

{tag} International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

[Volume 173](#)

-
[Number 4](#)

Year of Publication: 2017

Authors:

Manal Mostafa, Mohamed Gamal

10.5120/ijca2017915279

{bibtex}2017915279.bib{/bibtex}

Abstract

This paper presents High Probability Minimum Redundancy, HPMR, as a new algorithm for employing the most predictive features to contribute dimensionality reduction. The proposed algorithm is useful for finding new, optimal, and more informative features maintaining acceptable classification accuracy. A problem encountered in many large-scale information applications relevant to expert and intelligent systems such as pattern recognition, bioinformatics, social media content classification where data sets containing massive numbers of features. Implementing categorization on these huge, uneven, useless datasets with the overwhelming number of features is just a waste of time degrading the efficiency of classification algorithms and hence the results are not much accurate. HPMR controls the tradeoff between relevance and redundancy by selecting new feature subset that retains sufficient information to discriminate well among classes.

To emphasize the significance of HPMR, it has been relied upon to develop an intelligent system for Arabic sentiment analysis on social media. Additionally, the performance of such

algorithm is quantitatively compared with other traditional dimensionality reduction techniques in terms of performance accuracy, dataset reduction percentage, training time. Experimental results prove that HPMP cannot only diminish the feature vector but also can significantly enhance the performance of the well-known classifiers.

References

1. Abbasi; H. Chen; A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums" *ACM Trans. Information Systems*, Vol. 26, No. 3, Article 12, June 2008. .
2. Ahmed K. Farahat Ali Ghodsi Mohamed S. Kamel "An Efficient Greedy Method for Unsupervised Feature Selection" *IEEE 11th International Conference on Data Mining*. 2011.
3. Andrea Esuli; Fabrizio Sebastiani; Ahmed Abasi,"AI and Opinion Mining, Part 2", *IEEE Intelligent Systems*. Vol. 25, No.4, Pp. 72 - 79, 2010.
4. Anuj Sharma; Shubhamoy Dey "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis" *Special Issue of International Journal of Computer Applications on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA*, pp. 15- 20, June 2012.
5. Asha Gowda Karegowda, A. S. Manjunath & M.A.Jayaram "Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection" *International Journal of Information Technology and Knowledge Management*, Vol. 2, No. 2, pp. 271-277, July-Dec. 2010.
6. A. Alajmi; E. M. Saad; R. R. Darwish; "Toward an Arabic Stop-Words List Generation" *International Journal of Computer Applications*, Vol. 46, No.8, pp. 8 – 13, May 2012.
7. A.M. Martinez; A.C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 228-233, Feb. 2001.
8. Bangsuk Jantawan, Cheng-Fa Tsai "A Comparison of Filter and Wrapper Approaches with Data Mining Techniques for Categorical Variables Selection" *International Journal of Innovative Research in Computer and Communication Engineering*, pp. 4501- 4508, Vol. 2, No. 6, June 2014.
9. Daniel Jurafsky, James H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. 2nd Edition. Prentice-Hall, 2009.
10. Elhamifar, E.; Vidal, R. "Sparse Subspace Clustering: Algorithm, Theory and Applications" *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 35, No. 11, Pp. 2765 – 2781, 2013.
11. Erik Cambria; "Affective Computing and Sentiment Analysis" *IEEE Intelligent Systems*. Vol. 31, No. 2, pp. 102 -107, Mar - April 2016.
12. Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization" *ACM Computing Surveys*, Vol. 34, No.1, pp. 1–47, Mar. 2002.
13. Guangtao Wang; Qinbao Song; Baowen Xu; Yuming Zhou, "Selecting feature subset for high dimensional data via the propositional FOIL rules", *Pattern Recognition Elsevier*, Vol. 46, No. 1, pp. 199–214, Jan. 2013.
14. G.Vinodhini; RM.Chandrasekaran "Sentiment Analysis and Opinion Mining: A Survey" *International Journal of Advanced Research in Computer Science and Software Engineering*. Vol. 2, No. 6, pp 282 - 292. June 2012.

