

USING MATRIX FACTORIZATION FOR THE PREDICTION OF ELECTRICAL QUANTITIES

Fabian LECRON
UMONS - Belgium
fabian.lecron@umons.ac.be

Zacharie DE GRÈVE
UMONS - Belgium
zacharie.degreve@umons.ac.be

François VALLÉE
UMONS - Belgium
francois.vallee@umons.ac.be

Gerard MOR
CIMNE (UPC) - Spain
gmor@cimne.upc.edu

Daniel PÉREZ
CIMNE (UPC) - Spain
sanchez@cimne.upc.edu

Stoyan DANOV
CIMNE (UPC) - Spain
sdanov@cimne.upc.edu

Jordi CIPRIANO
CIMNE (UPC) - Spain
cipriano@cimne.upc.edu

ABSTRACT

The prediction task is attracting more and more attention among the power system community. Accurate predictions of electrical quantities up to a few hours ahead (e.g. renewable production, electrical load, etc.) are for instance crucial for Distribution System Operators (DSOs) to operate their network in presence of a high share of renewables, or for energy producers to maximize their profits by optimizing their portfolio management. In the literature, statistical approaches are usually proposed to predict electrical quantities. In the present paper, we present a novel method based on matrix factorization. Our approach is inspired by the literature on data mining and knowledge discovery and the methodologies involved in recommender systems. The idea is to transpose the problem of predicting ratings in a recommender system to a problem of forecasting electrical quantities in a power system. Preliminary results on a real wind speed dataset tend to show that the matrix factorization model provides similar results than ARIMA models in terms of accuracy (MAE and RMSE). Our approach is nevertheless highly scalable and can deal with noisy data (e.g. missing data).

INTRODUCTION

The prediction task is attracting more and more attention among the power system community. Accurate predictions of electrical quantities up to a few hours ahead (e.g. renewable production, electrical load, etc.) are for instance crucial for Distribution System Operators (DSOs) to operate their network in presence of a high share of renewables, or for energy producers to maximize their profits by optimizing their portfolio management. Even the importance of predicting unknown data in the past has been noted in [9].

In the literature, statistical approaches are usually proposed to predict electrical quantities such as electrical load or wind speed data (in relation with the prediction of renewable production). First predictive models were auto regressive (AR), auto regressive moving average (ARMA) and auto regressive integrated (ARIMA) models [3]. For instance, a periodic autoregressive model was used in [6] to

perform customer segmentation based on the forecast of short-term load. Wind speed data were also predicted by ARMA models in [5].

Other methods involving machine learning techniques can be identified. In this context, artificial neural networks have been extensively used in the literature [14, 4, 3, 7, 15]. Other specific approaches relying on Generalized Principal Component Analysis (GPCA) [10] or extreme learning machines [1] were also proposed.

In the present paper, we introduce a novel method based on matrix factorization to predict electrical quantities. Our approach is inspired by the literature on data mining and knowledge discovery and the methodologies involved in recommender systems.

Recommender systems are powerful tools that provide people with recommendations of items they will appreciate, based on their past preferences, history of purchase, etc. A standard way to perform this task is to consider a matrix \mathbf{X} of ratings where each row stands for a user and each column for an item (e.g. a movie). An element x_{ij} of the matrix is a rating given by the user i to the item j . The objective is to predict the unknown x_{ij} 's in the matrix \mathbf{X} . One of the most effective technique to do so is Matrix Factorization (MF). It has been popularized by the winners of the Netflix prize in [12]. The idea is to decompose the incomplete matrix \mathbf{X} into the product of two matrices \mathbf{W} and \mathbf{H} . Finally, the unknown values in the matrix \mathbf{X} can be obtained by computing \mathbf{WH} .

Our contribution is to propose a matrix factorization method to forecast electrical quantities by transposing the problem of predicting ratings in a recommender system to a problem of forecasting (e.g. electrical load or wind speed for estimating renewable production). As a consequence, each row in the matrix \mathbf{X} stands for a given period of time (a day, a week, etc.) and each column stands for the temporal granularity of the data (quarter-hourly data, hourly data, etc.). To illustrate our purpose, let us consider wind speed data (which can be transformed into the power produced by a wind farm using a power curve). The matrix \mathbf{X} is first built with previous measurements of wind speed. Then, the

last row is filled with the last known measurements of the current day. As a consequence, the last columns of the last row are unknown and can be forecasted by factorizing the matrix.

