IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Special Issue-2, March 2018 E-ISSN: 2347-2693

Ensemble Forecasting Through Learning Process

K.Amudha[1], S.Padmapriya[2], D.Saravanan[3]

[1] , Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous),
[2]Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous),
[3]Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous), Trichy

Available online at: www.ijcseonline.org

ABSTRACT- Ensemble forecasting is a widely-used numerical prediction method for modeling the progression of nonlinear dynamic systems. To calculate the future state of such systems, a set of ensemble member forecasts is generated from several runs of computer models, where each run is obtained by disquieting the starting condition or using a different model representation of the system. The ensemble mean or median is typically chosen as a point approximation for the ensemble member forecasts. These approaches are limited in that they assume each ensemble member is equally skillful and may not conserve the temporal autocorrelation of the predicted time series. To overcome these confines, new scheme present an online multi-task learning formation called ORION to estimate the optimal weights for combining the ensemble portion forecasts. Unlike other existing formulations, the proposed framework is original in that its learning algorithm must backtrack and modify its previous forecasts before making future predictions if the earlier forecasts were incorrect when verified against new observation data. We termed this strategy as online learning with restart. Our proposed framework employs a graph Laplacian regularize to ensure consistency of the predicted time series. It can also accommodate unusual types of loss functions, including insensitive and loss functions, the latter of which is particularly useful for extreme value prediction. A theoretical proof representative the convergence of our algorithm is also given. Tentative results on seasonal soil moisture forecasts from major river basins in North America exhibit the authority of ORION compared to other baseline algorithms.

KEYWORDS: Online learning, multi-task learning, ensemble forecasting

I. INTRODUCTION

Ensemble forecasting is a trendy numerical prediction method for modeling nonlinear vibrant systems, such as climate, agriculture, ecological, and hydrological systems. Specifically, the future states of the systems are predicted using computer models that simulate the physical processes governing the behavior of such systems. Since the models may not fully detain all the underlying processes as well as their parameterization accurately, their forecast errors tend to amplify with increasing lead time. Ensemble forecasting aims at obtaining more robust prediction results by combining outputs from multiple runs of the computer models. Each run is obtained by perturbing the starting stipulation or using a different model representation of the dynamic system. The forecast generated from each run corresponds to a evolution of predictions for a future time window, T, known as the forecast duration. There are altogether N sets of forecasts generated every 5 days, from April 5, 2011 to September 12, 2011. Each set of forecasts contains time series predictions generated by d ensemble members (x1; . . . ; xd) for a forecast duration T. The ultimate goal is to combine the ensemble member forecasts in such a way that produces an aggregated prediction that is consistent with the true inspection data, y. The ensemble mean or median is often used as a point estimate of the aggregated forecasts. These estimates assume that every ensemble member prediction is equally credible, and thus, their predictions should be weighted equally. However, as some ensemble members may fit the observed data less accurately than others, this may disgrace the overall analytical performance. The ensemble members were initially calibrated using soil moisture data from September 2, 2011. Their outputs for that day are therefore the same. Each ensemble member would then produce a set of forecasts for a 40-day time window, from September 7, 2011 to October 12, 2011. Though the forecasts were pretty similar at the beginning, they began to deviate with increasing lead time. Some ensemble member forecasts no longer fit the observed data well, thus affecting the accuracy of the ensemble median approach. This example illustrates the need to find out an optimal set of weights for combining the ensemble member forecasts. To deal with this need, this paper presents an online learning model that can update the weights of the ensemble members according to their predictive skills. Unlike conventional online learning, the ensemble forecasting task requires making multiple predictions for a time window of length T. As the forecast within the window are not free due to the temporal autocorrelation of the time series, the together forecasting task can be naturally cast as an online multitask regression problem. Multi-task education has been effectively used to solve multiple related learning troubles in much application, including computer vision healthcare recommender systems natural language processing and genomics. In particular, previous studies have shown that the predictive performance of various learning tasks can be improved by exploiting the commonalities among the tasks. Another difference between conventional online learning and the requirements of ensemble forecasting is that not all observation data are available when the model is updated. We call this problem online multi-task learning with partially observed data. Due to this property of the data, instead of updating the model from its most recent model, we need to backtrack and revise some of the older models when new observation data are available. In this paper, a novel approach develop a framework called ORION (which stands for Online Regularized multI-task regressiON) that uses an online learning with restart strategy to deal with the partially observed data. The framework also employs graph regularization constraints to ensure smoothness in the model parameters while taking into account the temporal autocorrelation of the predictions within each time window. A preliminary version of this work appeared in, where the ORION framework was introduced using the _-insensitive loss function to predict the weighted conditional mean of the ensemble member predictions. In this journal submission, the framework is extended to incorporate a quintile loss function, which is useful for predicting extreme values of a time series. Forecasting of extreme values is essential for applications such as climate change assessment and natural resource management due to their potential impact on human and natural systems. We also provide a theoretical proof demonstrating the convergence of our online learning algorithm. Finally, new experiments are added to evaluate sensitivity of the results to changes in the parameter setting of the ORION framework. The main contributions of this paper are as follows: and introduce the problem of online regularized multi-task regression with partially observed data and demonstrate its relevance to the ensemble forecasting task. And the present a novel framework called ORION, which uses an online learning with restart strategy to solve the problem. It also uses a graph Laplacian to capture relationships among the education tasks along with a passive forceful update scheme to optimize the insensitive loss function. To the best of our knowledge, ORION is the first multitask regression framework that has been tailored for extreme value prediction. It performed broad experiments using a real world soil moisture data set and showed that ORION outperforms several baseline algorithms, including the ensemble median, for the majority of the river basins in our data set.