15. Huan Liu; Lei Yu “Toward integrating feature selection algorithms for classification and clustering” *IEEE Trans. Knowledge and Data Engineering*. Vol. 17, No. 4, Pp. 491 – 502, 2005.
16. Jean-Charles Lamirel; Pascal Cuxac; Kafil Hajlaoui; Aneesh Sreevallabh Chivukula, “A new feature selection and feature contrasting approach based on quality metric: application to efficient classification of complex textual data” , *PAKDD 2013 Int. Workshops on Trends and Applications in Knowledge Discovery and Data Mining*. pp. 367–378, 2013.
17. Jiguang Liang; Xiaofei Zhou; Li Guo¹; Shuo Bai; “Feature Selection for Sentiment Classification Using Matrix Factorization”, pp. 63 – 64. Florence, Italy, ACM. May 2015,
18. Joseph D. Prusa; Taghi M. Khoshgoftaar; David J. Dittman “Impact of Feature Selection Techniques for Tweet Sentiment Classification” *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*. pp. 299 – 304. 2015.
19. Jun Yan, Benyu Zhang, Shuicheng Yan , Qiang Yang, Hua Li; Zheng Chen; Wensi Xi; Weiguo Fan; Wei-Ying Ma; Qiansheng Cheng “IMMC: Incremental Maximum Margin Criterion” *Aug.*, pp. 22-25, 2004, ACM, Aug. 2004.
20. Jung-Yi Jiang; Ren-Jia Liou; Shie-Jue Lee, “A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification” *IEEE Trans. Knowledge and Data Engineering*, Vol. 23, No. 3, Pp. 335 - 349 Jan. 2011.
21. Junshi Xia; Lionel Bombrun; Tülay Adalı; Yannick Berthoumieu; Christian Germain; “Spectral Spatial Classification of Hyperspectral Images Using ICA and Edge- Preserving Filter via an Ensemble Strategy” *IEEE Trans. Geoscience and Remote Sensing*, Vol. 54, No. 8, pp. 4971 – 4982, 2016.
22. J. Yan, B. Zhang, N. Liu, S. Yan, Q. Cheng, W. Fan, Q. Yang, W. Xi, and Z. Chen, “Effective and Efficient Dimensionality Reduction for Large-Scale and Streaming Data Preprocessing,” *IEEE Trans. Knowledge and Data Eng.*, Vol. 18, no. 2, pp. 1-14, Feb. 2006.
23. K.Mugunthadevi, S.C. Punitha, M. Punithavalli “Survey on Feature Selection in Document Clustering *International Journal on Computer Science and Engineering (IJCSE)*” Vol. 3 No. 3, Mar 2011.
24. L.Ladha, T.Deepa “Feature Selection Methods and Algorithms” *International Journal on Computer Science and Engineering (IJCSE)* Vol. 3 No. 5, pp. 1787 – 1797, May 2011.
25. Manal Mustafa; A.Shakour AlSamahi; Alaa Hamouda, “New Avenues in Arabic Sentiment Analysis”, *International Journal of Scientific & Engineering Research*, Vol. 8, No. 2, pp. 907 - 915, Feb. 2017.
26. Michael L. Raymer, William F. Punch, Erik D. Goodman, Leslie A. Kuhn, and Anil K. Jai “Dimensionality Reduction Using Genetic Algorithms” pp. 164 - 171 *IEEE Trans. Evolutionary Computation*, Vol. 4, No. 2, July 2000.
27. M.A. Jawale; D.N. Kyatanavar; A.B. Pawar, “Automated Sentiment or Opinion Discovery System” *International Journal of Computer Applications (IJCA)*, Vol. 106, No. 13, pp. 29- 35, Nov. 2014.
28. Nicola Falco, Jon Atli Benediktsson, Lorenzo Bruzzone” *Spectral and Spatial Classification of Hyperspectral Images Based on ICA and Reduced Morphological Attribute Profiles*” *IEEE Trans. Geoscience and Remote Sensing*. Vol. 53, No. 11, pp. 6223 – 6240, 2015.
29. Richard Jensen; Qiang Shen; “Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches”, *IEEE Trans. Knowledge and Data Engineering*, Vol. 17, No. 1, pp. 1 – 15, Jan. 2005.
30. Satoshi Niijima , Yasushi Okuno “Laplacian Linear Discriminant Analysis Approach to

Unsupervised Feature Selection” IEEE/ACM Trans. Computational Biology and Bioinformatics, Vol. 6, No. 4, pp. 605-614 , Oct.-Dec. 2009.

31. Tehseen Zia, Qaiser Abbas, Muhammad Pervez Akhtar “ Evaluation of Feature Selection Approaches for Urdu Text Categorization” I.J. Intelligent Systems and Applications, pp. 33-40, June 2015.

32. Yan Dang, Yulei Zhang, Hsinchun Chen “A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews” IEEE Intelligent Systems, Vol. 25, No. 4, pp. 46- 53, July /Aug. 2010.

33. Yijun Sun, Sinisa Todorovic, Steve Goodison, “Local Learning Based Feature Selection for High Dimensional Data Analysis”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, pp. 1610 – 1626. July 2010.

Index Terms

Computer Science

Information Sciences

Keywords

Feature Selection, HPMR, Chi-squared, Social Media, Arabic, SA, SVM.