学校编码：10384

学号：19020101152507

硕 士 学 位 论 文

基于模糊熵和模糊聚类的模糊时间序列模型的研究

Research of fuzzy time series model based on fuzzy entropy and fuzzy clustering

朱琼

指导教师：张志强

专业名称：概率论与数理统计

答辩日期：2013年6月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于:

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

(　　)2.不保密，适用上述授权。

(请在以上相应括号内打“✓”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名):

年　　月　　日

摘 要

时间序列预测是通过对有限个历史观测样本进行分析来建立模型，并利用模型来解释数据之间的统计规律，以期达到控制和预报目的的一门学科，在众多领域中都有非常广泛的应用。对于时间序列的建模和预测，目前已经有了许多成熟的技术和方法，但传统时间序列预测方法往往依赖大量的历史数据，而在实际问题中由于不确定性的广泛存在导致历史数据往往是不完整的、不准确的和含糊的，因而限制了传统预测模型的应用。为了解决这些问题，Song和Chissom提出了模糊时间序列的概念，其主要是在传统时间序列预测的基础上引入了模糊理论，通过建立相应的模糊逻辑关系进行预测。由于模糊时间序列在处理数据的不确定性和模糊性方面上所显示的优势，关于它的研究也得到了越来越多的关注。经过多年的研究和论证发现模糊时间序列模型中论域的划分、语言值个数选取、数据模糊化、模糊逻辑关系的构建、预测值的修正是提高模型预测精度的几个重要部分。目前已有众多学者针对上述某些方面进行改进与创新，提出了各种各样的模糊预测方法，但这些方法在某些方面依然存在不足，本文在前人工作的基础上针对以上几个方面提出新算法对模糊时间序列中存在的问题进行了研究。

在论域划分和语言值个数选取上，目前的研究表明合理地划分论域十分重要，间隔长度的确定将极大地影响预测结果，有效的论域划分有助于预测精度的提高。而间隔长度的确定也取决于区间划分的个数。在论域划分上，有些学者提出使用聚类算法，其中最具有典型、最受欢迎的是模糊C均值聚类算法，但模糊C均值聚类算法聚类数目具有人为主观性确定的缺点，基于此，本文提出引入模糊熵的概念来确定最优聚类数目，从而对论域进行有效地划分。

在数据模糊化上，目前的研究普遍采用主观定义模糊集的方法，该方法简单，经研究发现在建立基于模糊逻辑关系分组时变模型时，不同的定义方法对其没有影响，但在建立基于模糊逻辑关系方程时变模型时，结果出现数据钝化，不能真实反映数据的分布。基于此，本文在数据模糊化时提出了一种建立模糊集定义的方法---基于模糊等价关系聚类的隶属函数的构造，该方法不仅避免了主观定义模糊集合的方法，还与前面划分区间个数紧紧相关，突出了样本的内部结构。

在模糊逻辑关系的构建上，目前的研究表明逻辑关系的阶次极大的影响预测精度且并不是阶次越高，预测结果越好。已有的研究中，大部分学者通常采用一阶模型，这是因为一阶模型计算方便，为了给出逻辑关系阶次确定的合理解释，本文采用传统时间序列自相关函数的概念来确定模型的阶数。

为了提高预测结果的精度，目前众多学者将先进算法引入到模糊时间序列中，同时考虑到混合算法可以显著的提高整体模型的预测精度，因此，本文在最终预测结果的基础上提出使用残差GM(1，1)模型对预测值进行修正，实验结果表明预测精度得到提高。

最后，针对大样本和小样本数据分别将本文的方法用于阿拉巴马州州立大学入学人数和台湾1998/01-2001/12机械行业产品价值的预测，对本文方法的结果与现有模型的结果进行对比，发现本文模型的预测精度较高。

