

Voigt, Tobias; Fried, Roland; Backes, Michael; Rhode, Wolfgang

Threshold optimization for classification in imbalanced data in a problem of gamma-ray astronomy. (English) Zbl 1459.62231

Adv. Data Anal. Classif., ADAC 8, No. 2, 195-216 (2014).

Summary: We introduce a method to minimize the mean square error (MSE) of an estimator which is derived from a classification. The method chooses an optimal discrimination threshold in the outcome of a classification algorithm and deals with the problem of unequal and unknown misclassification costs and class imbalance. The approach is applied to data from the MAGIC experiment in astronomy for choosing an optimal threshold for signal-background-separation. In this application one is interested in estimating the number of signal events in a dataset with very unfavorable signal to background ratio. Minimizing the MSE of the estimation is a rather general approach which can be adapted to various other applications, in which one wants to derive an estimator from a classification. If the classification depends on other or additional parameters than the discrimination threshold, MSE minimization can be used to optimize these parameters as well. We illustrate this by optimizing the parameters of logistic regression, leading to relevant improvements of the current approach used in the MAGIC experiment.

MSC:

62P35 Applications of statistics to physics

62H30 Classification and discrimination; cluster analysis (statistical aspects)

85-08 Computational methods for problems pertaining to astronomy and astrophysics

Keywords:

classification; thresholding; magic; imbalanced data; unknown misclassification costs; random forest

Software:

CORSIKA

Full Text: [DOI](#)

References:

- [1] Aharonian FA (2004) Very high energy cosmic gamma radiation—a crucial window on the extreme universe. World Scientific Publishing Co.Pte. Ltd, Singapore· doi:10.1142/9789812561732
- [2] Albert J, Aliu E, Anderhub H, Antoranz P, Armada A, Asensio M, Baixeras C, Barrio JA, Bartko H, Bastieri D, Becker J, Bednarek W, Berger K, Bigongiari C, Biland A, Bock RK, Bordas P, Bosch-Ramon V, Bretz T, Britvitch I, Camara M, Carmona E, Chilingarian A, Ciprini S, Coarasa JA, Commichau S, Contreras JL, Cortina J, Costado MT, Curtef V, Danielyan V, Dazzi F, de Angelis A, Delgado C, de Lotto B, Domingo-Santamaría