



دانشکده فنی و مهندسی
گروه مهندسی کامپیوتر

خوشه‌بندی ترکیبی فازی مبتنی بر الگوریتم ژنتیک

پایان‌نامه برای دریافت درجه کارشناسی ارشد رشته مهندسی کامپیوتر- هوش مصنوعی

استاد راهنما:

جناب آقای دکتر محمدعلی بادامچی زاده

استاد مشاور:

جناب آقای مهندس کریم صمد زمینی

پژوهشگر

نگار راضی زاده

سال ۱۳۹۱

چکیده

در این پایان نامه سعی شده است، تا با استفاده از منطق فازی و الگوریتم ژنتیک با مسئله خوشه‌بندی ترکیبی برخورد شود. با استفاده از منطق فازی می‌توان اطلاعات بیشتری را از مجمع خوشه‌ها استخراج کرد که این امر می‌تواند منجر به ترکیب بهتر خوشه‌ها شود. همچنین الگوریتم ژنتیک می‌تواند با کاوش وسیع‌تر فضای مسئله نتیجه بهتری را نسبت به سایر روش‌های خوشه‌بندی ترکیبی ارائه کند. در حل مسئله با الگوریتم ژنتیک دو محور کلی می‌تواند مورد نظر قرار گیرد: الف) کد کردن مسئله خوشه‌بندی ترکیبی به گونه‌ای که بتوان با استفاده از الگوریتم ژنتیک آن را حل کرد. ب) طراحی تابع ارزیابی به گونه‌ای که راهنمای خوبی برای یافتن نتیجه بهتر برای مسئله خوشه‌بندی ترکیبی شود.

در این پایان‌نامه از نمایش رشته‌ای برای کد کردن مسئله استفاده شده است. همچنین یک تابع برازندگی فازی جدید برای الگوریتم ژنتیک طراحی شده است که به طور موثری منجر به کارایی بالای بهینه‌سازی می‌شود. نتایج آزمایشات نیز که روی مجموعه داده‌های استاندارد مختلف پیاده‌سازی شده است، نشان از کارایی بالای روش پیشنهادی دارند.

واژه های کلیدی: خوشه‌بندی، خوشه‌بندی ترکیبی، الگوریتم ژنتیک، منطق فازی.

فهرست مطالب

صفحه	عنوان
۱	فصل ۱
۱	مقدمه
۲	۱-۱ خوشه‌بندی داده‌ها
۴	۲-۱ خوشه‌بندی ترکیبی
۶	۱-۲-۱ ایجاد تنوع
۶	۲-۲-۱ تابع توافقی
۷	۳-۱ الگوریتم‌های خوشه‌بندی پایه
۸	۱-۳-۱ الگوریتم‌های خوشه‌بندی سلسله‌مراتبی
۱۲	۴-۱ ترکیب نتایج خوشه‌بندی‌ها
۱۳	۱-۴-۱ ماتریس همبستگی
۱۴	۲-۴-۱ الگوریتم ژنتیک
۱۶	فصل ۲
۱۶	مروری بر کارهای انجام شده
۱۷	۱-۲ مقدمه
۱۷	۲-۲ کارهای انجام شده پایه‌ای
۱۹	۳-۲ کارهای انجام شده با استفاده از مفهوم پایداری خوشه
۲۲	۴-۲ کارهای انجام شده در بهینه‌سازی
۲۳	فصل ۳
۲۳	روش پیشنهادی
۲۴	۱-۳ مقدمه
۲۴	۲-۳ چارچوب اصلی روش پیشنهادی
۲۷	۱-۳-۲ متغیرهای فازی
۲۹	۲-۳-۳ معرفی تابع برازندگی برای الگوریتم ژنتیک
۳۰	عملگر Mutation (جهش)
۳۱	عملگر crossover (ترکیب)
۳۲	عملگر selection (انتخاب)

۳۳	فصل ۴.....
۳۳	نتایج شبیه سازی.....
۳۴	۱-۴ مقدمه.....
۳۴	۲-۴ مجموعه داده.....
۳۶	۳-۴ پارامترهای شبیه سازی.....
۳۷	۴-۴ نتایج شبیه سازی.....
۴۰	فصل ۵.....
۴۰	نتیجه گیری و پیشنهادات.....
۴۳	پیشنهادات.....

