, pp. 493-497, 2000.
- [59] O. Takata, S. Miyamoto, and K. Umayahara, Clustering of Data with Uncertainties using Hausdorff Distance, *Proc. of the 2nd IEEE International Conference on Intelligent Processing Systems*, August 4-7, 1998, Gold Coast, Australia, pp. 67-71, 1998.
- [60] 高田 治, 宮本 定明, 区間データに基づく不確定性を含むデータのファジィクラスタリング, 日本ファジィ学会誌, Vol. 12, No. 5, pp. 686-695, 2000.
- [61] 竹村 和久, 意思決定の心理, 福村出版, 1996.
- [62] 田中 英夫, ファジィモデリングとその応用, 朝倉書店, 1990.
- [63] 田中 英夫, 石渕 久生, ソフトデータ解析, 朝倉書店, 1995.
- [64] P. Winkel, N. Tygstrup, and The Copenhagen Study Group for Liver Diseases, Classification of Cirrhosis. The Resolution of Data Modes and their Recovery in an Independent Material, *Computers and Biomedical Research*, Vol. 4, pp. 417-426, 1971.
- [65] L. A. Zadeh, Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338-353, 1965.

関連発表論文

学術雑誌に発表した論文

1. 高田 治, 宮本 定明,
区間データに基づく不確定性を含むデータのファジィクラスタリング,
日本ファジィ学会誌, Vol. 12, No. 5, pp. 686-695, 2000.
2. 高田 治, 宮本 定明,
ファジィデータに対する L_1 距離を用いたファジィクラスタリング,
日本ファジィ学会誌, Vol. 13, No. 6, pp. 689-698, 2001.

国際会議で発表した論文 (査読付き)

1. Sadaaki Miyamoto, Osamu Takata, and Kazutaka Umayahara,
Handling Missing Values in Fuzzy c -Means,
Proceedings of the 3rd Asian Fuzzy Systems Symposium,
June 18-21, 1998, Masan, Korea, pp. 139-142, 1998.
2. Osamu Takata, Sadaaki Miyamoto, and Kazutaka Umayahara,
Clustering of Data with Uncertainties using Hausdorff Distance,
Proceedings of the 2nd IEEE International Conference on Intelligent Processing Systems,
August 4-7, 1998, Gold Coast, Australia, pp. 67-71, 1998.
3. Osamu Takata, and Sadaaki Miyamoto,
Optimal Clustering using Metaheuristics and Automatic Evaluation for Esti-

mating Parameters in the Objective Function,
Proceedings of the 4th Asian Fuzzy Systems Symposium,
 May 31-June 3, 2000, Tsukuba, Japan, pp. 493-497, 2000.

4. Osamu Takata, Sadaaki Miyamoto, and Kazutaka Umayahara,
 Fuzzy Clustering of Data with Uncertainties using Minimum and Maximum
 Distances based on L_1 Metric,
*Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International
 Conference,*
 July 25-28, 2001, Vancouver, Canada, pp. 2511-2516, 2001.