Le Thi, Hoai An; **Nguyen, Manh Cuong**

DCA based algorithms for feature selection in multi-class support vector machine. (English)

Zbl 1404.68111

Ann. Oper. Res. 249, No. 1-2, 273-300 (2017).

Summary: This paper addresses the problem of feature selection for Multi-class Support Vector Machines. Two models involving the ℓ_0 (the zero norm) and the ℓ_2-ℓ_0 regularizations are considered for which two continuous approaches based on DC (Difference of Convex functions) programming and DCA (DC Algorithms) are investigated. The first is DC approximation via several sparse inducing functions and the second is an exact reformulation approach using penalty techniques. Twelve versions of DCA based algorithms are developed on which empirical computational experiments are fully performed. Numerical results on real-world datasets show the efficiency and the superiority of our methods versus one of the best standard algorithms on both feature selection and classification.

MSC:

68T05 Learning and adaptive systems in artificial intelligence

90C26 Nonconvex programming, global optimization

Keywords:

feature selection; MSVM; DC programming; DCA; DC approximation; exact penalty

Full Text: DOI

References:

[1] Bradley, PS; Mangasarian, OL; Shavlik, J (ed.), Feature selection via concave minimization and support vector machines, 82-90, (1998), San Francisco

[2] Cai, X., Nie, F., Huang, H., & Ding, C. (2011). Multi-class $\boxtimes _{2,1}$-norm support vector machine. In \textit{Data mining (ICDM), 2011 IEEE 11th International Conference} (pp. 91-100). · Zbl 0913.65054

[3] Candès, EJ; Wakin, MB; Boyd, SP, Enhancing sparsity by reweighted $\boxtimes _{1}$ minimization, Journal of Fourier Analysis and Applications, 14, 877-905, (2008) · Zbl 1176.94014 · doi:10.1007/s00041-008-9045-x

[4] Chapelle, O. (2008). \textit{Multi-class feature selection with support vector machines}. Technical report YR-2008-002. · Zbl 1108.62059

[5] Chen, Y. W., & Lin, C. J. (2006). Combining SVMs with various feature selection strategies. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), \textit{Feature extraction. Studies in Fuzziness and Soft Computing} (Vol. 207, pp. 315-324). Berlin: Springer. · Zbl 1102.68605

[6] Chen, Y., Li, Y., Cheng, X-Q., & Guo, L. (2006). Survey and taxonomy of feature selection algorithms in intrusion detection system. In \textit{Proceedings of Inscrypt 2006, LNCS 4318} (pp. 153-167). · Zbl 1172.94613

[7] Chen, X; Zeng, X; Alphen, DV, Multi-class feature selection for texture classification, Pattern Recognition Letters, 27, 1685-1691, (2006)· doi:10.1016/j.patrec.2006.03.013

[8] Collobert, R; Sinz, F; Weston, J; Bottou, L, Large scale transductive svms, Journal of Machine Learning Research, 7, 1687-1712, (2006) · Zbl 1222.68173

[9] Deng, S; Xu, Y; Li, L; Li, X; He, Y, A feature-selection algorithm based on support vector machine-multiclass for hyperspectral visible spectral analysis, Journal of Food Engineering, 119, 159-166, (2013)· doi:10.1016/j.jfoodeng.2013.05.024

[10] Duan, KB; Rajapakse, JC; Wang, H; Azuaje, F, Multiple SVM-RFE for genne selection in cancer classification with expression data, IEEE Transactions on Nanobioscience, 4, 228-234, (2005)· doi:10.1109/TNB.2005.853657

[11] Fan, J; Li, R, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association, 96, 1348-1360, (2001) · Zbl 1073.62547 · doi:10.1198/016214501753382273

[12] Gribonval, R; Nielsen, M, Sparse representation in union of bases, IEEE Transactions on Information Theory, 49, 3320-73325, (2003) · Zbl 1286.94032 · doi:10.1109/TIT.2003.820031

[13] Guyon, I; Elisseeff, A, An introduction to variable and feature selection, Journal of Machine Learning Research, 3, 1157-1182, (2003) · Zbl 1102.68556

[14] Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2006). \textit{Feature extraction, foundations and applications}. Berlin: Springer. · Zbl 1114.68059 · doi:10.1007/978-3-540-35488-8

[15] Hermes, L., & Buhmann, J. M. (2000). Feature selection for support vector machines. \textit{Proceedings of the 15th International Conference on Pattern Recognition, vol. 2} (pp. 712-715). · Zbl 1446.62179

[16] Hsu, CW; Lin, CJ, A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks, 13, 415-425, (2002)· doi:10.1109/72.991427

[17] Huang, J; Ma, S; Zhang, CH, Adaptive lasso for sparse high-dimentional regression models, Statistica Sinica, 18, 1603-1618, (2008) · Zbl 1255.62198

[18] Huang, L; Zhang, HH; Zeng, ZB; Bushel, PR, Improved sparse multi-class SVM and its application for gene selection in cancer classification, Cancer Inform, 12, 143-153, (2013)· doi:10.4137/CIN.S10212