A very few works are dedicated to the prediction of electrical quantities based on matrix factorization. Authors in [11] used sparse matrix factorization for that matter. Very recently, temporal dependency was included into a regularized matrix factorization procedure [16]. The interest of these approaches is to be highly scalable and to deal with noisy data (e.g. missing data).

In this work, a real dataset of wind speed data was extracted from the Royal Netherlands Meteorological Institute database. The prediction accuracy was evaluated by computing the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). The method has been compared with the persistence model and ARIMA prediction models. First results tend to show that the matrix factorization model provides similar results than ARIMA models in terms of accuracy (MAE and RMSE).

METHODS

Matrix Factorization

Matrix Factorization allows to decompose a given matrix into the product of two matrices.

It can be proven (cf. [13]) that the rank factorization of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ of rank f is a factorization :

$$\mathbf{X} = \mathbf{W}\mathbf{H}, \quad (1)$$

where $\mathbf{W} \in \mathbb{R}^{m \times f}$, $\mathbf{H} \in \mathbb{R}^{f \times n}$, and $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{W}) = \text{rank}(\mathbf{H}) = f$.

Based on that, we can define the low-rank matrix approximation of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$. The idea is to find the best approximation $\hat{\mathbf{X}} \approx \mathbf{X}$ such that $\text{rank}(\hat{\mathbf{X}}) \leq f$. This problem can be written as [8]:

$$\min_{\hat{\mathbf{X}}} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 \quad \text{such that } \text{rank}(\hat{\mathbf{X}}) \leq f, \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm.

Since $\text{rank}(\hat{\mathbf{X}}) \leq f$, $\hat{\mathbf{X}}$ can be factorized following equation (1). The problem defined at equation (2) becomes:

$$\min_{\mathbf{W}, \mathbf{H}} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2, \quad (3)$$

where $\hat{\mathbf{X}} = \mathbf{W}\mathbf{H}$.

Matrix Factorization for Missing Data

When data are missing in the matrix \mathbf{X} , it is possible to approximate \mathbf{X} and then recover the missing data. In the collaborative filtering community, matrix factorization techniques are widely used to recover missing data, for instance

to build recommender systems. Therefore, we will first introduce recommender systems and how missing data are recovered and then transpose these concepts to the problem of prediction of electrical quantities.

Recommender systems try to provide people with recommendations of items they will appreciate, based on their past preferences, history of purchase, and demographic information. Data in most standard recommender systems are often represented with a matrix since it is characterized by three sets: a set of users $U = \{u_1, u_2, \dots, u_m\}$ and a set of items $I = \{i_1, i_2, \dots, i_n\}$ and a set of evaluations. For instance, if we want to recommend movies, evaluation could be a rating r_{ui} given by a user u to a movie i . Therefore, a $m \times n$ matrix \mathbf{R} is built with all the known ratings. By predicting the unknown ratings, one can know the predicted interest of one user for an item and make recommendations accordingly. Let us consider a matrix \mathbf{R} with 3 users and 3 movies. User 1 has seen movie 2, user 2 has seen movie 3, and user 3 has seen movie 1. The matrix \mathbf{R} would be:

$$\mathbf{R} = \begin{pmatrix} ? & r_{u_1 i_2} & ? \\ ? & ? & r_{u_2 i_3} \\ r_{u_3 i_1} & ? & ? \end{pmatrix}, \quad (4)$$

where “?” are unknown ratings.

In [12], Koren et al. present a basic matrix factorization model based on equation (3). With this technique, a rating r_{ui} can be approximated by \hat{r}_{ui} , the inner product of user-item interactions. If we consider matrices \mathbf{W} and \mathbf{H} described in equation (3), we can write that:

$$\hat{r}_{ui} = w_u h_i, \quad (5)$$

where w_u is a row of \mathbf{W} representing user u and h_i is a column of \mathbf{H} representing item i .

