

REFERENCES

- Aladdin Knowledge Systems, Anti-spam white paper, <<http://www.eAladdin.com>>.
- Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., & Spyropoulos, C. D. (2000). An evaluation of naive bayesian anti-spam filtering. *arXiv preprint cs/0006013*.
- Androutsopoulos, I., Paliouras, G., & Michelakis, E. (2004). *Learning to filter unsolicited commercial e-mail*. "DEMOKRITOS", National Center for Scientific Research.
- Abdulsalam, H., Skillicorn, D. B., & Martin, P. (2011). Classification using streaming random forests. *Knowledge and Data Engineering, IEEE Transactions on*, 23(1), 22-36.
- Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007, October). A comparison of machine learning techniques for phishing detection. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 60-69). ACM.
- Ali, S., & Xiang, Y. (2007, July). Spam classification using adaptive boosting algorithm. In *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on* (pp. 972-976). IEEE.
- Assassin, S. The Apache SpamAssasin Project, 2 pages, downloaded from <http://spamassasin.apache.org> on Aug. 25, 2005. *USA cited by other*
- Awad, W. A., & ELseuofi S. M., (2011) Machine Learning Methods for Spam E-Mail Classification, *International Journal of Computer Science & Information Technology*, Vol 3, No 1, pp 173 –184. 2011.
- Azam, N. (2002). Comparative Study of Features Space Reduction Techniques for Spam Detection (*Doctoral dissertation, National University of Sciences and Technology*).
- Bartell, B. T., Cottrell, G. W., & Belew, R. K. (1994, August). Automatic combination of multiple ranked retrieval systems. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 173-181). Springer-Verlag New York, Inc..
- Bashiri, H., Oroumchian, F., & Moeini, A. (2005). Persian Email Classification Based on Rocchio and K-Nearest Neighbor Approach. *University of Wollongong in Dubai-Papers*, 19.
- Birrel, A, The Penny Black Project. <http://research.microsoft.com/research/sv/PennyBlack/>.
- Blanzieri, E., & Bryl, A. (2007). Evaluation of the highest probability SVM nearest neighbor classifier with variable relative error cost.

- Blanzieri, E., & Bryl, A. (2007, May). Instance-Based Spam Filtering Using SVM Nearest Neighbor Classifier. In *FLAIRS Conference* (pp. 441-442).
- Blanzieri, E., & Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, 29(1), 63-92.
- BogoFilter. <http://bogofilter.sourceforge.net/>.
- Boykin, P. O., & Roychowdhury, V. P. (2005). Leveraging social networks to fight spam. *Computer*, 38(4), 61-68.
- Bratko, A., Filipič, B., Cormack, G. V., Lynam, T. R., & Zupan, B. (2006). Spam filtering using statistical data compression models. *The Journal of Machine Learning Research*, 7, 2673-2698.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140
- Cai, J., Tang, Y., & Hu, R. (2008, July). Spam filter for short messages using winnow. In *Advanced Language Processing and Web Information Technology, 2008. ALPIT'08. International Conference*
- Cao, J., Chen, A., & Huang, X. (2005). Almost periodic attractor of delayed neural networks with variable coefficients. *Physics letters A*, 340(1), 104-120. (pp. 454-459). IEEE.
- Caliendo, M., Clement, M., Papiés, D., & Scheel-Kopeinig, S. (2008). The cost impact of spam filters: Measuring the effect of information system technologies in organisations.
- Carpineto, C., Romano, G., & Giannini, V. (2002). Improving retrieval feedback with multiple term-ranking function combination. *ACM Transactions on Information Systems (TOIS)*, 20(3), 259-290
- Carreras, X., & Marquez, L. (2001). Boosting trees for anti-spam email filtering. *arXiv preprint cs/0109015*.
- Carreras, X., & Márquez, L. (2001, July). Boosting trees for clause splitting. In *Proceedings of the 2001 workshop on Computational Natural Language Learning-Volume 7* (p. 26). Association for Computational Linguistics.
- Castillo, C., Chellapilla, K., & Denoyer, L. (2008). Web spam challenge 2008. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*.
- Castillo, C., Donato, D., Gionis, A., Murdock, V., & Silvestri, F. (2007, July). Know your neighbors: Web spam detection using the web topology. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 423-430). ACM.