[19] Hui, Z, The adaptive lasso and its oracle properties, Journal of the American Statistical Association, 101, 1418-1429, (2006) · Zbl 1171.62326 · doi:10.1198/016214506000000735

[20] Krause, N., & Singer, Y. (2004). Leveraging the margin more carefully. In \textit{Proceeding of ICML '04} (pp. 63-71). NY, USA. · Zbl 1089.62511

[21] Le Thi, H. A. (2005). DC programming and DCA. Available on http://lita.sciences.univ-metz.fr/~lethi/DCA.html. · Zbl 1116.90122

[22] Le Thi, H. A. (2012). \textit{A new approximation for the}$☒ _{0}$\textit{-norm}. Research Report LITA EA 3097, University of Lorraine. · Zbl 1119.62341

[23] Le Thi, H. A., & Phan, D. N. (2016). DC programming and DCA for sparse fisher linear discriminant analysis. \textit{Neural Computing and Applications}, doi:10.1007/s00521-016-2216-9. · Zbl 1284.90057

[24] Le Thi, HA; Belghiti, T; Pham Dinh, T, A new efficient algorithm based on DC programming and DCA for clustering, Journal of Global Optimization, 37, 593-608, (2006) · Zbl 1198.90327

[25] Le Thi, HA; Hoai, M; Dinh, T Pham, Feature selection in machine learning: an exact penalty approach using a difference of convex function algorithm, Machine Learning, 101, 163-186, (2015) · Zbl 1343.68201 · doi:10.1007/s10994-014-5455-y

[26] Le Thi, HA; Hoai, M; Nguyen, VV; Pham Dinh, T, A DC programming approach for feature selection in support vector machines learning, Journal of Advances in Data Analysis and Classification, 2, 259-278, (2008) · Zbl 1284.90057 · doi:10.1007/s11634-008-0030-7

[27] Le Thi, HA; Hoai, M; Pham Dinh, T, Optimization based DC programming and DCA for hierarchical clustering, European Journal of Operational Research, 183, 1067-1085, (2007) · Zbl 1149.90117 · doi:10.1016/j.ejor.2005.07.028

[28] Le Thi, HA; Huynh, VN; Pham Dinh, T, Exact penalty and error bounds in DC programming, Journal of Global Optimization, 52, 509-535, (2012) · Zbl 1242.49037 · doi:10.1007/s10898-011-9765-3

[29] Le Thi, HA; Nguyen, VV; Ouchani, S; Tang, C (ed.); Ling, CX (ed.); Zhou, X (ed.); Cercone, NJ (ed.); Li, X (ed.), Gene selection for cancer classification using DCA, No. 5139, 62-72, (2008), Heidelberg

[30] Le Thi, HA; Pham Dinh, T, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, Annals of Operations Research, 133, 23-46, (2005) · Zbl 1116.90122 · doi:10.1007/s10479-004-5022-1

[31] Le Thi, HA; Pham Dinh, T; Le Hoai, M; Vo, XT, DC approximation approaches for sparse optimization, European Journal of Operational Research, 244, 26-46, (2015) · Zbl 1346.90819 · doi:10.1016/j.ejor.2014.11.031

[32] Le Thi, HA; Phan, DN, DC programming and DCA for sparse optimal scoring problem, Neurocomputing, 186, 170-181, (2016)· doi:10.1016/j.neucom.2015.12.068

[33] Lee, Y; Kim, Y; Lee, S; Koo, J, Structured multicategory support vector machines with analysis of variance decomposition, Biometrika, 93, 555-71, (2006) · Zbl 1108.62059 · doi:10.1093/biomet/93.3.555

[34] Lee, Y; Lin, Y; Wahba, G, Multicategory support vector machines, theory, and application to the classification of microarray data and satellite radiance data, Journal of the American Statistical Association, 99, 67-81, (2004) · Zbl 1089.62511 · doi:10.1198/016214504000000098

[35] Li, G. Z., Yang, J., Liu, G. P., & Xue, L. (2004). Feature selection for multi-class problems using support vector machines. In \textit{PRICAI 2004: Trends in artificial intelligence, lecture notes in computer science 3157} (pp. 292-300). Berlin: Springer. · Zbl 1171.62326

[36] Liu, D; Qian, H; Dai, G; Zhang, Z, An iterative SVM approach to feature selection and classification in high-dimensional datasets, Pattern Recognition, 46, 2531-2537, (2013) · Zbl 1323.68459 · doi:10.1016/j.patcog.2013.02.007

[37] Liu, Y; Shen, X, Multicategory $Ψ $-learning, Journal of the American Statistical Association, 101, 500-509, (2006) · Zbl 1119.62341 · doi:10.1198/016214505000000781

[38] Liu, Y; Shen, X; Doss, H, Multicategory $ψ $-learning and support vector machine: computational tools, Journal of Computational and Graphical Statistics, 14, 219-236, (2005)· doi:10.1198/106186005X37238

