Krakovsky, Yury; Luzgin, Aleksandr

Robust interval forecasting algorithm based on a probabilistic cluster model. (English)

Zbl 07192660
J. Stat. Comput. Simulation 88, No. 12, 2309-2324 (2018)

Summary: For substantiation of managerial decisions the forecasting results of dynamic indicators are used. Therefore, forecasting accuracy of these indicators must be acceptable. Consequently, forecasting algorithms are constantly improved to get the acceptable accuracy. This paper considers a variant of the method of forecasting binary outcomes. This method allows prediction of whether or not a future value of the indicator exceeds a predetermined value. This method 'interval forecasting' was named. In this paper a robust interval forecasting algorithm based on a probabilistic cluster model is proposed. The algorithm's accuracy was compared with an algorithm based on logistic regression. The indicators with different statistical properties were chosen. The obtained results have shown the accuracy of both the algorithms is approximately similar in most cases. However, the cases when the algorithm based on logistic regression demonstrated unacceptable accuracy, unlike the presented algorithm have been identified. Thus, this new algorithm is more accurate.

MSC:

62-XX Statistics

Keywords:

interval forecasting; forecasting binary outcomes; time series; dynamic indicators; probabilistic cluster model; logistic regression

Software:

extRemes; CRAN; astsa; SwissAir; mAr; R

Full Text: DOI

References:

[1] Pinson P, Nielsen HA, Moller JK, et al. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy. 2017;10(6):497-587. doi: 10.1002/we.230[Crossref], [Google Scholar]

[2] Elliot G, Lieli RP.Predicting binary outcomes. J Econom. 2013;174:15-26. doi: 10.1016/j.jeconom.2013.01.003[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1277.62043

[3] Lahiri K, Yang L.Forecasting binary outcomes. Handb Econ Forecast. 2013;2:1025-1106. doi: 10.1016/B978-0-444-62731-5.00019-1[Crossref], [Google Scholar]

[4] Fenlon C, Grandy L, Doherty L, et al. A discussion of calibration techniques for evaluating binary and categorical predictive models. Prev Vet Med.. 2018;149:107-114. doi: 10.1016/j.prevetmed.2017.11.018[Crossref], [PubMed], [Web of Science ®], [Google Scholar]

[5] Kubkowski M, Mielniczuk J.Projections of a general binary model on a logistic regression. Linear Algebra Appl. 2018;536:152-173. doi: 10.1016/j.laa.2017.09.013[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1403.62136

[6] Shumway RH, Stoffer DS. Time series analysis and its applications with R examples. 3rd ed. New York (NY): Springer; 2011. [Crossref], [Google Scholar] · Zbl 1276.62054

[7] Wang H, Li G, Wang G, et al. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy. 2017;88(15):56-70. doi: 10.1016/j.apenergy.2016.11.111[Crossref], [Google Scholar]

[8] Vernay M, Lafaysse M, Merindol L, et al. Ensemble forecasting of snowpack conditions and avalanche hazard. Cold Reg Sci Tech. 2015;120:251-262. doi: 10.1016/j.coldregions.2015.04.010[Crossref], [Web of Science ®], [Google Scholar]

[9] Krakovsky YM, Luzgin AN.Interval forecasting algorithm for dynamic indicators based on robust probabilistic cluster model. Sci Educ Bauman MSTU. 2016;11:113-126. [Google Scholar] · Zbl 07192660

[10] Cui M, Ke D, Sun Y, et al. Wind power ramp event forecasting using a stochastic scenario generation method. In: Bikash CP, editor. Proceedings of the IEEE transactions on sustainable energy. London: IEEE; 2015. p. 422-433. [Google Scholar]

[11] Yoder M, Hering AS, Navidi WC, et al. Short-termforecasting of categorical changes in wind power with Markov Chain models. Wind Energy. 2014;17:1425-1439. [Web of Science ®], [Google Scholar]

[12] Arbues F, Villanúa I.Determinants of behavior toward selective collection of batteries in Spain. A bivariate Probit model. Conserv Recycl. 2016;106:1-8. doi: 10.1016/j.resconrec.2015.11.004[Crossref], [Web of Science ®], [Google Scholar]

[13] Taylor JW.Probabilistic forecasting of wind power ramp events using autoregressive logit models. Eur J Oper Res. 2016;259:703-712. doi: 10.1016/j.ejor.2016.10.041[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1395.62350

[14] Murata A, Fujii Y, Naitoh K.Multinomial logistic regression model for predicting driver's drowsiness using behavioral measures. Procedia Manuf. 2015;3:2426-2433. doi: 10.1016/j.promfg.2015.07.502[Crossref], [Google Scholar]

[15] Kliestik T, Kocisova K, Misankova M.Logit and Probit model used for prediction of financial health of company. Procedia Econ Financ. 2015;23:850-855. doi: 10.1016/S2212-5671(15)00485-2[Crossref], [Google Scholar]

[16] Gutierrez-Perez GL, Mesones-Tello JP, Medina-Conde BB, et al. About conditions of application of logistic regression. Revista de Psiquiatría y Salud Mental (English Edition). 2017;10(4):1-216. [PubMed], [Web of Science ®], [Google Scholar]

[17] Tsangaratos P, Ilia I.Comparison of a logistic regression and naive Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. CATENA. 2016;145:164-179. doi: 10.1016/j.catena.2016.06.004[Crossref], [Web of Science ®], [Google Scholar]

[18] Villarubia G, Paz JF, Chamoso P, et al. Artificial neural networks used in optimization problems. Neurocomputing. 2018;272:10-16. doi: 10.1016/j.neucom.2017.04.075[Crossref], [Web of Science ®], [Google Scholar]