The challenge is then to compute all the w_u and h_i . For that matter, it is described in [12] that w_u and h_i can be obtained by modeling directly the known ratings only, while avoiding overfitting thanks to a regularized model. As a consequence, the system proposed in [12] will perform a matrix factorization by only considering known ratings and the objective function to be minimized is written:

$$\min_{\mathbf{w}^*, \mathbf{h}^*} \sum_{(u,i) \in \kappa} (r_{ui} - w_u h_i)^2 + \lambda (\|w_u\|^2 + \|h_i\|^2) \quad (6)$$

where κ is the set of the (u, i) pairs for which r_{ui} is known, λ is a parameter controlling the regularization to avoid overfitting.

Finally, computing all the w_u and h_i allows to obtain the matrix \mathbf{W} and \mathbf{H} . The product between \mathbf{W} and \mathbf{H} then gives an approximation of matrix \mathbf{R} .

Problem defined at equation (6) can be solved with Stochastic Gradient Descent (SGD). Further details on how to implement it are provided in [12].

Prediction of Electrical Quantities

To transpose the problem of recommender systems to the problem of predicting electrical quantities, we first have to adapt the data. For recommender systems, we had a matrix with r_{ui} , a rating given by a user to an item. Now, we have an electrical quantity (e.g. electrical load), a period of time (e.g. a day, a week, etc.) and a temporal granularity of the data (e.g. quarter-hourly data, hourly data, etc.).

Let us imagine that today is Saturday and that we want to predict the electrical load of one user for the next day, i.e. Sunday. Furthermore, we have recorded the electrical load corresponding to this user for the last three weeks. As a consequence, we have a small matrix with 3 rows (the last row is corresponding to the current week) and 7 columns (column 1 is Monday and column 7 is Sunday). We know all the r_{ui} in the matrix, except the one at the intersection of the last row and the last column. Our matrix \mathbf{R} is therefore written:

$$\mathbf{R} = \begin{pmatrix} r_{u_1 i_1} & r_{u_1 i_2} & \cdots & r_{u_1 i_6} & r_{u_1 i_7} \\ r_{u_2 i_1} & r_{u_2 i_2} & \cdots & r_{u_2 i_6} & r_{u_2 i_7} \\ r_{u_3 i_1} & r_{u_3 i_2} & \cdots & r_{u_3 i_6} & ? \end{pmatrix}, \quad (7)$$

where “?” is the unknown electrical load.

Actually, matrix at equation (7) has a similar structure than matrix at equation (4) which characterizes data of a recommender system. The only difference is that in the case of the prediction of electrical quantities, only one row and the last columns are concerned by unknown values, since the data to be predicted are in the future.

As a consequence, the unknown electrical load of matrix \mathbf{R} at equation (7) can be found with the factorization of \mathbf{R} , following the equation (6).

Finally, we provide an illustration of a specific behavior when we predict electrical quantities with matrix factorization. If we consider the matrix at equation (7), u_3 is the current week, i_1 is Monday and i_7 is Sunday. As we proposed, let us consider that we are on Saturday and that we want to predict the electrical load for Sunday but also for the Monday of the next week. We could add a fourth row in our matrix but this row would be characterized by only unknown values. Therefore, we have to shift the data in the matrix in such a way that the Monday of the fourth week appears on the third row. Such a shift implies:

$$\mathbf{R} = \begin{pmatrix} r_{u_1 i_2} & r_{u_1 i_3} & \cdots & r_{u_1 i_7} & r_{u_2 i_1} \\ r_{u_2 i_3} & r_{u_2 i_4} & \cdots & r_{u_2 i_7} & r_{u_3 i_1} \\ r_{u_3 i_2} & r_{u_3 i_3} & \cdots & ? & ? \end{pmatrix}, \quad (8)$$

where “?” are the unknown electrical loads.

EXPERIMENTS

Data and Methods

In the present paper, we decided to perform experiments on wind speed data as a first application.