E, Dorner D, Doro M, Errando M, Fagiolini M, Ferenc D, Fernández E, Firpo R, Flix J, Fonseca MV, Font L, Fuchs M, Galante N, García-López RJ, Garczarczyk M, Gaug M, Giller M, Goebel F, Hakobyan D, Hayashida M, Hengstebeck T, Herrero A, Höhne D, Hose J, Hsu CC, Jacon P, Jogler T, Kosyra R, Kranich D, Kritzer R, Laille A, Lindfors E, Lombardi S, Longo F, López J, López M, Lorenz E, Majumdar P, Maneva G, Mannheim K, Mansutti O, Mariotti M, Martínez M, Mazin D, Merck C, Meucci M, Meyer M, Miranda JM, Mirzoyan R, Mizobuchi S, Moralejo A, Nieto D, Nilsson K, Ninkovic J, Oña-Wilhelmi E, Otte N, Oya I, Panniello M, Paoletti R, Paredes JM, Pasanen M, Pascoli D, Pauss F, Pegna R, Persic M, Peruzzo L, Piccioli A, Puchades N, Prandini E, Raymers A, Rhode W, Ribó M, Rico J, Rissi M, Robert A, Rügamer S, Saggion A, Saito T, Sánchez A, Sartori P, Scalzotto V, Scapin V, Schmitt R, Schweizer T, Shayduk M, Shinozaki K, Shore SN, Sidro N, Sillanpää A, Sobczynska D, Stamerra A, Stark LS, Takalo L, Temnikov P, Tescaro D, Teshima M, Torres DF, Turini N, Vankov H, Vitale V, Wagner RM, Wibig T, Wittek W, Zandanel F, Zanin R, Zapatero J (2007) Unfolding of differential energy spectra in the MAGIC experiment. *Nucl Instrum Methods Phys Res A* 583:494-506. doi:10.1016/j.nima.2007.09.048, 0707.2453
- [3] Albert J, Aliu E, Anderhub H, Antoranz P, Armada A, Asensio M, Baixeras C, Barrio JA, Bartko H, Bastieri D, Becker J, Bednarek W, Berger K, Bigongiari C, Biland A, Bock RK, Bordas P, Bosch-Ramon V, Bretz T, Britvitch I, Camara M, Carmona E, Chilingarian A, Ciprini S, Coarasa JA, Commichau S, Contreras JL, Cortina J, Costado MT, Curtef V, Danielyan V, Dazzi F, de Angelis A, Delgado C, de Lotto B, Domingo-Santamaría E, Dorner D, Doro M, Errando M, Fagiolini M, Ferenc D, Fernández E, Firpo R, Flix J, Fonseca MV, Font L, Fuchs M, Galante N, García-López RJ, Garczarczyk M, Gaug M, Giller M, Goebel F, Hakobyan D, Hayashida M, Hengstebeck T, Herrero A, Höhne D, Hose J, Huber S, Hsu CC, Jacon P, Jogler T, Kosyra R, Kranich D, Kritzer R, Laille A, Lindfors E, Lombardi S, Longo F, López J, López M, Lorenz E, Majumdar P, Maneva G, Mannheim K, Mariotti M, Martínez M, Mazin D, Merck C, Meucci M, Meyer M, Miranda JM, Mirzoyan R, Mizobuchi S, Moralejo A, Nieto D, Nilsson K, Ninkovic J, Oña-Wilhelmi E, Otte N, Oya I, Panniello M, Paoletti R, Paredes