- Cavallanti, G., Cesa-Bianchi, N., & Gentile, C. (2010). Linear algorithms for online multitask classification. *The Journal of Machine Learning Research*, 9999, 2901-2934.
- Chan, T. Y., Ji, J., & Zhao, Q. Learning to Detect Spam: Naive-Euclidean Approach. *International Journal of Signal Processing*, 1.
- Chang, M. W., Yih, W. T., & Meek, C. (2008, August). Partitioned logistic regression for spam filtering. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 97-105). ACM.
- Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759-771.
- Cheng, V., & Li, C. H. (2006, December). Personalized spam filtering with semi-supervised classifier ensemble. In *Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence* (pp. 195-201). IEEE Computer Society.
- Chellapilla, K., Larson, K., Simard, P., & Czerwinski, M. (2005, April). Designing human friendly human interaction proofs (HIPs). In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 711-720). ACM.
- Chellapilla, K., Larson, K., Simard, P. Y., & Czerwinski, M. (2005, July). Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs). In *CEAS*.
- Chuan, Z., Xianliang, L., Mengshu, H., & Xu, Z. (2005). A LVQ-based neural network anti-spam email approach. *ACM SIGOPS Operating Systems Review*, 39(1), 34-39.
- Clark, K. P. (2008). A Survey of Content-based Spam Classifiers.
- Coitweb.uncc.edu/~wwang22/Teaching/2013Fall.../IntroPhishing.ppt
- Cole, W. K. (2007). Blacklists, Blocklists, DNSBL's, and survival. *Retrieved on*, 01-26.
- Cormack, G. V. (2007). Email spam filtering: A systematic review. *Foundations and Trends in Information Retrieval*, 1(4), 335-455.
- Cormack, G. V., Gómez Hidalgo, J. M., & Sánz, E. P. (2007, November). Spam filtering for short messages. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 313-320). ACM
- Cormack, G. V., Hidalgo, J. M. G., & Sánz, E. P. (2007, July). Feature engineering for mobile (SMS) spam filtering. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 871-872). ACM
- Cormack, G. V., & Lynam, T. R. (2005, November). TREC 2005 Spam Track Overview. In *TREC*.

- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1), 21-27.
- Cunningham, P., & Delany, S. J. (2007). k-Nearest neighbour classifiers. *Multiple Classifier Systems*, 1-17.
- Dantu, R., & Kolan, P. (2005, July). Detecting spam in VoIP networks. In *Proceedings of the steps to reducing unwanted traffic on the internet on steps to reducing unwanted traffic on the internet workshop* (pp. 5-5). USENIX Association
- Dasarathy, B. V. (1991). Nearest Neighbor ({NN}) Norms: {NN} Pattern Classification Techniques.
- DeBarr, D., & Wechsler, H. (2009). Spam detection using clustering, random forests, and active learning. In *Sixth Conference on Email and Anti-Spam. Mountain View, California*.
- De Guerre, J. (2007). The mechanics of Vipul's Razor technology. *Network Security*, 2007(9), 15-17.
- Delany, S. J., Cunningham, P., & Tsymbal, A. (2006). A Comparison of Ensemble and Case-Base Maintenance Techniques for Handling Concept Drift in Spam Filtering. In *FLAIRS Conference* (pp. 340-345).
- Delany, S. J., Cunningham, P., Doyle, D., & Zamolotskikh, A. (2005). Generating estimates of classification confidence for a case-based spam filter. In *Case-Based Reasoning Research and Development* (pp. 177-190). Springer Berlin Heidelberg.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1-15). Springer Berlin Heidelberg.
- Dimmock, N., & Maddison, I. (2004). Peer-to-peer collaborative spam detection. *Crossroads*, 11(2), 4-4.
- Falkinger, J. 2007. Attention economies. *Journal of Economic Theory*. **133** 266-294.
- Drucker, H., Wu, D., & Vapnik, V. N. (1999). Support vector machines for spam categorization. *Neural Networks, IEEE Transactions on*, 10(5), 1048-1054.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.
- Erdélyi, M., Garzó, A., & Benczúr, A. A. (2011, March). Web spam classification: a few features worth more. In *Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality* (pp. 27-34). ACM.
- Espejo, P. G., Ventura, S., & Herrera, F. (2010). A survey on the application of genetic programming to classification. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(2), 121-144.