关键词：模糊时间序列；模糊熵；模糊聚类；残差GM(1,1)；预测

Abstract

Time series forecasting is modeled by limited historical observations sample, it is a technology of using the model to explain the statistical regularity of data in order to achieve the purpose of control and forecast and having a wide range of applications in many fields. For time series modeling and forecasting, there are many mature technologies and methods. The traditional time series prediction model is dependent on a large number of historical data, but the historical data is often incomplete, inaccurate and vague due to the widespread presence of uncertainty in practical problems, and thus limit the application of traditional prediction model. In order to solve these problems, Song and Chissom proposed the concept of fuzzy time series which is mainly the introduction of traditional time series prediction based on fuzzy theory, then establish fuzzy logic relationship to predict. For the advantage of fuzzy time series to handle data of uncertainty and ambiguity, it has been gotten more and more attention. After years of research and demonstration, we found that the division of domain, number of linguistic values selected, data fuzzification, building of fuzzy logic relationship, amendments of predictive value are several important parts to improve prediction accuracy in fuzzy time series model. Many scholars have been made improvements and innovation for some of these at present, they proposed variety of fuzzy forecasting methods, but these methods are still inadequate for the above aspects. This paper is on the basis of previous work and propose a new algorithm for fuzzy time series.

On the division of the domain and the number of linguistic values selected, the current study shows that reasonably divided the universe is very important, the determination of the interval length will greatly affect the predicted results, the domain of division make a contribution to the improvement of the prediction accuracy. The determination of the length of the interval also depends on the

number of interval division. On the divided of domain, some scholars have proposed to use clustering algorithms, one of the most typical and popular is fuzzy C-means clustering algorithm, but it has the shortcomings of human subjectivity determine on the number of clusters, based on this, this paper proposed to introduce the concept of fuzzy entropy to determine the optimal number of clusters which effectively divided the domain .

On the data fuzzification, the current study is commonly used subjective definition of fuzzy set method which is simple. By study we found that different definitions of sets have no effect in the time-variant model based on fuzzy logic relationship group, but in fuzzy logic relationship equation model, the result of data is passivated which cannot truly reflect the distribution of the data. Thus this paper puts forward an universal method of definition of fuzzy sets based on this data fuzzification---the structure of the membership function based on fuzzy equivalence relation clustering, the method is not only avoid the subjective definition of fuzzy sets, but also related to the number of intervals in the previous division and highlights the internal structure of the sample.

On the establishment of fuzzy logic relationships ,current research show that the order of the logical relationships greatly affect prediction accuracy and not the higher order, the better prediction results. most commonly used first-order model in existing studies, for the calculations of first-order model is convenient. In order to give a reasonable explanation to the order determined of logical relationships, this paper use the concept of correlation function of traditional time to determine the order of the model in fuzzy time series.

In order to improve the accuracy of the predicted results, many scholars introduced some advanced algorithm into fuzzy time series and took into account that hybrid algorithm can significantly improve the prediction accuracy of the overall model, therefore, this paper uses the residual GM (1,1) model to amend the prediction value on the basis of fuzzy time series forecasting results, the

experimental results show that the prediction accuracy is improved.

Finally, for large sample and small sample data, our proposed method is used for Alabama State University enrollment and Taiwan machinery industry product value of 1998/01-2001/12 forecast, the results of our method and the results of existing models are compared and find that the proposed model with higher prediction accuracy.

Keywords: Fuzzy time series; Fuzzy entropy; Fuzzy clustering; Residual GM(1,1); Forecasting