- JM, Pasanen M, Pascoli D, Pauss F, Pegna R, Persic M, Peruzzo L, Piccioli A, Puchades N, Prandini E, Raymers A, Rhode W, Ribó M, Rico J, Rissi M, Robert A, Rügamer S, Saggion A, Saito TY, Sánchez A, Sartori P, Scalzotto V, Scapin V, Schmitt R, Schweizer T, Shayduk M, Shinozaki K, Shore SN, Sidro N, Sillanpää A, Sobczynska D, Spanier F, Stamerra A, Stark LS, Takalo L, Temnikov P, Tescaro D, Teshima M, Torres DF, Turini N, Vankov H, Venturini A, Vitale V, Wagner RM, Wibig T, Wittek W, Zandanel F, Zanin R, Zapatero J (2008) Implementation of the random forest method for the imaging atmospheric Cherenkov telescope MAGIC. *Nucl Instrum Methods Phys Res A* 588:424-432. doi:10.1016/j.nima.2007.11.068, 0709.3719
- [4] Aleksić J, Anderhub H, Antonelli LA, Antoranz P, Backes M, Baixeras C, Balestra S, Barrio JA, Bastieri D, Becerra González J, Becker JK, Bednarek W, Berdyugin A, Berger K, Bernardini E, Biland A, Bock RK, Bonnoli G, Bordas P, Borla Tridon D, Bosch-Ramon V, Bose D, Braun I, Bretz T, Britzger D, Camara M, Carmona E, Carosi A, Colin P, Commichau S, Contreras JL, Cortina J, Costado MT, Covino S, Dazzi F, de Angelis A, de Cea Del Pozo E, de Los Reyes R, de Lotto B, de Maria M, de Sabata F, Delgado Mendez C, Doert M, Domínguez A, Dominis Prester D, Dorner D, Doro M, Elsaesser D, Errando M, Ferenc D, Fonseca MV, Font L, García López RJ, Garczarczyk M, Gaug M, Godinovic N, Hadasch D, Herrero A, Hildebrand D, Höhne-Mönch D, Hise J, Hrupec D, Hsu CC, Jogler T, Klepser S, Krähenbühl T, Kranich D, La Barbera A, Laille A, Leonardo E, Lindfors E, Lombardi S, Longo F, López M, Lorenz E, Majumdar P, Maneva G, Mankuzhiyil N, Mannheim K, Maraschi L, Mariotti M, Martínez M, Mazin D, Meucci M, Miranda JM, Mirzoyan R, Miyamoto H, Moldón J, Moles M, Moralejo A, Nieto D, Nilsson K, Ninkovic J, Orito R, Oya I, Paoletti R, Paredes JM, Partini S, Pasanen M, Pascoli D, Pauss F, Pegna RG, Perez-Torres MA, Persic M, Peruzzo L, Prada F, Prandini E, Puchades N, Puljak I, Reichardt I, Rhode W, Ribó M, Rico J, Rissi M, Rügamer S, Saggion A, Saito TY, Salvati M, Sánchez-Conde M, Satalecka K, Scalzotto V, Scapin V, Schweizer T, Shayduk M, Shore SN, Sierpowska-Bartosik A, Sillanpää A, Sitarek J, Sobczynska D, Spanier F, Spiro S, Stamerra A, Steinke B, Strah N, Struebig JC, Suric T, Takalo L, Tavecchio F, Temnikov P, Tescaro D, Teshima M, Torres DF, Vankov H, Wagner RM, Zabalza V, Zandanel F, Zanin R, MAGIC Collaboration (2010) MAGIC TeV gamma-ray observations of Markarian 421 during multiwavelength campaigns in 2006. *Astron Astrophys* 519:A32+. doi:10.1051/0004-6361/200913945
- [5] Aleksić J, Alvarez EA, Antonelli LA, Antoranz P, Asensio M, Backes M, Barrio JA, Bastieri D, Bednarek W, Berdyugin A, Berger K, Bernardini E, Biland A, Blanch O, Bock RK, Boller A, Bonnoli G, Braun I, Bretz T, Cañellas A, Carmona E, Carosi A, Colin P, Colombo E, Contreras JL, Cortina J, Cossio L, Covino S, Dazzi F, de Angelis A, de Caneva G, de Cea Del Pozo E, de Lotto B, Delgado Mendez C, Diago Ortega A, Doert M, Domínguez A, Dominis Prester D, Dorner D, Doro M, Elsaesser D, Ferenc D, Fonseca MV, Font L, Fruck C, Garczarczyk M, Garrido D, Giavitto G, Godinovic N, Hadasch D, Häfner D, Herrero A, Hildebrand D, Höhne-Mönch D, Hise J, Hrupec D, Huber B, Jogler T, Kellermann H, Klepser S, Krähenbühl T, Krause J, La Barbera A, Lelas D, Leonardo E, Lindfors E, Lombardi S, López M, López-Oramas A, Lorenz E, Makariev M, Maneva G, Mankuzhiyil N, Mannheim K, Maraschi L, Mariotti M, Martínez M, Mazin D, Meucci M, Miranda JM, Mirzoyan R, Miyamoto H, Moldón J, Moralejo A, Munar-Adrover P, Nieto D, Nilsson