- Etzold, D. (2003). Improving spam filtering by combining Naive Bayes with simple k-nearest neighbor searches. *arXiv preprint cs/0312004*.
- Evers. Spam fighter hit with \$11.7 million judgement. Cnet News, September 14, 2006. http://news.cnet.com/2100-7350_3-6116009.html?part=rss&tag=6116009&subj=news.
- Exarchos, T. P., Tsipouras, M. G., Exarchos, C. P., Papaloukas, C., Fotiadis, D. I., & Michalis, L. K. (2007). A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. *Artificial Intelligence in Medicine*, 40(3), 187-200.
- Fox, E. A., & Shaw, J. A. (1994). Combination of multiple searches. *NIST SPECIAL PUBLICATION SP*, 243-243.
- Firte, L., Lemnaru, C., & Potolea, R. (2010, August). Spam detection filter using KNN algorithm and resampling. In *Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on* (pp. 27-33). IEEE.
- Freund, Y., & Schapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine learning*, 37(3), 277-296.
- Gadat, S., & Younes, L. (2007). A stochastic algorithm for feature selection in pattern recognition. *The Journal of Machine Learning Research*, 8, 509-547.
- Gallant, S. I. (1990). Perceptron-based learning algorithms. *Neural Networks, IEEE Transactions on*, 1(2), 179-191
- Geng, G. G., Wang, C. H., Li, Q. D., Xu, L., & Jin, X. B. (2007, August). Boosting the performance of web spam detection with ensemble under-sampling classification. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on* (Vol. 4, pp. 583-587). IEEE.
- Gkanogiannis, A., & Kalamboukis, T., 2009., A modified and fast Perceptron learning rule and its use for Tag Recommendations in Social Bookmarking Systems. *ECML PKDD Discovery Challenge 2009 (DC09)*
- Goh, K. L., Singh, A. K., & Lim, K. H. Multilayer Perceptrons : Neural Network Based on Web Spam Detection Application..
- Gómez Hidalgo, J. M., Bringas, G. C., Sáenz, E. P., & García, F. C. (2006, October). Content based SMS spam filtering. In *Proceedings of the 2006 ACM symposium on Document engineering* (pp. 107-114). ACM.
- Goodman, J., Cormack, G. V., & Heckerman, D. (2007). Spam and the ongoing battle for the inbox. *Communications of the ACM*, 50(2), 24-33.

- Graham-Cumming, J. (2005). Why I hate challenge-response. *JGC's Anti-Spam Newsletter*, 14.
- Graham, P. (2008). *Hackers & painters: big ideas from the computer age*. O'Reilly.
- Graham, P. (2003). A plan for spam, 2002, 2004. Available from World Wide Web: <http://www.paulgraham.com/spam.html>.
- Graham, P. Better bayesian filtering, 2003. URL <http://www.paulgraham.com/better.html>.
- Graham-Cumming, J. (2005, January). People and spam. In The Spam Conference (Vol. 15).
- Gulyás, C. (2006). Creation of a Bayesian network-based meta spam filter, using the analysis of different spam filters (*Doctoral dissertation, Master Thesis, Budapest, 16th May*).
- Gyongyi, Z., & Garcia-Molina, H. (2005). Web spam taxonomy. In *First international workshop on adversarial information retrieval on the web (AIRWeb 2005)*.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning* (Vol. 2, No. 1). New York: Springer.
- Hall, M. A., & Smith, L. A. (1998). Practical feature subset selection for machine learning.
- Harris, E. (2009). The next step in the spam control war: Greylisting, 2003. URL <http://projects.puremagic.com/greylisting/whitepaper.html>
- He, P., Sun, Y., Zheng, W., & Wen, X. (2008, January). Filtering short message spam of group sending using CAPTCHA. In *Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on* (pp. 558-561). IEEE
- Hershkop, S., & Stolfo, S. J. (2005, August). Combining email models for false positive reduction. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 98-107). ACM.
- Hidalgo, J. M. G. (2002, March). Evaluating cost-sensitive unsolicited bulk email categorization. In *Proceedings of the 2002 ACM symposium on Applied computing* (pp. 615-620). ACM.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press.
- Internet society.,(2012) Combating Spam: Policy, Technical and Industry Approaches.
- Islam, M. R., Chowdhury, M. U., & Zhou, W. (2005, November). An innovative spam filtering model based on support vector machine. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on* (Vol. 2, pp. 348-353). IEEE.