The dataset used in these experiments contains wind speeds measured in Rotterdam, the Netherlands, from 1986 to 1995. Data were obtained on the website of the Koninklijk Nederlands Meteorologisch Instituut (KNMI). The wind speeds were measured every hour. As a consequence, the resulting matrix \mathbf{R} has 3,652 rows and 24 columns. The dataset was divided into one training set containing the first 9 years of data (from 1986 to 1994). The last year (1995) was used as a test set (see Fig. 1). The matrix for the training set ($\mathbf{R}_{\text{training}}$) and the matrix for the test set (\mathbf{R}_{test}) have the following form:

$$\mathbf{R}_{\text{training}} = \begin{pmatrix} r_{u_1 i_1} & r_{u_1 i_2} & \cdots & r_{u_1 i_{24}} \\ r_{u_2 i_1} & r_{u_2 i_2} & \cdots & r_{u_2 i_{24}} \\ \vdots & \vdots & \cdots & \vdots \\ r_{u_{3287} i_1} & r_{u_{3287} i_2} & \cdots & r_{u_{3287} i_{24}} \end{pmatrix}, \quad (9)$$

$$\mathbf{R}_{\text{test}} = \begin{pmatrix} r_{u_{3288} i_1} & r_{u_{3288} i_2} & \cdots & r_{u_{3288} i_{24}} \\ r_{u_{3289} i_1} & r_{u_{3289} i_2} & \cdots & r_{u_{3289} i_{24}} \\ \vdots & \vdots & \cdots & \vdots \\ r_{u_{3652} i_1} & r_{u_{3652} i_2} & \cdots & r_{u_{3652} i_{24}} \end{pmatrix}, \quad (10)$$

First, equation (6) is used to factorize matrix $\mathbf{R}_{\text{training}}$, according to:

$$\mathbf{R}_{\text{training}} = \mathbf{W}_{\text{training}} \mathbf{H}_{\text{training}}, \quad (11)$$

where $\mathbf{W}_{\text{training}} \in \mathbb{R}^{3287 \times f}$ and $\mathbf{H}_{\text{training}} \in \mathbb{R}^{f \times 24}$.

If we consider that we want to predict the 24th hour of the first day of 1995 (i.e. we want to predict $r_{u_{3288} i_{24}}$ in matrix \mathbf{R}_{test}), we need to create a new matrix by adding the first row of matrix \mathbf{R}_{test} to matrix $\mathbf{R}_{\text{training}}$. We obtain:

$$\mathbf{R}_{\text{training,new}} = \begin{pmatrix} r_{u_1 i_1} & r_{u_1 i_2} & \cdots & r_{u_1 i_{24}} \\ r_{u_2 i_1} & r_{u_2 i_2} & \cdots & r_{u_2 i_{24}} \\ \vdots & \vdots & \cdots & \vdots \\ r_{u_{3288} i_1} & r_{u_{3288} i_2} & \cdots & ? \end{pmatrix}, \quad (12)$$

where “?” is the unknown wind speed we want to predict.

On this new matrix, we only need to use equation (6) to obtain the last row of the new matrix \mathbf{W} which corresponds to the row we have added to $\mathbf{R}_{\text{training}}$. This process is repeated for all the values contained in \mathbf{R}_{test} .

The approach described in the present work is mainly influenced by three parameters: the number of factors conditioning the rank of the approximated matrix \mathbf{WH} , the regularization parameter λ in the optimization problem at equation (6) and the number of iterations needed by the gradient descent algorithm to converge to an adequate solution. The optimal value of these parameters has been calculated by performing an exhaustive search among values manually defined. The optimal values were chosen as the ones providing the lowest MAE resulting from the experiments.