II. RELATED WORK

[1] Numerical weather calculation models as well as the atmosphere itself can be viewed as nonlinear dynamical systems in which the progression depends sensitively on the initial conditions. The fact that estimates of the present state are mistaken and that mathematical models have inadequacies, leads to forecast errors that grow with mounting forecast lead time. The growth of issues depends on the flow itself. Ensemble forecasting aims at quantifying this flow-dependent forecast ambiguity. The sources of ambiguity in weather forecasting are discussed. Then, an overview is given on evaluating probabilistic forecasts and their utility compared with single forecasts. Thereafter, the symbol of doubts in ensemble forecasts is reviewed with a reputation on the preliminary condition perturbations. The review is complemented by a exhaustive explanation of the methodology to produce initial condition perturbations of the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). These perturbations are based on the foremost part of the singular value decomposition of the operator recounting the linearised dynamics over a finite time interval. The perturbations are flow-dependent as the linearization is performed with respect to a way out of the nonlinear forecast model. The extent to which the current ECMWF ensemble prediction system is proficient of predicting flow-dependent variations in uncertainty is assessed for the large-scale flow in mid-latitudes.

[2] And address the problem of computing joint sparse representation of visual signal across various kernel-based representations. Such a problem occurs logically in supervised visual detection applications where one aims to renovate a test sample with multiple features from as few training subjects as feasible. It cast the linear version of this problem into a multi-task joint covariate selection model, which can be very resourcefully optimized via kernelizable accelerated proximal gradient method. Furthermore, two kernel-view extensions of this method are provided to handle the situations where descriptors and similarity functions are in the form of kernel matrices. And the system investigate into two applications of our algorithm to feature combination: 1) fusing gray-level and LBP features for face recognition, and 2) combining multiple kernels for object categorization. Experimental results on demanding real world datasets show that the feature grouping capability of our proposed algorithm is competitive to the state-of-the art multiple kernel learning methods.

[3] This project suggests a novel algorithm for multi-task education with boosted decision trees. We learn several different learning tasks with a joint model, overtly addressing the essentials of each learning task with task-specific parameters and the commonalities between them through mutual parameters. This enables contained data

sharing and regularization. We assess our learning method on web-search ranking data sets from several countries. Here, multitask learning is particularly helpful as data sets from different countries vary largely in size because of the cost of editorial judgments. Our experiments authorize that education various tasks jointly can lead to significant improvements in performance with surprising reliability.

