Explainable Markov Chain Based Pattern Forecasting

Debdeep Paul
Singapore Technology Center
Panasonic Industrial Devices Pte. Ltd., Singapore
debdeep.paul@sg.panasonic.com

Chandra Wijaya
Singapore Technology Center
Panasonic Industrial Devices Pte. Ltd., Singapore
chandra.wijaya@sg.panasonic.com

Sahim Yamaura
Singapore Technology Center
Panasonic Industrial Devices Pte. Ltd., Singapore
sahim.yamaura@sg.panasonic.com

Koji Miura
Engineering Division
Panasonic Industry Co., Ltd
Japan
miura.koji@jp.panasonic.com

Yosuke Tajika
Engineering Division
Panasonic Industry Co., Ltd
Japan
tajika.yosuke@jp.panasonic.com

***Abstract*—We consider the trend or pattern forecasting for demand timeseries in a business-to-business supply chain where demand exhibits high volatilities, non-stationarities, and skewness. We develop a pattern forecasting system by designing a data driven, feature dependent Markov chain-based framework. To increase adoption of AI based techniques among the various stakeholders we address the aspect of explainability. We define two metrices to evaluate the quality of explainability. To provide guidelines on selecting different attributes of our pipeline, we compare between feature selection methods from two families, one advanced and one traditional. We evaluate the proposed strategy on a real dataset and observe a sparsity promoting feature selection show a sparse feature selection method outperforms a conventional decision tree-based feature selection method.**

Keywords*—*Markov chain, explainable artificial intelligence.

I. Introduction

This paper deals with the pattern forecasting of demand timeseries for B2B industrial products [1], which is less explored and often remain esoteric to the practitioners as compared to retail B2C products. Typically, the demand for different types of B2B products is highly volatile, nonstationary and comes from a wide class of statistical distributions. The supply chain managers require the pattern forecast that is used in upstream decision support systems as only numerical forecast may have unacceptable errors. On the other hand, for companies with well-established traditional business processes implementing black-box machine learning (ML) based approaches is particularly challenging. This can be addressed by providing explainability to the business users to establish connection with their domain expertise and thereby build trust in the advanced ML technologies [2]. Our contributions are twofold. Firstly, we propose a method for encoding a continuous variable to a discrete number of states and model the transitions across different states through a data driven *feature dependent Markov chain* (*FDMC*). Secondly, we propose a strategy for quantitative modeling of the pattern of interest through a score as a function of the respective entries in the *transition probability matrix* (TPM) corresponding to the Markov chain. Thereafter based on that score we quantify the likelihood of emergence of the pattern of interest and propose a method to explain that likelihood based on the input features to provide additional insights to domain experts and gain their trust. Fig.1 given below describes the system architecture.

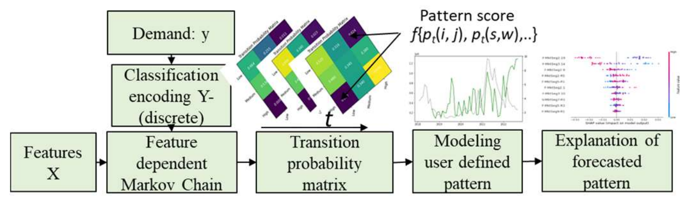


Fig. 1. System architecture

II. Feature dependend Markov chain based pattern modeling

We now present the framework to model different user defined patterns through versatile, widely adopted discrete time Markov chain (DTMC) model. More formally, a DTMC is a stochastic process $\{Y_t, t=0,1,...T\}$, where Y_t is the state at time step t and such that, $\forall t \geq 0, \forall i, j$, and $\forall i_0, i_1 ... i_{t-1}$,

$$\begin{aligned} &\mathrm{P}\{Y_{t+1}=s_j \mid Y_t=s_i, Y_{t-1}=s_{i_{t-1}}, ..., Y_0=s_{i_0}\} \\ &= \mathrm{P}\{Y_{t+1}=s_j \mid Y_t=s_i\}=P_{i,j}, \end{aligned} \tag{1}$$

where $P_{i,j}$ is independent of the timestep and of past history. The transition probability matrix (TPM) associated with any DTMC is a matrix, $\mathbf{P}_t$, for timeslot t, whose $(i, j)^{\text{th}}$ entry $P_{i,j}$ represents the probability of going to state j on the next transition given the current state is i. We adopt the method proposed in [3] to compute the TPM through a set of features that may be considered as a data driven approach with a specified feature matrix for every timeslot under consideration, referred to as *FDMC* for the rest of this paper.

III. Explainability

This paper provides two metrices to evaluate the quality of the explainability namely, i) relevance, and, ii) informativeness. Relevance is defined as the extent to which the explanation is consistent with the inherent physical process and is evaluated based on user feedback/annotations. Informativeness is defined as the amount of information contained in the explanation. We measure this by the sharpness of feature attributes such as the variance in the importance score. Higher the variance higher is the information content. A popular framework for interpreting predictions, namely SHAP is used for explainability [4]. Our input data is high dimensional where the number of columns is much higher than number of rows, therefore the value of the appropriateness in choosing the feature selection method is significantly high. We leverage a recently developed lasso type feature selection method that enforces feature sparsity, controllability, namely LassoNet [5], referred to as *LN-FS* and compared the performance with a conventional method based on decision tree; referred to as *DT-FS* in terms of accuracy and explainability.