参考资料

[1] L. A. Zadeh, Fuzzy sets, Information and Control [M], 8 (1965)338-353.
[2] Q. Song, B.S. Chissom , Fuzzy time series and its models [J], Fuzzy Sets and Systems 54 (1993a)269-277.
[3] Q. Song, B.S. Chissom , Forecasting enrollments with fuzzy time series—Part I [J], Fuzzy Sets and Systems 54(1993b)1-10.
[4] Q. Song, B.S. Chissom , Forecasting enrollments with fuzzy time series—Part II [J], Fuzzy Sets and Systems 62 (1994)1-8.
[5] S. M. Chen, Forecasting enrollments based on fuzzy time series [J], Fuzzy Sets and Systems 81 (1996)311-319.
[6] F. M. Tseng, G. H. Tzeng , H. C. Yu, B. J. C. Yuan, Fuzzy ARIMA model for forecasting the foreign exchange market [J], Fuzzy Sets and Systems 118(2001)9-19.
[7] F. M. Tseng, G. H. Tzeng , A fuzzy seasonal ARIMA model for forecasting [J], Fuzzy Sets and Systems 126(2002)367-376.
[8] 吴今培,模糊时间序列建模及应用 [J],系统工程(2002) 第20卷第4期.
[9] 常娜娜,石永奎,苏莉,模糊时间序列预测在动态经济系统预测中的应用 [J],山东大学自然科学学报(2007) 第26卷第5期.
[10] J. R. Hwang, S. M. Chen, C. H. Lee, Handing forecasting problems using fuzzy time series [J], Fuzzy Sets and Systems 100(1998)217-228.
[11] S. Melike, Y. D. Konstantin, Forecasting enrollment model based on first-order fuzzy time series, in proc. International Conference on Computational Intelligence, Istanbul, Turkey,2004.
[12] J. R. Wang, J. W. Liu, Weighted fuzzy time series forecasting model [C], Intelligent Information and Database Systems - Second International Conference, Part I, LNAI, Hue City, Vietnam,2010,5990:408-415.
[13] M. Stevenson, J. E. Porter, Fuzzy time series forecasting using percentage change as the universe of discourse [J], In: World Academy of Science ,Engineering and Technology 55(2009)154-156.
[14] K. Huarng , Effective lengths of intervals to improve forecasting in fuzzy time series [J], Fuzzy Sets and Systems 123(2001)387-394.
[15] K. H. Huarng , T. H. K. Yu, Ratio-based lengths of intervals to improve fuzzy time series forecasting [J], IEEE Transactions on systems, Man and Cybernetics-Part B:Cybernetics 36(2006)328-340.
[16] T. A. Jilani, S. M. A. Burney, C. Ardil ,Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning, In: Proceedings of World Academy of Science , Engineering and Technology 23(2007)1307-6884.
[17] R. C. Tsaur, T. C. Kuo, The adaptive fuzzy time series model with an application to Taiwan ’s tourism demand [J], Expert Systems with Applications 38 (2011)9164-9171.
[18] S. M. Chen, C. C. Hsu, A new method to forecasting enrollments using fuzzy time series [J],International Journal of Applied Science and Engineering 3(2004)234-244.
[19] C. H. Cheng, J. R. Chang, C. A. Yeh,Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost [J],Technological Forecasting and Social Change73(2006) 524-542.
[20] K. Huarng, H. K. Yu ,A dynamic approach to adjusting lengths of intervals in fuzzy time series forecasting [J],Intelligent Data Analysis 8(2004)3-27.
[21] U. Yolcu, E. Egrioglu, V. R. Uslu, M. A. Basaran, C. H. Aladag, A new approach for determining the length of intervals for fuzzy time series [J],Applied Soft Computing 9(2009)647-651.
[22] E. Egrioglu, C. H. Aladag, U. Yolcu, V. R. Uslu, M. A. Basaran, Finding an optimal interval length in high order fuzzy time series [J], Expert Systems with Applications 37 (2010)5052-5055.
[23] S.M. Chen, N. Y. Chung, Forecasting enrollments using high-order fuzzy time series and genetic algorithms [J], International journal of Intelligent Systems 21(2006)485-501.