- Janecek, A., Gansterer, W. N., Demel, M., & Ecker, G. (2008). On the Relationship Between Feature Selection and Classification Accuracy. *Journal of Machine Learning Research-Proceedings Track*, 4, 90-105.
- Joachims, T. (1998). *Text categorization with support vector machines: Learning with many relevant features* (pp. 137-142). Springer Berlin Heidelberg
- Joseph, K. and A. Thevaranjan. 2008. Investigating pricing solutions to combat spam: Postage stamp and bonded senders. *Journal of Interactive Marketing*. Forthcoming.
- Kishore, J. K., Patnaik, L. M., Mani, V., & Agrawal, V. K. (2000). Application of genetic programming for multiclass pattern classification. *Evolutionary Computation, IEEE Transactions on*, 4(3), 242-258.
- Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3), 226-239.
- Kolcz, A., Chowdhury, A., & Alspector, J. (2004). The impact of feature selection on signature-driven spam detection. In *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS-2004)*.
- Koprinska, I., Poon, J., Clark, J., & Chan, J. (2007). Learning to classify e-mail. *Information Sciences*, 177(10), 2167-2187.
- Korada, N. K., Kumar, N. S. P., & Deekshitulu, Y. V. N. H. (2012). Implementation of Naive Bayesian Classifier and Ada-Boost Algorithm Using Maize Expert System. *International Journal of Information*, 2(3).
- Kolcz, A., & Chowdhury, A. (2007). Hardening fingerprinting by context. *CEAS'07*.
- Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection* (Vol. 1). MIT press.
- Kufandirimbwa, O., & Gatora, R. (2012). Spam Detection Using Artificial Neural Networks (Perceptron Learning Rule). *Online Journal of Physical and Environmental Science Research*, 1(2), 22-29.
- Kumar, R. K., Poonkuzhali, G., & Sudhakar, P. (2012). Comparative Study on Email Spam Classifier using Data Mining Techniques. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists* (Vol. 1).
- Lai, C. C. (2007). An empirical study of three machine learning methods for spam filtering. *Knowledge Based Systems*, 20(3), 249-254
- Lai, Chih-Chin, and Ming-Chi Tsai. "An empirical performance comparison of machine learning methods for spam e-mail categorization." *Hybrid Intelligent Systems, 2004. HIS'04. Fourth International Conference on*. IEEE, 2004.

