Smiti, Abir; **Elouedi, Zied**

WCOID-DG: an approach for case base maintenance based on weighting, clustering, outliers, internal detection and DBsan-Gmeans. (English) Zbl 1311.68155

J. Comput. Syst. Sci. 80, No. 1, 27-38 (2014).

Summary: The success of the Case Based Reasoning system depends on the quality of the case data and the speed of the retrieval process that can be costly in time, especially when the number of cases gets bulky. To guarantee the system's quality, maintaining the contents of a case base (CB) becomes unavoidably. In this paper, we propose a novel case base maintenance policy named WCOID-DG: Weighting, Clustering, Outliers and Internal cases Detection based on Dbscan and Gaussian means. Our WCOID-DG policy uses in addition to feature weights and outliers detection methods, a new efficient clustering technique, named DBSCAN-GM (DG) which is a combination of DBSCAN and Gaussian-Means algorithms. The purpose of our WCOID-GM is to reduce both the storage requirements and search time and to focus on balancing case retrieval efficiency and competence for a CB. WCOID-GM is mainly based on the idea that a large CB with weighted features is transformed to a small CB with improving its quality. We support our approach with empirical evaluation using different benchmark data sets to show its competence in terms of shrinking the size of the CB and the research time, as well as, getting satisfying classification accuracy.

MSC:

68T37 Reasoning under uncertainty in the context of artificial intelligence
68T05 Learning and adaptive systems in artificial intelligence

Keywords:

case based reasoning; case base maintenance; Gaussian-means clustering; density based clustering; outliers detection

Software:

R; Stata; UCI-ml

Full Text: DOI

References:

[1] Aamodt, A.; Plaza, E., Case-based reasoning: foundational issues, methodological variations, and system approaches, AI Commun., 7, 1, 39-52, (1994)

[2] Leake, D. B., Case-based reasoning: experiences, lessons and future directions, (1996), MIT Press Cambridge, MA, USA

[3] Varshavskii, P.; Eremeev, A., Analogy-based search for solutions in intelligent systems of decision support integrated models and flexible calculations in artificial intelligence, J. Comput. Syst. Sci. Int., 44, 1, 90-101, (2005) · Zbl 1126.68606

[4] Vagin, V. N.; Oskin, P. V., Multiagent simulation subsystem of diagnostic complexes based on device models, J. Comput. Syst. Sci. Int., 45, 6, 970-982, (2006) · Zbl 1263.68165

[5] Arshadi, N.; Jurisica, I., Data mining for case-based reasoning in high-dimensional biological domains, IEEE Trans. Knowl. Data Eng., 17, 1127-1137, (2005)

[6] Sun, Z.; Finnie, G., A unified logical model for cbr-based e-commerce systems, Int. J. Intell. Syst., 1-28, (2005)

[7] Leake, D. B.; Wilson, D. C., Maintaining case-based reasoners: dimensions and directions, Comput. Intell., 17, 196-213, (2001)

[8] Q. Yang, J. Wu, Keep it simple: A case-base maintenance policy based on clustering and information theory, in: Canadian Conference on Artificial Intelligence, 2000, pp. 102-114. · Zbl 0987.68745

[9] G. Cao, S.C.K. Shiu, X. Wang, A fuzzy-rough approach for case base maintenance, in: International Conference on Case Based Reasoning, 2001, pp. 118-130. · Zbl 0982.68509

[10] M.K. Haouchine, B. Chebel-Morello, N. Zerhouni, Competence-preserving case-deletion strategy for case-base maintenance, in: 9th European Conference on Case-Based Reasoning, 2008, pp. 171-184.

[11] B. Smyth, M.T. Keane, Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems, in: 14th International Joint Conference on Artificial Intelligence, 1995, pp. 377-382.

[12] B. Smyth, Case-base maintenance, in: 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE, 1998, pp. 507-516.

[13] Smiti, A.; Elouedi, Z., Wcoid: maintaining case-based reasoning systems using weighting, clustering, outliers and internal cases detection, (International Conference on Intelligent Systems Design and Applications (ISDA), (2011), IEEE Computer Society), 356-361

[14] Asuncion, A.; Newman, D., UCI machine learning repository, (2007)

[15] Smiti, A.; Elouedi, Z., Overview of maintenance for case based reasoning systems, Int. J. Comput. Appl., 32, 49-56, (2011)

[16] Aha, D. W.; Kibler, D.; Albert, M. K., Instance-based learning algorithms, Mach. Learn., 37-66, (1991)

[17] Salamó, M.; Golobardes, E., Hybrid deletion policies for case base maintenance, (The Florida Artificial Intelligence Research Society FLAIRS-2003, (2003), AAAI Press), 150-154