[24] L. W. Lee, L. H. Wang, S. M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms [J], Expert Systems with Applications 33(2007)539-550.
[25] S. Davari, M. H. F. Zarandi, An improved fuzzy time series forecasting model based on particle swarm intervalization [J],In: The 28th North American Fuzzy Information Processing Society Annual Conference,2009.
[26] 付芳萍..基于信息熵及粒子群优化算法的模糊时间序列预测模型研究[D].硕士论文,昆明：昆明理工大学,2011年.
[27] 常玉杰.粒子群优化算法在模糊时间序列预测问题中的应用 [D], 硕士论文,大连：大连海事大学,2012年.
[28] I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin ,et al ,An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization [J], Expert Systems with Applications 36(2009)6108-6117.
[29] I. H. Kuo, S. J. Horng, Y. H. Chen, R. S. Run ,et al ,Forecasting TAIFEX based on fuzzy time series and particle swarm optimization [J], Expert Systems with Applications 37(2010)1494-1502.
[30] Y. L. Huang, S. J. Horng, M. X. He, etal, A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization [J], Expert Systems with Applications 38(2011)8014-8023.
[31] S. M. Chen, N. Y. Wang ,J. S. Pan, Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships [J], Expert Systems with Applications 36 (2009)11070-11076.
[32] S. M. Chen, K. Tanuwijaya, Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques [J], Expert Systems with Applications 38 (2011)15425-15437.
[33] N. Y. Wang, S. M. Chen, Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series [J], Expert Systems with Applications 36(2009)2143-2154.
[34] C. H. Cheng, G. W. Cheng, J. W. Wang, Multi-attribute fuzzy time series method based on fuzzy clustering [J], Expert Systems with Applications 34 (2008)1235-1242.
[35] K. Chi, F. P. Fu, W. G. Che, A novel model of fuzzy time series based on K-means clustering.[J] Proceeding of 2nd International Workshop on Education Technology and Computer Science ,2010,1:223-225.
[36] 曲宏巍.模糊时间序列模型相关理论的研究[D].硕士论文,大连：大连海事大学,2012年.
[37] H. T. Liu, An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers [J],Fuzzy Optim Decis Making(2007)6：63-80.
[38] K. Huarng , Heuristic models of fuzzy time series for forecasting [J] ,Fuzzy Sets and Systems 123 (2001)369-386.
[39] C. M. Own, P. T. Yu, Forecasting fuzzy time series on a heuristic high-order model [J], Cybernetics and Systems 36(2005)705-717.
[40] C. H. Cheng, T. L. Chen, H. J. Teoh, C. H. Chiang, Fuzzy time-series based on adaptive expectation model for TAIEX forecasting [J], Expert Systems with Applications34(2008) 1126-1132.
[41] E. J. Bai, W. K. Wong, W. C. Chu, ct al. A heuristic time-invariant model for fuzzy time series forecasting [J], Expert Systems with Applications 38 (2011)2701-2707.
[42] R. C. Tsaur, J. C. O. Yang, H. F. Wang, Fuzzy relation analysis in fuzzy time series model [J], Computers and Mathematics with Applications 49(2005)539-548.
[43] H. K. Yu, A refined fuzzy time series model for forecasting [J], Physica A 346(2005)657-681.
[44] H. K. Yu, Weighted fuzzy time series models for TAIEX forecasting [J], Physica A 349(2005)609-624.
[45] C. H. Cheng, T. L. Chen, C. H. Chiang, Trend-weighted fuzzy time series model for TAIEX forecasting [C],King et al.:ICONIP 2006,Part III,LNCS 4234,PP.469-477,2006.
[46]T. A. Jilani, S. M. A. Burney, M-factor high order fuzzy time series forecasting for road accident data [C], P. Melin et al. (Eds.): Anal. and Des. of Intel. Sys. using SC Tech., ASC 41, pp. 246 – 254, 2007.
[47]沈斌、姚敏、易文晟.基于二乘支持向量机的模糊时序分析方法[D],浙江大学学报(工学版)39(8)：1142-1146.
[48] S. M. Chen, Forecasting enrollments based on high-order fuzzy time series [J], Cybernetics and Systems 33 (2002)1-16.