[39] Liu, Y; Zhang, HH; Park, C; Ahn, J, Support vector machines with adaptive $☒ _q$ penalty, Computational Statistics & Data Analysis, 51, 6380-6394, (2007) · Zbl 1446.62179 · doi:10.1016/j.csda.2007.02.006

[40] Maldonado, S; Weber, R; Basak, J, Simultaneous feature selection and classification using kernel-penalized support vector machines, Information Sciences, 181, 115-128, (2011)· doi:10.1016/j.ins.2010.08.047

[41] Neumann, J; Schnörr, C; Steidl, G, Combined SVM-based feature selection and classification, Machine Learning, 61, 129-150, (2005) · Zbl 1137.90643 · doi:10.1007/s10994-005-1505-9

[42] Ong, CS; Le Thi, HA, Learning sparse classifiers with difference of convex functions algorithms, Optimization Methods and Software, 28, 4, (2013) · Zbl 1282.90181 · doi:10.1080/10556788.2011.652630

[43] Peleg, D; Meir, R, A bilinear formulation for vector sparsity optimization, Signal Processing, 8, 375-389, (2008) · Zbl 1186.94273 · doi:10.1016/j.sigpro.2007.08.015

[44] Pham Dinh, T., & Le Thi, H. A. (2014). Recent advances on DC programming and DCA. In \textit{Transactions on compu-

tational intelligence XIII, Lecture Notes in Computer Science Vol. 8342} (pp. 1-37).

[45] Pham Dinh, T; Le Thi, HA, Convex analysis approach to D.C. programming: theory, algorithm and applications, Acta Mathematica Vietnamica, 22, 289-355, (1997) · Zbl 0895.90152

[46] Pham Dinh, T; Le Thi, HA, Optimization algorithms for solving the trust region subproblem, SIAMJ. Optimization, 2, 476-505, (1998) · Zbl 0913.65054

[47] Rakotomamonjy, A, Variable selection using SVM-based criteria, Journal of Machine Learning Research, 3, 1357-1370, (2003) · Zbl 1102.68583

[48] Ramona, M; Richard, G; David, B, Multiclass feature selection with kernel Gram-matrix-based criteria, IEEE Transactions on Neural Networks and Learning Systems, 23, 1611-1623, (2012)· doi:10.1109/TNNLS.2012.2201748

[49] Ronan, C., Fabian, S., Jason, W., & Lé, B. (2006). Trading convexity for scalability. In \textit{Proceedings of the 23rd international conference on machine learning ICML 2006} (pp. 201-208). Pittsburgh, Pennsylvania. · Zbl 1323.68459

[50] Tibshirani, R, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, 46, 431-439, (1996) · Zbl 0850.62538

[51] Wang, H; Li, G; Jiang, G, Robust regression shrinkage and consistent variable selection via the LAD-LASSO, Journal of Business & Economics Statistics, 25, 347-355, (2007)· doi:10.1198/073500106000000251

[52] Wang, L; Shen, X, On $\boxtimes$_1$-norm multi-class support vector machine: methodology and theory, Journal of the American Statistical Association, 102, 583-594, (2003) · Zbl 1172.62317 · doi:10.1198/016214506000001383

[53] Weston, J., & Watkins, C. (1999). Support vector machines for multi-class pattern recognition. In \textit{Proceedings-European symposium on artificial neural networks, ESANN 1999} (pp. 219-224). D-Facto public. · Zbl 1135.62056

[54] Weston, J; Elisseeff, A; Schölkopf, B, Use of zero-norm with linear models and kernel methods, Journal of Machine Learning Research, 3, 1439-1461, (2003) · Zbl 1102.68605

[55] Wu, K., Lu, B., Uchiyama, M. & Isahara, H. (2007). A probabilistic approach to feature selection for multi-class text categorization. In D. Liu et al. (Eds.), \textit{ISNN 2007, Part I, LNCS 4491} (pp. 1310-1317).

[56] Yeh, Y., Chung, Y., Lin, T., & Wang, Y. (2011). Group lasso regularized multiple kernel learning for heterogeneous feature selection. In \textit{The 2011 international joint conference on neural networks (IJCNN)} (pp. 2570-2577). · Zbl 1171.62326

[57] Zhang, HH; Liu, Y; Wu, Y; Zhu, J, Variable selection for the multicategory SVM via adaptive sup-norm regularization, Journal of Statistics, 2, 149-167, (2008) · Zbl 1135.62056

[58] Zhou, Y., Jin, R. & Hoi, S. C. (2010). Exclusive lasso for multi-task feature selection. In \textit{AISTATS 9}.

[59] Zhou, X; Tuck, DP, MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data, Bioinformatics, 23, 1106-1114, (2007)· doi:10.1093/bioinformatics/btm036

[60] Zou, H, The adaptive lasso and its oracle properties, Journal of the American Statistical Association, 101, 1418-71429, (2006) · Zbl 1171.62326 · doi:10.1198/016214506000000735