K, Orito R, Oya I, Paneque D, Paoletti R, Pardo S, Paredes JM, Partini S, Pasanen M, Pauss F, Perez-Torres MA, Persic M, Peruzzo L, Pilia M, Pochon J, Prada F, Prandini E, Puljak I, Reichardt I, Reinthal R, Rhode W, Ribó M, Rico J, Rügamer S, Saggion A, Saito K, Saito TY, Salvati M, Satalecka K, Scalzotto V, Scapin V, Schultz C, Schweizer T, Shayduk M, Shore SN, Sillanpää A, Sitarek J, Snidarić I, Sobczynska D, Spanier F, Spiro S, Stamatescu V, Stamerra A, Steinke B, Storz J, Strah N, Surić T, Takalo L, Takami H, Tavecchio F, Temnikov P, Terzić T, Tescaro D, Teshima M, Tibolla O, Torres DF, Treves A, Uellenbeck M, Vankov H, Vogler P, Wagner RM, Weitzel Q, Zabalza V, Zandanel F, Zanin R (2012) Performance of the MAGIC stereo system obtained with Crab Nebula data. *Astropart Phys* 35:435-448. doi:10.1016/j.astropartphys.2011.11.007, 1108.1477
- [6] Aliu E et al (2009) Improving the performance of the single-dish Cherenkov telescope MAGIC through the use of signal timing. *Astropart Phys* 30:293-305. doi:10.1016/j.astropartphys.2008.10.003, 0810.3568
- [7] Becherini Y, Djannati-Ataï A, Marandon V, Punch M, Pita S (2011) A new analysis strategy for detection of faint γ -ray sources with imaging atmospheric Cherenkov telescopes. *Astropart Phys* 34:858-870. doi:10.1016/j.astropartphys.2011.03.005, 1104.5359
- [8] Bock RK, Chilingarian A, Gaug M, Hakl F, Hengstebeck T, Jiřina M, Klaschka J, Kotrč E, Savický P, Towers S, Vaiculis A, Wittek W (2004) Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope. *Nucl Instrum Methods Phys Res A* 516:511-528. doi:10.1016/j.nima.2003.08.157 · doi:10.1016/j.nima.2003.08.157
- [9] Boinee P, Barbarino F, de Angelis A, Saggion A, Zucchello M (2006) Neural networks for gamma-hadron separation in MAGIC. In: Sidharth BG, Honsell F, de Angeles A (eds) *Frontiers of fundamental and computational physics*, p 297. arXiv:astro-ph/0503539
- [10] Breiman L (2001) Random forests. *Mach Learn* 45:5 · Zbl 1007.68152 · doi:10.1023/A:1010933404324
- [11] Carmona E, Majumdar P, Moralejo A, Vitale V, Sobczynska D, Haffke M, Bigongiari C, Cabras G, de Maria M, de Sabata F, for the MAGIC collaboration (2008) Monte carlo simulation for the MAGIC-II system. In: *Proceedings of the 30th international cosmic ray conference, international cosmic ray conference, vol 3*, pp 1373-1376 (0709.2959)
- [12] Chadwick PM, Latham IJ, Nolan SJ (2008) TOPICAL REVIEW: TeV gamma-ray astronomy. *JPhys G Nucl Phys* 35(3):033201+. doi:10.1088/0954-3899/35/3/033201
- [13] Cherenkov PA (1934) Visible emission of clean liquids by action of gamma radiation. *Doklady Akademii Nauk SSSR* 2:451+. <http://ufn.ru/en/articles/2007/4/g/>
- [14] Domingo-Santamaria E, Flix J, Rico J, Scalzotto V, Wittek W (2005) The DISP analysis method for point-like or extended gamma source searches/studies with the MAGIC telescope. In: *Proceedings of the 29th international cosmic ray conference, international cosmic ray conference, vol 5*, pp 363-366
- [15] Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861-874. doi:10.1016/j.patrec.2005.10.010
- [16] Fegan DJ (1997) Topical review: gamma/hadron separation at TeV energies. *J Phys G Nucl Phys* 23:1013-1060. doi:10.1088/0954-3899/23/9/004 · doi:10.1088/0954-3899/23/9/004
- [17] Firpo Curcoll R, Delfino M, Neissner C, Reichardt I, Rico J, Tallada P, Tonello N (2011) The MAGIC data processing pipeline. *J Phys Conf Ser* 331(3):032,040. doi:10.1088/1742-6596/331/3/032040 · doi:10.1088/1742-6596/331/3/032040
- [18] Fomin VP, Stepanian AA, Lamb RC, Lewis DA, Punch M, Weekes TC (1994) New methods of atmospheric Cherenkov