[49] 迟凯.基于差分启发信息的模糊时间序列预测模型研究 [D],硕士学位论文,昆明：昆明理工大学,2010年.
[50] S. R. Singh, A robust method of forecasting based on fuzzy time series [J], Applied Mathematics and Computation 188(2007)472-484.
[51] S. R. Singh, A simple method of forecasting based on fuzzy time series [J], Applied Mathematics and Computation 186(2007)330-339.
[52] S. R. Singh, A computational method of forecasting based on high-order fuzzy time series [J], Expert Systems with Applications 36(2009)10551-10559.
[53] L. W. Lee, L. H. Wang, S. M. Chen, Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques [J], Expert Systems with Applications 34,1(2008)328-336.
[54] C. H. Aladag, M. A. Basaran, E. Egrioglu, et al, Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations [J], Expert Systems with Applications 36(2009)4228-4231.
[55] E. Egrioglu, C. H. Aladag, U. Yolcu, et al, A new approach based on artificial neural networks for high order multivariate fuzzy time series [J], Expert Systems with Applications 36(2009)10589-10594.
[56] S. M. Chen, C. D. Chen, Handing forecasting problems based on high-order fuzzy logical relationships [J], Expert Systems with Applications 38(2011)3857-3864.
[57] Q. Song, A note on fuzzy time series model selection with sample autocorrelation functions [J], Cybernetics and Systems 34(2003)93-107.
[58] S. T. Li, Y. C. Cheng, Deterministic fuzzy time series model for forecasting enrollments [J], Computers & Mathematics with Applications 53(2007)1904-1920.
[60] T. H. K. Yu, K. H. Huarng, A bivariate fuzzy time series model to forecast the TAIEX [J], Expert Systems with Applications 34(2008)2945-2952.
[59] S. M. Chen, J. R. Hwang, Temperature prediction using fuzzy time series, IEEE Transactions on systems ,Man and Cybernetics-Part B:Cybernetics 30(2000)263-275.
[61] 曾淑惠.多变量模糊时间序列之分析及应用 [D],博士论文,厦门：厦门大学,2003.
[62] 邱望仁.刘晓东.基于证据理论的模糊时间序列模型 [J],控制与决策 27(2012)99-103.
[63] 蔺玉佩.杨一文.基于模糊时间序列模型的股票市场预测 [J],统计与决策 8(2010)34-37.
[64] 石慧.王玉兰.翁福利.BP神经网络和模糊时间序列组合预测模型及其应用 [J],计算机应用,第31卷增刊2,2011.
[65] 邱望仁. 模糊时间序列模型及其在股指趋势分析中的应用研究[D],博士论文,大连：大连理工大学,2012年.
[66] K. H. Huarng, T. H. Yu, Y. W. Hsu, A multivariate heuristic model for fuzzy time series forecasting [J],IEEE Transaction on Systems, Man and Cybernetics-Part B:Cybernetics 37(4),836-846,2007.
[67] Q. Song, B.S. Chissom, Seasonal forecasting in fuzzy time series [J],Fuzzy Sets and Systems,107(2),235-238,1999.
[68] H. T. Liu, An integrated fuzzy time series forecasting system [J], Expert Systems with Applications 36(2009)10045-10053.
[69] H. T. Liu, M. L. Wei, An improved fuzzy forecasting method for seasonal time series [J], Expert Systems with Applications 37(2010)6310-6318.
[70] Suhartono, M. H. Lee, H. Javedani, A weighted fuzzy integrated time series for forecasting tourist arrivals [J], A. Abd Manaf et al. (Eds.): ICIEIS 2011, Part II, CCIS 252, pp. 206 – 217, 2011.
[71] T. L. Chen, , C. H. Cheng, H. J. Teoh, Fuzzy time-series based on Fibonacci sequence for stock price forecasting [J], Physica A – Statistical Mechanics and Its Applications 380(2007)377 – 390.
[72] T. L. Chen, C. H. Cheng, H. J. Teoh, High-order fuzzy time series based on multi-period adaptation model for forecasting stock markets [J], Physica A – Statistical Mechanics and Its Applications 387(2008)876-888.
[73] 吴铭峰.蒋勋.基于模糊时间序列的预测模型以上证指数为例 [J],价值工程,(2008)165-168.
[74] L. Sun, Average-based fuzzy time series models for forecasting Shanghai Compound Index [J],

International Journal of Applied Science and Engineering,2005,234-244.
[75] 张钰敏.张羽.沈晓羽.国际石油期货价格的模糊时间序列预测 [J],石油天气学报,第33卷第5期,2011.
[76] 王中宇.张海滨.刘智敏.剔除离群值的学生化残差新方法 [J],仪器仪表学报,第26卷第6期,2006.
[77] 吴柏林,张钿富,廖敏志,模糊时间序列与台湾地区中学教师人数需求之预测 [J],国立政治大学学报(73):287-312,1996.
[78] 邓聚龙.灰预测与灰决策 [M],华中科技大学出版社,2002年9月.
[79] H. S. Lee, M. T. Chou, Fuzzy forecasting based on fuzzy time series [J],International Journal of Computer Mathematics,81(7),781-789,2004.