- Labuschagne, P., du Toit J. V.(2012), Spam Classification using a Generalized Additive NeuralNetwork Approach., SATNAC, 2012.
- Lakshmi, R. D., & Radha, N. (2010, September). Spam classification using supervised learning techniques. In *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India* (p. 66). ACM.
- Lee, L. H., Wan, C. H., Yong, T. F., & Kok, H. M. (2010). A review of nearest neighbor-support vector machines hybrid classification models. *Journal of Applied Sciences*, 10, 1841-1858.
- Leiba, B., Ossher, J., Rajan, V. T., Segal, R., & Wegman, M. N. (2005, July). SMTP Path Analysis. In *CEAS*.
- Levine, J. R. (2005, July). Experiences with Greylisting. In *CEAS*.
- Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *Machine learning: ECML-98* (pp. 4-15). Springer Berlin Heidelberg.
- Li, & Ping (2009), The e-mail filtering system based on improved genetic algorithm. *Proceedings of the 2009 International Workshop on Information Security and Application (IWISA 2009)*, ISBN 978-952-5726-06-0
- Liu, B., McKay, B., & Abbass, H. A. (2003, December). Improving genetic classifiers with a boosting algorithm. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on* (Vol. 4, pp. 2596-2602). IEEE.
- Liu, H., & Setiono, R. (1995, November). Chi2: Feature selection and discretization of numeric attributes. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence* (pp. 388-388). IEEE Computer Society.
- Lowd, D., & Meek, C. (2005, July). Good Word Attacks on Statistical Spam Filters. In *CEAS*.
- Luo, X., & Zincir-Heywood, A. N. (2005). Evaluation of two systems on multi-class multi-label document classification. In *Foundations of Intelligent Systems*(pp. 161-169). Springer Berlin Heidelberg
- Lynam, T. R., Buckley, C., Clarke, C. L., & Cormack, G. V. (2004, November). A multi-system analysis of document and term selection for blind feedback. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (pp. 261-269). ACM
- MAAWG. 2007. *Email metrics program - the network operators' perspective, report #5 – first quarter 2007*. (accessed Jan. 12, 2008), [available at http://www.maawg.org/about/MAAWG20071Q_Metrics_Report.pdf].
- Manjusha, K., & Kumar, R. (2010, November). Spam Mail Classification Using Combined Approach of Bayesian and Neural Network. In *Computational Intelligence and*

Communication Networks (CICN), 2010 International Conference on (pp. 145-149). IEEE.

- McCue, R. (2009). A Comparison of the Accuracy of Support Vector Machine and Naive Bayes Algorithms In Spam Classification.
- Melville, N., A. Stevens, R.K. Plice, and O.V. Pavlov. 2006. Unsolicited commercial e-mail: Empirical analysis of a digital commons. *International Journal of Electronic Commerce*. **10**(4) 143-168.
- Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006, July). Spam filtering with naive bayes-which naive bayes?. In *CEAS* (pp. 27-28).
- Micro, T. (2005). Mail abuse prevention system. <http://www.mail-abuse.com/>.
- Minsky, M. L., & Papert, S. A. (1987). *Perceptrons - Expanded Edition: An Introduction to Computational Geometry*. Boston, MA.: MIT press.
- Mishne, G., Carmel, D., & Lempel, R. (2005, May). Blocking Blog Spam with Language Model Disagreement. In *AIRWeb* (Vol. 5, pp. 1-6)
- Montague, M., & Aslam, J. A. (2002, November). Condorcet fusion for improved retrieval. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 538-548). ACM.
- Neumayer, R. (2006). Clustering based ensemble classification for spam filtering. In *Proceedings of the 7th Workshop on Data Analysis (WDA'06)* (pp. 11-22).
- Ntoulas, A., Najork, M., Manasse, M., & Fetterly, D. (2006, May). Detecting spam web pages through content analysis. In *Proceedings of the 15th international conference on World Wide Web* (pp. 83-92). ACM.
- O'Brien, C., & Vogel, C. (2003, September). Spam filters: bayes vs. chi-Square; letters vs. words. In *Proceedings of the 1st international symposium on Information and communication technologies* (pp. 291-296). Trinity College Dublin.
- OECD. 2005. *Spam issues in developing countries*, <http://www.oecd.org/dataoecd/5/47/34935342.pdf>.
- Pantel, P., & Lin, D. (1998, March). Spamcop: A spam classification & organization program. In *Proceedings of AAAI-98 Workshop on Learning for Text Categorization* (pp. 95-98).
- Prakash, V. V., & O'Donnell, A. (2005). Fighting spam with reputation systems. *Queue*, **3**(9), 36-41
- Qiu, Y., Xu, Y., & Li, D. (2011). Application of Linear Classifier on Chinese Spam Filtering. *Journal of Software*, **6**(1), 116-123.
- Rahman, A., & Verma, B. (2011). Novel layered clustering-based approach for generating ensemble of classifiers. *Neural Networks, IEEE Transactions on*, **22**(5), 781-792.