- imaging for gamma-ray astronomy. I. The false source method. *Astropart Phys* 2:137-150. doi:10.1016/0927-6505(94)90036-1
doi:10.1016/0927-6505(94)90036-1
- [19] Hadasch D (2008) Study of the MAGIC performance at high zenith angles and application of the results on a very high energy gamma ray flare of the blazar PKS 2155-304. Diplomarbeit, Technische Universitaet Dortmund
- [20] Heck D, Knapp J (2010) EAS simulation with CORSIKA: a user's manual. Forschungszentrum Karlsruhe. <http://www-ik.fzk.de/corsika>
- [21] Hillas AM (1985) Cherenkov light images of EAS produced by primary gamma. In: Jones FC (ed) 19th international cosmic ray conference ICRC, San Diego, USA, International Cosmic Ray Conference, vol 3, p 445
- [22] Hinton J (2009) Ground-based gamma-ray astronomy with Cherenkov telescopes. *New J Phys* 11(5):055005-+. doi:10.1088/1367-2630/11/5/055005 (0803.1609)
- [23] Hinton JA, Hofmann W (2009) Teraelectronvolt astronomy. *Annu Rev Astron Astrophys* 47:523-565. doi:10.1146/annurev-astro-082708-101816, 1006.5210
- [24] Jogler T (2009) Detailed study of the binary system LS I +61o303 in VHE gamma-rays with the MAGIC telescope. Ph.D. thesis, Technische Universitaet Muenchen
- [25] Kohnle A, Aharonian F, Akhperjanian A, Bradbury S, Daum A, Deckers T, Fernandez J, Fonseca V, Hemberger M, Hermann G, Hess M, Heusler A, Hofmann W, Kankanian R, Köhler C, Konopelko A, Lorenz E, Mirzoyan R, Müller N, Panter M, Petry D, Plyasheshnikov A, Rauterberg G, Samorski M, Stamm W, Ulrich M, Völk HJ, Wiedner CA, Wirth H (1996) Stereoscopic imaging of air showers with the first two HEGRA Cherenkov telescopes. *Astropart Phys* 5:119-131. doi:10.1016/0927-6505(96)00011-4 · doi:10.1016/0927-6505(96)00011-4
- [26] Lessard RW, Buckley JH, Connaughton V, Le Bohec S (2001) A new analysis method for reconstructing the arrival direction of TeV gamma rays using a single imaging atmospheric Cherenkov telescope. *Astropart Phys* 15:1-18. doi:10.1016/S0927-6505(00)00133-X, arXiv:astro-ph/0005468
- [27] Li TP, Ma YQ (1983) Analysis methods for results in gamma-ray astronomy. *Astrophys J* 272:317-324. doi:10.1086/161295-
doi:10.1086/161295
- [28] Maier G, Knapp J (2007) Cosmic-ray events as background in imaging atmospheric Cherenkov telescopes. *Astropart Phys* 28:72-81. doi:10.1016/j.astropartphys.2007.04.009, 0704.3567
- [29] Majumdar P, Moralejo A, Bigongiari C, Blanch O, Sobczynska D, for the MAGIC collaboration (2005) Monte Carlo simulation for the MAGIC telescope. In: Proceedings of the 29th international cosmic ray conference, international cosmic ray conference, vol 5, p 203. arXiv:astro-ph/0508274
- [30] Mazin D (2007) A study of very high energy gamma-ray emission from AGNs and constraints on the extragalactic background light. Ph.D. thesis, Technische Universitaet Muenchen
- [31] Milke N, Rhode W, Ruhe T (2011) Studies on the unfolding of the atmospheric neutrino spectrum with IceCube 59 using the TRUUE algorithm. In: Proceedings of the 32nd international cosmic ray conference, international cosmic ray conference (1111.2736)
- [32] Milke N, Doert M, Klepser S, Mazin D, Blobel V, Rhode W (2012) Solving inverse problems with the unfolding program TRUue: examples in astroparticle physics
- [33] Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7(4):308-313. doi:10.1093/comjnl/7.4.308 · Zbl 0229.65053 · doi:10.1093/comjnl/7.4.308
- [34] Ohm S, van Eldik C, Egberts K (2009) γ /hadron separation in very-high-energy γ -ray astronomy using a multivariate analysis method. *Astropart Phys* 31:383-391. doi: 10.1016/j.astropartphys.2009.04.001 · doi:10.1016/j.astropartphys.2009.04.001
- [35] Quionero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND (2009) Dataset shift in machine learning. The MIT Press, Cambridge, MA, USA
- [36] Schlickeiser R (2002) Cosmic ray astrophysics. Springer, Berlin, Heidelberg · doi:10.1007/978-3-662-04814-6
- [37] Sheng V, Ling C (2006) Thresholding for making classifiers cost-sensitive. In: Proceedings of the 21st national conference on artificial intelligence, vol 1. AAAI Press, pp 476-481
- [38] Sobczynska D (2007) Natural limit on the γ /hadron separation for a stand alone air Cherenkov telescope. *J Phys G Nucl Phys* 34:2279-2288. doi:10.1088/0954-3889/34/11/005, arXiv:astro-ph/0702562
- [39] Voigt T (2010) Exploration und Vorverarbeitung von MAGIC-Daten zur Gamma-Hadron-Separation. Diplomarbeit, Technische Universitaet Dortmund, Germany
- [40] Weekes T (2003) Very high energy gamma-ray astronomy. Institute of Physics Publishing, Bristol, Philadelphia · doi:10.1887/0750306580

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.