IV. EXPERIEMNTAL SETUP AND RESULTS

We consider the sales of a generic purpose industrial component as the variable whose trend is to be predicted. In our dataset we have 56 months demand and 96 related indices used as features, referred to as X_1 to X_{96}. The first 50 months are used as training set and remaining 6 are used as testing set in which the performance of the model is evaluated. A three state *FDMC* is leveraged where the TPM for timeslot t is a 3×3 matrix. First 9 months is used to train the *FDMC*. In this experiment we model the nature or intensity of the change that we call as, i) steady state, ii) moderate fluctuation, and, iii) drastic fluctuation, respectively.

$$\mu_t^1 = \gamma^1 e^{(p_t(1,1)+p_t(2,2)+p_t(3,3))}, \text{ for } \forall t. \quad (2)$$

$$\mu_t^2 = \gamma^2 e^{(p_t(1,2)+p_t(2,1)+p_t(2,3)+p_t(3,2))}, \text{ for } \forall t. \quad (3)$$

$$\mu_t^3 = \gamma^3 e^{(p_t(1,3)+p_t(3,1))}, \text{ for } \forall t, \quad (4)$$

where the constants γ^i is used to rationalize the scores. Equation (2) is a function of diagonal entries of $\mathbf{P}_t$ that corresponds to transition to same state. Equation (3) is a function of probabilities corresponding to transitions between adjacent states. Finally, equation (4) represents a function of probabilities corresponding to transitions between non-adjacent states. The scores corresponding to different states are computed by normalizing with respect to the total score and is obtained by:

$$\Delta_t^S = \frac{\mu_t^1}{\sum_{i=1}^{3}\mu_t^i} \quad \Delta_t^M = \frac{\mu_t^2}{\sum_{i=1}^{3}\mu_t^i} \quad \Delta_t^D = \frac{\mu_t^3}{\sum_{i=1}^{3}\mu_t^i}, \text{ for } \forall t \quad (5)$$

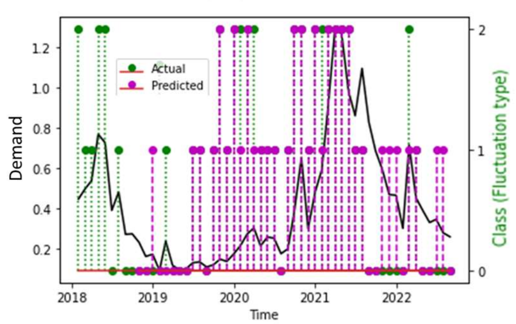
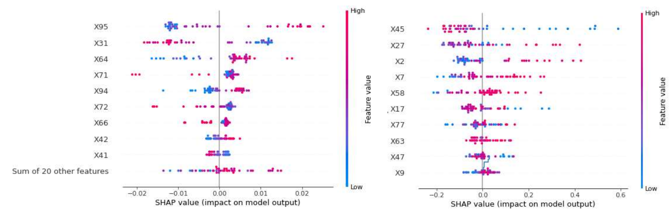


Fig. 2. Actual vs prediction of classes through FDMC

where Δ_t^S, Δ_t^M and Δ_t^D denotes the scores corresponding to steady state, moderate fluctuation, and drastic fluctuation, respectively. We make the prediction based on the maximum among these in (5). Fig.2 shows the actual demand, respective fluctuation type and the fluctuation type predicted, that can be considered as classes. The prediction accuracy obtained with *DT-FS* and *LN-FS* are 0.5 and 0.67, respectively.



(a) (b)

Fig. 3. Experiment-1: SHAP values of features: (a) *DT-FS*, (b) *LN-FS*

TABLE I. COMPARISON BETWEEN DIFFERENT FEATURE SELECTION METHODS

Feature selection	*SHAP-Variance*	*Accuracy*	*F1-Score*
DT-FS	8.36×10^{-6}	0.5	0.43
LN-FS	0.38	0.67	0.51

Fig. 3 shows the contributions from top influencing features, where X95, X45 are the indices related to global semiconductor market that is very relevant to the product under consideration. The feature contributions are quite relevant and informative. The variance in the SHAP values is much higher with *LN-FS* than *DT-FS*, that corresponds to informativeness of explainability. We observe that both in terms of explainability and prediction accuracy *LN-FS* outperforms *DT-FS*.

V. CONCLUSIONS

In this paper we focused on pattern modeling and forecasting for demand timeseries of B2B products by encoding the continuous variable (demand) to discrete number of states and modeling through a data driven feature dependent Markov chain. From the entries of the transition probability metrices (one for each timeslot), the score corresponding to a specific pattern of interest is appropriately quantified. Thereafter we developed methods for explaining the forecasting of specific user defined patterns through a popular method for post hoc explainability. We performed a case study with real world dataset and compared two feature selection methods for fitting the feature dependent Markov chain and show a sparse feature selection method outperforms a conventional decision tree-based feature selection method.

REFERENCES

[1] M. C. Cohen, P.-E. Gras, A. Pentecoste, and R. Zhang, *Demand prediction in retail: A practical guide to leverage data and predictive analytics*. Springer, 2022.

[2] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:.08608,* 2017.

[3] S. Barratt and S. Boyd, "Fitting feature-dependent Markov chains," *Journal of Global Optimization,* pp. 1-18, 2022.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems,* vol. 30, 2017.

[5] I. Lemhadri, F. Ruan, L. Abraham, and R. Tibshirani, "Lassonet: A neural network with feature sparsity," *The Journal of Machine Learning Research,* vol. 22, no. 1, pp. 5633-5661, 2021.