- Rao, J. M., & Reiley, D. H. (2012). The economics of spam. *The Journal of Economic Perspectives*, 26(3), 87-110.
- Rennie, J. (2000, August). ifile: An application of machine learning to e-mail filtering. In *Proc. KDD 2000 Workshop on Text Mining, Boston, MA*.
- Rigoutsos, I., & Huynh, T. (2004, July). Chung-Kwei: a Pattern-discovery-based System for the Automatic Identification of Unsolicited E-mail Messages (SPAM). In *CEAS*.
- Rios, G., & Zha, H. (2004, July). Exploring Support Vector Machines and Random Forests for Spam Detection. In *CEAS*.
- Robinson, G. (2002). Spam detection. URL <http://radio.weblogs.com/0101454/stories/2002/09/16/spamDetection.html> accessed, 22, 00.
- Robinson, G. (2003). A statistical approach to the spam problem. *Linux journal*, 2003(107), 3.
- Robinson, G. (2004). Gary Robinson's spam rants, <http://radio.weblogs.com/0101454/categories/spam>.
- Saberi, A., Vahidi, M., & Bidgoli, B. M. (2007, November). Learn to detect phishing scams using learning and ensemble? methods. In *Web Intelligence and Intelligent Agent Technology Workshops, 2007 IEEE/WIC/ACM International Conferences on* (pp. 311-314). IEEE.
- Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In *Learning for Text Categorization: Papers from the 1998 workshop* (Vol. 62, pp. 98-105).
- Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., & Stamatopoulos, P. (2001). Stacking classifiers for anti-spam filtering of e-mail. *arXiv preprint cs/0106040*.
- Sasaki, M., & Shinnou, H. (2005, November). Spam detection using text clustering. In *Cyberworlds, 2005. International Conference on* (pp. 4-pp). IEEE.
- Schapire, R. E. (1997, July). Using output codes to boost multiclass learning problems. In *ICML* (Vol. 97, pp. 313-321).
- Schapire, R. E. (2003). The boosting approach to machine learning: An overview. *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG*
- Sculley, D. (2007, August). Online Active Learning Methods for Fast Label-Efficient Spam Filtering. In *CEAS*.
- Sculley, D., Wachman, G., & Brodley, C. E. (2006). Spam Filtering Using Inexact String Matching in Explicit Feature Space with On-Line Linear Classifiers. In *TREC*.

- Sculley, D., & Wachman, G. M. (2007, July). Relaxed online SVMs for spam filtering. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 415-422). ACM.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- Shengen, L., Xiaofei, N., Peiqi, L., & Lin, W. (2011, March). Generating new features using genetic programming to detect link spam. In *Intelligent Computation Technology and Automation (ICICTA), 2011 International Conference on* (Vol. 1, pp. 135-138). IEEE.
- Siefkes, C., Assis, F., Chhabra, S., & Yerazunis, W. S. (2004). Combining winnow and orthogonal sparse bigrams for incremental spam filtering. In *Knowledge Discovery in Databases: PKDD 2004* (pp. 410-421). Springer Berlin Heidelberg.
- Simard, P. Y., Szeliski, R., Benaloh, J., Couvreur, J., & Calinov, I. (2003, August). Using character recognition and segmentation to tell computer from humans. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on* (pp. 418-423). IEEE.
- Siponen, M., & Stucke, C. (2006, January). Effective anti-spam strategies in companies: An international study. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on* (Vol. 6, pp. 127c-127c). IEEE.
- Soonthornphisaj, N., Chaikulseriwat, K., & Tang-On, P. (2002, August). Anti-spam filtering: a centroid-based classification approach. In *Signal Processing, 2002 6th International Conference on* (Vol. 2, pp. 1096-1099). IEEE.
- Tang, Y., Krasser, S., He, Y., Yang, W., & Alperovitch, D. (2008, November). Support vector machines and random forests modeling for spam senders behavior analysis. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE* (pp. 1-5). IEEE.
- The Spamhaus Project - The Definition Of Spam". Spamhaus.org. Retrieved 2013-09-03.
- Tretyakov, K. (2004, May). Machine learning techniques in spam filtering. In *Data Mining Problem-oriented Seminar, MTAT* (Vol. 3, pp. 60-79).
- Trudgian, D. C. (2004). Spam classification using nearest neighbour techniques. In *Intelligent Data Engineering and Automated Learning-IDEAL 2004* (pp. 578-585). Springer Berlin Heidelberg.
- Vahora, S., Hasan, M., & Lakhani, R. (2011, December). Novel approach: Naïve Bayes with Vector space model for spam classification. In *Engineering (NUiCONE), 2011 Nirma University International Conference on* (pp. 1-5). IEEE.