مراجع

- [۱] داودی س. ر. ، محمدی م. ، رحمانی ع. *خوشه‌بندی ترکیبی توزیع شده به روش K-means* دوازدهمین کنفرانس بین‌المللی انجمن کامپیوتر، تهران، ایران. ۱۳۸۵
- [۲] علیزاده ح.، مینایی بیدگلی ب. *بررسی روش‌های ارزیابی خوشه‌بندی*، کنفرانس ملی مهندسی نرم‌افزار و کاربردهای آن، لاهیجان، ایران. ۱۳۸۷
- [۳] علیزاده ح.، مشکي م.، پروین ح. مینایی بیدگلی ب.، *خوشه‌بندی ترکیبی مبتنی بر زیرمجموعه‌ای از نتایج اولیه*، مجله پردازش علائم و داده‌ها، مرکز تحقیقات پردازش هوشمند علائم، ایران. ۱۳۸۹
- [۴] عظیمی ج. *بررسی پراکندگی در خوشه‌بندی ترکیبی*، پایان نامه کارشناسی ارشد، دانشگاه علم و صنعت ایران. ۱۳۸۶
- [۵] محمدی م. ، عظیمی ج.، داودی ر. ، رحمانی ع. *ارایه یک روش بر مبنای الگوریتم‌های تکاملی برای خوشه‌بندی ترکیبی*، دوازدهمین کنفرانس بین‌المللی انجمن کامپیوتر، تهران، ایران. ۱۳۸۵
- [۶] Brandsma T. and Buishand T.A., “*Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling*”, *Hydrology And Earth System Sciences* ۲, pp. ۱۹۵-۲۰۹. (۱۹۹۸)
- [۷] Alizadeh H., Minaei-Bidgoli B., Parvin H: “*A New Criterion for Clusters Validation*”, *Eann/aiai* ۲۰۱۱, Part II, IFIP AICT ۳۶۴, pp. ۱۱۰-۱۱۵. ۲۰۱۱
- [۸] Alizadeh H., Parvin H., Moshki M., Minaei-Bidgoli B.: “*A New Clustering Ensemble Framework*”, *INCT, CCIS* ۲۴۱, pp. ۲۱۶-۲۲۴. ۲۰۱۱
- [۹] Ayad H. and Kamel M., *Cluster-based cumulative ensembles*. In N. Oza and R. Polikar, editors. *The 7th INTL. Workshop On Multiple Classifier Systems*, pp. ۲۳۶-۲۴۵. LNCS ۳۵۴۱. ۲۰۰۵
- [۱۰] Baumgartner R., Somorjai R., Summers R., Richter W., Ryner L., and Jarmasz M., *Resampling as a Cluster Validation Technique in fMRI*, *Journal Of Magnetic Resonance Imaging* ۱۱: pp. ۲۲۸-۲۳۱. (۲۰۰۰)
- [۱۱] Banfield C.F. “*Ultrametric Distances for a Single Linkage Dendrogram*”, *JSTOR: Applied Statistics, Statistical Algorithms*, Vol. ۲۵, No. ۳, pp. ۳۱۳-۳۱۵. (۱۹۷۶)
- [۱۲] Ben-Hur A., Elisseeff A. and Guyon I., *A stability based method for discovering structure in clustered data*, In *Pacific Symposium On Biocomputing*, VOL. ۷, pp. ۶-۱۷. (۲۰۰۲)
- [۱۳] Alizadeh H., Minaei-Bidgoli B., Parvin H., Moshki M: *A New Asymmetric Criterion for Cluster Validation*. *CIarp* ۲۰۱۱, LNCS ۷۰۴۲, pp. ۳۲۰-۳۳۰. ۲۰۱۱
- [۱۴] Breckenridge J., *Replicating cluster analysis: Method, consistency and validity*, *Multivariate Behavioral Research*. (۱۹۸۹)
- [۱۵] Breiman, L. “*Arcing classifiers*”, *The Annals Of Statistics*, ۲۶(۳), pp. ۸۰۱-۸۴۹. . (۱۹۹۸)
- [۱۶] Brossier G.. *Piecewise hierarchical clustering*, *Journal Of Classification, Springer NEW YORK*, Vol. ۷, No. ۲, pp. ۱۹۷-۲۱۶. ۱۹۹۰