[4] Alzheimer's Disease (AD), the most common type of dementia, is a brutal neurodegenerative disorder. Identifying markers that can follow the progress of the disease has recently received increasing attentions in AD research. A definitive diagnosis of AD requires autopsy verification, thus many clinical/cognitive procedures counting Mini Mental State Examination (MMSE) and Alzheimer's disease Assessment Scale cognitive subscale (ADAS-Cog) have been designed to assess the cognitive status of the patients and used as significant criteria for clinical diagnosis of probable AD. This scheme suggests a multi-task schooling formulation for predicting the disease development measured by the cognitive get and choose markers predictive of the progression. Specifically, we formulate the prediction problem as a multi-task regression problem by considering the prediction at each time point as a task. We capture the intrinsic relatedness among different tasks by a temporal group Lasso regularize. The regularize consists of two components including an $\ell 2,1$-norm penalty on the regression weight vectors, which ensures that a small subset of features will be selected for the regression models at all time points, and a temporal smoothness term which ensures a small deviation between two regression models at following time points. We have performed extensive estimate using various types of data at the baseline from the Alzheimer's disease. Neuroimaging Initiative (ADNI) databases for forecast the future MMSE and ADAS-Cog scores. Our tentative studies demonstrate the effectiveness of the proposed algorithm for capturing the progression trend and the cross-sectional group differences of AD severity. Results also show that most markers selected by the planned algorithm are reliable with findings from existing cross-sectional studies.

[5] We address the problem of learning classifiers for a large number of tasks. We derive a solution that produces resampling weights which match the pool of all examples to the target distribution of any given task. Our work is forced by the problem of predicting the outcome of a therapy attempt for a patient who carries an HIV virus with a set of observed genetic properties. Such predictions need to be made for hundreds of possible combinations of drugs, some of which use similar biochemical mechanisms. Multi-task educations allow us to make prediction even for drug combinations with few or no preparation examples and significantly improve the overall prediction accuracy

III. METHODOLOGY

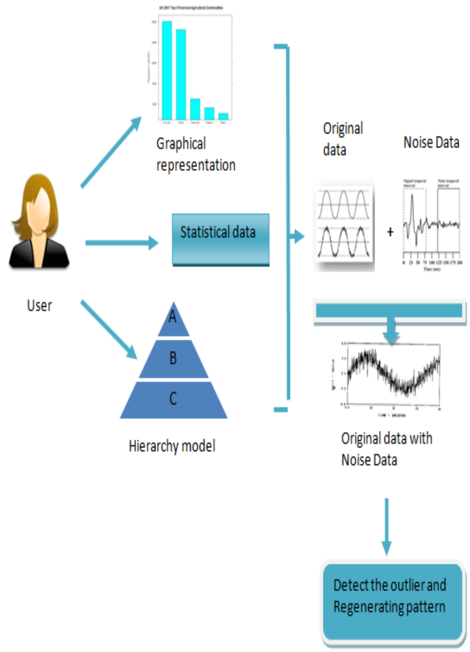
EXISINTG PROCESS

The framework is comprehensive to incorporate a quintile loss function, which is valuable for predicting excessive values of a time series. Forecasting of tremendous Values is necessary for applications such as climate change Assessment and natural resource management due to their Potential impact on human and expected systems. We also Provide a speculative proof demonstrating the convergence Of our online learning algorithm.