- Webb, S., Caverlee, J., & Pu, C. (2006, July). Introducing the Webb Spam Corpus: Using Email Spam to Identify Web Spam Automatically. In *CEAS*
- Wei, C. P., Chen, H. C., & Cheng, T. H. (2008). Effective spam filtering: A single-class learning and ensemble approach. *Decision Support Systems*, 45(3), 491-503.
- Whittaker, S., Bellotti, V., & Moody, P. (2005). Introduction to this special issue on revisiting and reinventing e-mail. *Human-Computer Interaction*, 20(1), 1-9.
- Wiki E-mail spam, http://en.wikipedia.org/wiki/E-mail_spam
- Wiki Spam (electronic), http://en.wikipedia.org/wiki/Spam_%28electronic%29
- Woitaszek, M., Shaaban, M., & Czernikowski, R. (2003, January). Identifying junk electronic mail in microsoft outlook with a support vector machine. In *Applications and the Internet, 2003. Proceedings. 2003 Symposium on* (pp. 166-169). IEEE.
- Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2), 241-259.
- Xu, Z., Weinberger, K., & Chapelle, O. (2012). The greedy miser: Learning under test-time budgets. *arXiv preprint arXiv:1206.6451*.
- Yang, Z., Nie, X., Xu, W., & Guo, J. (2006, October). An approach to spam detection by naive Bayes ensemble based on decision induction. In *Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on* (Vol. 2, pp. 861-866). IEEE.
- Yerazunis, W., Siefkes, C., & Chhabra, S. Sorting Spam with K-nearest-neighbor and Hyperspace Classifiers.
- Yerazunis, B. (2002). Better Than Human. *Web document*, URL: <http://www.paulgraham.com/wsy.html>
- Yerazunis, W. S. (2004, January). The spam-filtering accuracy plateau at 99.9% accuracy and how to get past it. In *Proceedings of the 2004 MIT Spam Conference*.
- Ying, K. C., Lin, S. W., Lee, Z. J., & Lin, Y. T. (2010). An ensemble approach applied to classify spam e-mails. *Expert Systems with Applications*, 37(3), 2197-2201.
- Youn, S., & McLeod, D. (2007, April). Efficient spam email filtering using adaptive ontology. In *Information Technology, 2007. ITNG'07. Fourth International Conference on* (pp. 249-254). IEEE.
- Zdziarski, J. A. (2005). *Ending spam: Bayesian content filtering and the art of statistical language classification*. No Starch Press.
- Zhang, C. X., & Zhang, J. S. (2008). A local boosting algorithm for solving classification problems. *Computational Statistics & Data Analysis*, 52(4), 1928-1941.

- Zhang, L. 2005. The CAN-SPAM act: An insufficient response to the growing spam problem. *Berkeley Technology Law Journal*. **20** 301-332.
- Zhang, L., Zhu, J., & Yao, T. (2004). An evaluation of statistical spam filtering techniques. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(4), 243-269.
- Zhao, W., & Zhang, Z. (2005, May). An email classification model based on rough set theory. In *Active Media Technology, 2005.(AMT 2005). Proceedings of the 2005 International Conference on* (pp. 403-408). IEEE.
- Zhou, Y., Mulekar, M. S., & Nerellapalli, P. (2007). Adaptive spam filtering using dynamic feature spaces. *International Journal on Artificial Intelligence Tools*, 16(04), 627-646.
- Zinman, A., & Donath, J. S. (2007, August). Is britney spears spam?. In *CEAS*
- Zorkadis, V., Panayotou, M., & Karras, D. A. (2005). Improved spam e-mail filtering based on committee machines and information theoretic feature extraction. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on* (Vol. 1, pp. 179-184). IEEE.
- Zorkadis, V., & Karras, D. A. (2006). Efficient information theoretic extraction of higher order features for improving neural network-based spam e-mail categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, 18(4), 523-534.