- [17] Castro E. V. and Yang J., *Cluster Validity Using Support Vector Machines*, DaWaK 2007, LNCS 4737, pp. 244-256, 2007
- [18] Christou I. T., *Coordination of Cluster Ensembles via Exact Methods*, IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL. 33, NO. pp. 12-23, 2011
- [19] Das A.K. and Sil J., *Cluster Validation using Splitting and Merging Technique*, IN . Of INT. Conference. On Computational Intelligence And Multimedia Applications, ICCIMA. 2007
- [20] Duda, R.O., Hart, P.E. and Stork D.G. *Pattern Classification*. 2nd Edition, John Wiley & Sons, Inc., New York NY. 2001
- [21] Fern, X. and Brodley, C. E. “*Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach*”, In . 20th INT. Conference. On Machine Learning, ICMLpp. 13-17, 2003.
- [22] Fern X. and Lin W., *Cluster Ensemble Selection*, *SIAM International Conference on Data Mining (SDM'04)*, pp. 123-140, 2004
- [23] Fred, A. and Jain, A. K. *Combining multiple clusterings using evidence accumulation*. IEEE Transactions. on Pattern Analysis and Machine Intelligence, pp. 133-152, 2000
- [24] Fred A. and Jain A.K., *Learning Pairwise Similarity for Data Clustering*, IN . OF THE 18TH INT. Conference. On Pattern Recognition (ICPR'07). 2007
- [25] Fridlyand J. and Dudoit S.. *Applications of resampling methods to estimate the number of clusters and to improve the accuracy of a clustering method*. STAT. Berkeley Tech Report. NO. 700, 2001
- [26] Inokuchi R., Nakamura T. and Miyamoto S., *Kernelized Cluster Validity Measures and Application to Evaluation of Different Clustering Algorithms*, In . Of The IEEE INT. Conference. On Fuzzy Systems, Canada, JULY 16-21, 2006
- [27] Jain A.K. and Dubes R.C., *Algorithms for Clustering Data*. Prentice Hall. 1988
- [28] Jain, A.K. *Data clustering: 20 years beyond K-means*. Pattern Recognition Lett. 2009
- [29] Kuncheva L.I. and Hadjitodorov S. *USING DIVERSITY IN CLUSTER ENSEMBLES*. In . Of IEEE INTL. Conference On Systems, Man And Cybernetics, pp. 1214-1219, 2004.
- [30] Lange T., Braun M.L., Roth V., and Buhmann J.M.. *Stability-based model selection*. In Advances In Neural Information essing Systems 10. MIT PRESS. 2003
- [31] Lange T., Roth V., Braun M.L., and Buhmann J.M.. *Stability-based validation of clustering solutions*. *Neural Computation*, 16(7):1299-1323, 2004
- [32] Lapointe F.J. and Legendre P. *The generation of random ultrametric matrices representing dendrograms*. *Journal Of Classification*, Springer New York, VOL. 8, NO. 2, pp 177-200, 1991
- [33] Law M.H.C., Topchy A.P., and Jain A.K.. *Multiobjective data clustering*. In . of IEEE Conference On Computer Vision And Pattern Recognition, VOL 2, pp. 424-430, Washington D.C. 2004
- [34] Levine E., Domany E., Resampling Method for Unsupervised Estimation of Cluster Validity. *Neural Computation* 12:pp. 2073-2093, 2001
- [35] Luxburg U.V. and Ben-David S., *Towards a statistical theory of clustering*, *Technical Report, Pascal Workshop On Clustering, London*. 2000

- [36] Minaei-Bidgoli, B., Punch, W. F. “*Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System*”, Gecco, pp. 2202-2213, 2003
- [37] Moller U., Radke D., *A Cluster Validity Approach based on Nearest-Neighbor Resampling*, In . Of The 14TH INT. Conference. On Pattern Recognition (ICPR'07), 2007
- [38] Parvin H., Minaei-Bidgoli B., Alizadeh H. “*A New Clustering Algorithm With The Convergence Proof*” , Part I, LNAI 6881, pp. 21-31, 2011
- [39] Rakhlin A. and Caponnetto A. *Stability of k-means clustering*, In *Advances In Neural Information Processing Systems 19*, Cambridge, MA, 2007
- [40] Roth V., Lange T., Braun M., and Buhmann J. *A Resampling Approach to Cluster Validation*, b Coneference. On Computational Statistics, Compstat, 2002
- [41] Roth V., Braun M.L., Lange T., and Buhmann J.M., *Stability-Based Model Order Selection in Clustering with Applications to Gene Expression Data*, ICANN 2002, LNCS 2410, pp. 607-612, 2002
- [42] Roth V. and Lange T., *Feature Selection in Clustering Problems*, In *Advances In Neural Information Processing Systems*, 2004
- [43] Shamiry O., Tishby N, *Cluster Stability for Finite Samples*, 1ST Annual Conference On Neural Information Processing Systems, 2007
- [44] Singh V., Mukherjee L., Peng J., Xu J, *Ensemble clustering using semidefinite programming with applications*, MACH LEARN 99: 177-200, 2000
- [45] Strehl A. and Ghosh J., *Cluster ensembles - a knowledge reuse framework for combining multiple partitions*. *Journal Of Machine Learning Research*, 3(Dec):pp 583-617, 2002
- [46] Topchy, A., Jain, A. K. and Punch, W. F. “*Combining Multiple Weak Clusterings*”, . 7D IEEE INTL. Conference. On Data Mining, PP. 331-338, 2003
- [47] Xie X.L., Beni G., *A Validity measure for Fuzzy Clustering*, *IEEE Transaction. On Pattern Analysis And Machine Intelligence*, Vol. 13, No. 4, pp. 841-846, 1991
- [48] Minaei-Bidgoli B., Parvin H., Alinejad H., Alizadeh H., Punch W., *Effects of resampling method and adaptation on clustering ensemble efficacy*, *Artif Intell Rev*, Published Online, DOI 10.1007/s10462-011-9290-x, 2011
- [49] Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*. Addison Wesley. p. 41, 1989