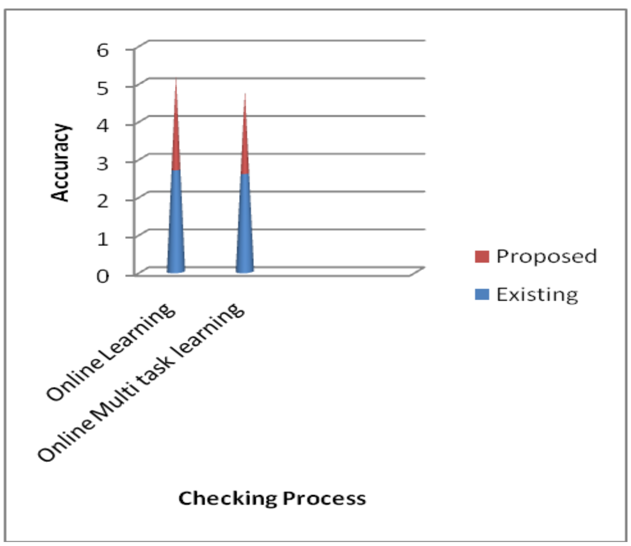
PROPOSED PROCESS

Multi-task education is an approach designed to improve analytical performance by education from multiple related tasks at the same time, taking into account the relationships and information shared among the different tasks. Depending on how information is shared among the different tasks, multi-task learning can be broadly classified into four categories. (a) Low-rank representation: This category of methods assumes that the models for different tasks share a low-rank representation, either additively or multiplicatively and incorporated both additive and multiplicative representation into their formulation with a low-dimensional feature map shared across the different tasks. (b) Explicit task relationship: This category of methods explicitly incorporates the task relationship into the multi-task learning formulation. It employed a task relationship to ensure smoothness between time series predictions. Predicting extreme value events are important for applications such as weather and hydrological forecasting due to their adverse impacts on both human and natural systems. Unfortunately, most of the existing work on multi-task learning has measured only squared or center loss functions, and thus, are not suitable for extreme value prediction. In this section, we describe an extension of the ORION framework to predict extreme values in a time series by incorporating a quintile loss function. To describe the approach, we first present the quintile regression (QR) method and introduce the quintile loss function. QR is a statistical method for estimating the conditional quintiles of a target variable as a function of its predictor variables. By focusing on the upper or lower quintiles of the distribution, this may help bias the algorithm towards learning the extreme values of the target distribution. And the proposed a regularization formulation to simultaneously learn the task relationship and task models.

ARCHITECTURE



PERFORMANCE ANALYSIS



IV. CONCLUSION

This paper presents an online regularized multi-task deterioration framework for ensemble forecasting tasks. Our framework is inimitable in that it uses an online learning with restart policy to update its models. The proposed framework is also flexible in that it can provide accommodation both insensitive and quintile loss functions. Tentative results confirm the superiority of the proposed framework compared to several baseline methods.

REFERENCES

[1] C. Tebaldi and R. Knutti, “The use of the multi-model ensemble in probabilistic climate projections,” Philosoph. Trans. Roy. Soc. London A: Math. Phys. Eng. Sci., vol. 365, no. 1857, pp. 2053–2075, 2007.

[2] P. Cantelaube and J.-M. Terres, “Seasonal weather forecasts for crop yield modeling in Europe,” Tellus A, vol. 57, no. 3, pp. 476– 487, 2005.

[3] M. B. Araujo and M. New, “Ensemble forecasting of species distributions,” Trends Ecology Evol., vol. 22, no. 1, pp. 42–47, 2007.

[4] L. Luo and E. F. Wood, “Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the Eastern United States,” J. Hydrometeorology, vol. 9, pp. 866– 884, 2008.

[5] M. Leutbecher and T. Palmer, “Ensemble forecasting,” J. Comput. Phys., vol. 227, pp. 3515–3539, 2008.

[6] X.-T. Yuan and S. Yan, “Visual classification with multi-task joint sparse representation,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 3493–3500.

[7] X. Wang, C. Zhang, and Z. Zhang, “Boosted multi-task learning for face verification with applications to Web image and video search,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 142–149.

[8] J. Zhou, L. Yuan, J. Liu, and J. Ye, “A multi-task learning formulation for predicting disease progression,” in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 814– 822.

[9] S. Bickel, J. Bogojeska, T. Lengauer, and T. Scheffer, “Multi-task learning for HIV therapy screening,” in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 56–63.

[10] X. Ning and G. Karypis, “Multi-task learning for recommender system,” in Proc. 2nd Asian Conf. Mach. Learn., 2010, pp. 269–284.

[11] J. Wang, S. C. Hoi, P. Zhao, and Z.-Y. Liu, “Online multi-task collaborative filtering for on-the-fly recommender systems,” in Proc. 7th ACM Conf. Recommender Syst., 2013, pp. 237–244. [12] R. K. Ando and T. Zhang, “A framework for learning predictive structures from multiple tasks and unlabeled data,” J. Mach. Learm. Res., vol. 6, pp. 1817–1853, Dec. 2005.

[13] G. Obozinski, B. Taskar, and M. Jordan, “Joint covariate selection and joint subspace selection for multiple classification problems,” Statist. Comput., vol. 20, no. 2, pp. 231–252, 2010.

[14] J. Xu, P.-N. Tan, and L. Luo, “ORION: Online regularized multItask regressiON and its application to ensemble forecasting,” in Proc. IEEE Int. Conf. Data Mining, 2014, pp. 1061–1066.

[15] H. Cheng, P.-N. Tan, J. Jing, and J. Scripps, “Multi-step ahead time series prediction,” in Proc. 10th Pacific-Asia Conf. Advances Knowl. Discovery Data Mining, 2006, pp. 765–774.