[18] Leake, D. B.; Wilson, D. C., Remembering why to remember: performance-guided case-base maintenance, (Proceedings of the 5th European Workshop on Advances in Case-Based Reasoning, vol. 1898, (2000), Springer-Verlag), 161-172

[19] Shiu, S. C.K.; Yeung, D. S.; Sun, C. H.; Wang, X., Transferring case knowledge to adaptation knowledge: an approach for case-base maintenance, Comput. Intell., 17, 2, 295-314, (2001)

[20] Smiti, A.; Elouedi, Z., Coid: maintaining case method based on clustering, outliers and internal detection, (Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2010, SNPD'10, vol. 295, (2010), Springer Berlin/Heidelberg), 39-52

[21] E. McKenna, B. Smyth, A competence model for case-based reasoning, in: 9th Irish Conference on Artificial Intelligence and Cognitive Science, 1998, pp. 208-220.

[22] Markovitch, S.; Scott, P. D., The role of forgetting in learning, (Fifth International Conference on Machine Learning, (1988), Morgan Kaufmann), 459-465

[23] Minton, S., Qualitative results concerning the utility of explanation-based learning, Artificial Intelligence, 42, 363-391, (1990)

[24] Chou, C. H.; Kuo, B. H.; Chang, F., The generalized condensed nearest neighbor rule as a data reduction method, Int. Conf. Pattern Recognit., 2, 556-559, (2006)

[25] Manry, M. Y.J.; Wilson, D., Prototype classifier design with pruning, Int. J. Artif. Intell. Tools, 261-280, (2005)

[26] Wilson, D. L., Asymptotic properties of nearest neighbor rules using edited data, IEEE Trans. Syst. Man Cybern., 2, 3, 408-421, (1972) · Zbl 0276.62060

[27] Aha, D. W., Feature weighting for lazy learning algorithms, vol. SECS 453, (1998), Kluwer Academic Boston

[28] Manoharan, C.; Lakshmi, N. R., Classification of micro calcifications in mammogram using combined feature set with svm, Int. J. Comput. Appl., 11, 10, 30-34, (2010)

[29] Ester, M.; Kriegel, H. Peter; Xu, J. S.X., A density-based algorithm for discovering clusters in large spatial databases with noise, 226-231, (1996), AAAI Press

[30] Smiti, A.; Elouedi, Z., Dbscan-gm: an improved clustering method based on Gaussian means and dbscan techniques, (International Conference on Intelligent Engineering Systems (INES), (2012), IEEE Computer Society), 573-578

[31] Hamerly, G.; Elkan, C., Learning the \textit{k} in \textit{k}-means, vol. 17, (2003), MIT Press

[32] MacQueen, J. B., Some methods for classification and analysis of multivariate observations, (Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, (1967), University of California Press), 281-297

[33] Li, Y.; Zhao, K.; Chu, X.; Liu, J., Speeding up \textit{k}-means algorithm by gpus, J. Comput. System Sci., 115-122, (2010)

[34] Moon, T. K., The expectation-maximization algorithm, IEEE Signal Process. Mag., 13, 6, 47-60, (1996)

[35] M. Dash, H. Liu, X. Xu, '$1 + 1 > 2$': Merging distance and density based clustering, in: International Conference on Database Systems for Advanced Applications, 2001, pp. 32-39.

[36] Muenchen, A. R.; Hilbe, J. M., R for stata users (statistics and computing), (2010), Springer · Zbl 1269.68001

[37] Bussian, B. M.; Härdle, W., Robust smoothing applied to white noise and single outlier contaminated Raman spectra, Appl. Spectrosc., 38, 3, 309-313, (1984)

[38] Filzmoser, P.; Garrett, R. G.; Reimann, C., Multivariate outlier detection in exploration geochemistry, Comput. Geosci., 31, 579-587, (2005)

[39] M. Wölfel, H.K. Ekenel, Feature weighted Mahalanobis distance: Improved robustness for Gaussian classifiers, in: 13th European Signal Processing Conference, EUSIPCO, Citeseer, 2005, pp. 208-212.

[40] MacQueen, J. B., Some methods for classification and analysis of multivariate observations, (Cam, L. M.L.; Neyman, J., Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, (1967), University of California Press), 281-297

[41] Sabharwal, Y.; Sharma, N.; Sen, S., Nearest neighbors search using point location in balls with applications to approximate Voronoi decompositions, J. Comput. System Sci., 72, 6, 955-977, (2006) · Zbl 1100.68632

[42] Smiti, A.; Chelly, Z.; Elouedi, Z., Coid-fdcm: the fuzzy maintained dendritic cell classification method, (Artificial Intelligence and Soft Computing, Lecture Notes in Comput. Sci., vol. 7268, (2012), Springer), 233-241