[19] Mansourian S, Darbandi EI, Mohassel MH, et al. Comparison of artificial neural networks and logistic regression as potential methods for predicting weed populations on dryland chickpea and winter wheat fields of Kurdistan province, Iran. Crop Protection. 2017;93:43-51. doi: 10.1016/j.cropro.2016.11.015[Crossref], [Web of Science ®], [Google Scholar]

[20] Munoz-Mas R, Fukuda S, Portoles J, et al. Revisiting probabilistic neural networks: a comparative study with support vectormachines and the microhabitat suitability for the Eastern Iberian chub (Squalius valentinus). Ecol Inform. 2018;43:24-37. doi: 10.1016/j.ecoinf.2017.10.008[Crossref], [Web of Science ®], [Google Scholar]

[21] Naik N, Purohit S.Comparative study of binary classification methods to analyze a massive dataset on virtual machine. Procedia Comput Sci. 2017;112:1863-1870. doi: 10.1016/j.procs.2017.08.232[Crossref], [Google Scholar]

[22] Panapakidis IP.Clustering based day-ahead and hour-ahead bus load forecasting models. Int J Electr Power Energy Syst. 2016;80:171-178. doi: 10.1016/j.ijepes.2016.01.035[Crossref], [Web of Science ®], [Google Scholar]

[23] Jimenez-Perez PF, Mora-Lopez L.Modeling and forecasting hourly global solar radiation using clustering and classification techniques. Orig Res Artic Sol Energy. 2016;135:682-691. doi: 10.1016/j.solener.2016.06.039[Crossref], [Web of Science ®], [Google Scholar]

[24] Jousselme A, Maupin P.Distances in evidence theory: comprehensive survey and generalizations. Int J Approx Reason. 2012;53(2):118-145. doi: 10.1016/j.ijar.2011.07.006[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1280.68258

[25] Todeschini R, Ballabio D, Consonni V, et al. A new concept of higher-order similarity and the role of distance/similarity measures in local classification methods. Chemometrics Intell Lab Syst. 2016;157:50-57. doi: 10.1016/j.chemolab.2016.06.013[Crossref], [Web of Science ®], [Google Scholar]

[26] Yang L, Dong L, Bi X.An improved location difference of multiple distances based nearest neighbors searching algorithm. Intern J Light Electron Optics. 2016;127(22):10838-10843. doi: 10.1016/j.ijleo.2016.08.091[Crossref], [Google Scholar]

[27] Shevlyakov G, Oja H. Robust correlation: theory and applications. New York: John Wiley and Sons; 2016. [Crossref], [Google Scholar] · Zbl 1381.62007

[28] Shevlyakov G, Smirnov P.Robust estimation of the correlation coefficient: An attempt of survey. Aust J Stat. 2011;40:147-156. [Google Scholar]

[29] Smirnov PO, Shevlyakov GL.Fast highly efficient and robust one-step M-estimators of scale based on Qn. Comput Stat Data Anal. 2014;78:153-158. doi: 10.1016/j.csda.2014.04.013[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1506.62169

[30] Smirnov PO, Shirokov IS, Shevlyakov GL.High-efficiency and robust M-estimates of the scale parameter on the Q-estimate basis. St Petersburg Polytech University J: Phys Math. 2017;3(3):239-241. doi: 10.1016/j.spjpm.2017.09.012[Crossref], [Google Scholar]

[31] Burger M, Repisky J.Problems of linear least square regression. In: Mokrys M, Lieskovsky A, editors. Proceedings of the 1th advanced research in scientific areas virtual conference. Slovakia: EDIS; 2012. p. 257-262. [Google Scholar]

[32] Ohlson JA, Kim S.Linear valuation without OLS: the Theil-Sen estimation approach. Rev Acc Stud. 2015;20(1):395-435. doi: 10.1007/s11142-014-9300-0[Crossref], [Web of Science ®], [Google Scholar]

[33] Wilcox RR, Clark F.Comparing robust regression lines associated with two dependent groups when there is heteroscedasticity. Comput Stat. 2014;29(5):1175-1186. doi: 10.1007/s00180-014-0485-2[Crossref], [Web of Science ®], [Google Scholar] · Zbl 1306.65150

[34] Shah SH, Rehman A, Rashid T, et al. A comparative study of ordinary least squares regression and Theil-Sen regression through simulation in the presence of outliers. U J Sci Techl. 2016;5:137-142. [Google Scholar]

[35] Barbosa SM.2012. R functions for multivariate autoregressive analysis. R package version 1.1;[cited 2017 Jun 15]. Available from: https://cran.r-project.org/web/packages/mAr[Google Scholar]

[36] Gilleland E.2016. Functions for performing extreme value analysis. R package version 2.0;[cited 2017 Feb 3]. Available from: https://cran.r-project.org/web/packages/extRemes[Google Scholar]

[37] Datamarket.com [Internet]. Reykjavik (IC): Basis ehf; [cited 2016 Dec 9]. Available from: https://datamarket.com/data/[Google Scholar]

[38] Locher R.2012. SwissAir: Air Quality Data of Switzerland for one year in 30 min Resolution. R package version 1.1.4;[cited 2016 Dec 25]. Available from: https://cran.r-project.org/web/packages/SwissAir[Google Scholar]

[39] Merkle EC, Steyvers M, Mellers B, et al. A neglected dimension of good forecasting judgment: the questions we choose also matter. Int J Forecast. 2017;33(4):817-832. doi: 10.1016/j.ijforecast.2017.04.002[Crossref], [Web of Science ®], [Google Scholar]

[40] Cran.r-project.org [Internet]. Vein (AT): Alfred Nagl; [cited 2017 Sep 5]. Available from: https://cran.r-project.org/[Google Scholar]