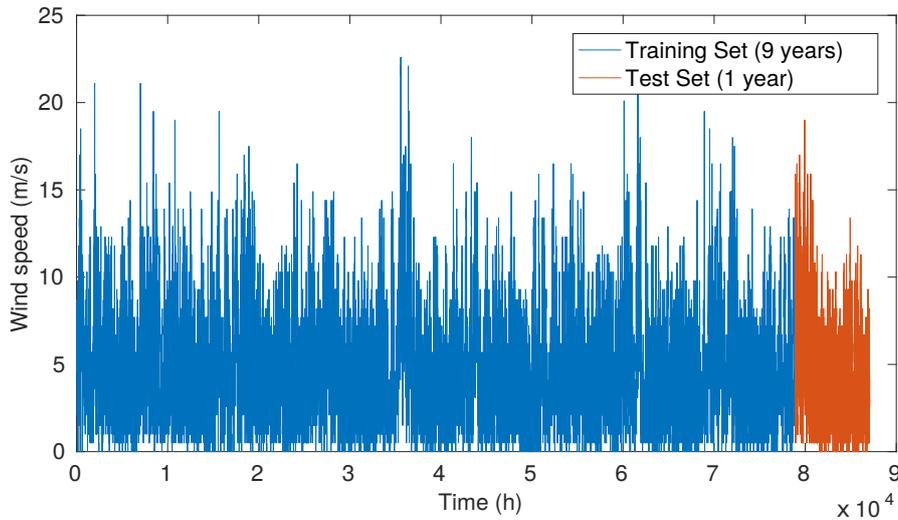


Fig. 1: Wind speed dataset used for the experiments

Results

In this section, we evaluate the behavior of our prediction approach based on matrix factorization regarding two criteria:

1. Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |\hat{r}_t - r_t| \quad (13)$$

2. Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n |\hat{r}_t - r_t|^2} \quad (14)$$

In equations (13) and (14), n is the number of values in the test set, \hat{r}_t is a predicted value, and r_t is the real value corresponding to \hat{r}_t .

Two methods were used in order to evaluate and compare our approach:

1. Persistence model: it assumes that the future value to be predicted is equal to the last known value.
2. ARIMA model [2]: we used the ARIMA forecast tool in Matlab to implement this approach.

The analysis of Tables 1, 2, and 3 shows that at a short time horizon (1h), the ARIMA model is the best but the persistence model presents an interesting precision too. From an horizon time of 3h, our approach is very competitive with the ARIMA model in terms of accuracy (MAE and RMSE).

Persistence Model					
Horizon	1h	2h	3h	4h	5h
MAE	0.8061	1.0466	1.2405	1.4276	1.5960
RMSE	1.1445	1.4589	1.7074	1.9338	2.1376

Table 1: MAE and RMSE values for the persistence model

ARIMA Model					
Horizon	1h	2h	3h	4h	5h
MAE	0.8041	1.0290	1.2081	1.3731	1.5162
RMSE	1.1077	1.4011	1.6240	1.8163	1.9799

Table 2: MAE and RMSE values for the ARIMA model

CONCLUSION

In this paper, we proposed a novel method based on matrix factorization to predict electrical quantities. Our approach is inspired by the literature on data mining and knowledge discovery and the methodologies involved in recommender systems. Our contribution was to propose a matrix factorization method by transposing the problem of predicting ratings in a recommender system to a problem of forecasting. Experiments were performed on a real wind speed dataset from the Koninklijk Nederlands Meteorologisch Instituut (KNMI). Results showed that our approach is competitive with ARIMA models. However, an interesting characteristic of matrix factorization is its ability to be highly scalable and to deal with noisy data or with missing data. In the future, it would be interesting to evaluate the impact with larger datasets and with noisy data on

Our approach					
Horizon	1h	2h	3h	4h	5h
MAE	0.8954	1.0692	1.2197	1.3593	1.4844
RMSE	1.2155	1.4445	1.6361	1.8103	1.9687

Table 3: MAE and RMSE values for our approach

the precision of prediction. Another interesting property of matrix factorization is the possibility to introduce an a priori knowledge into the mathematical formulation (e.g. the results of an independent ARIMA forecast). By doing so, accuracy is therefore expected to improve, placing matrix factorization as a promising alternative to forecast electrical quantities.

REFERENCES

- [1] R. Ak, O. Fink, and E. Zio. Two machine learning approaches for short-term wind speed time-series prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 27(8):1734–1747, 2016.
- [2] G. Box, G. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control*. NJ: Prentice Hall, 3rd edition, 1994.
- [3] N. Ding, C. Benoit, G. Foggia, Y. Bsanger, and F. Wurtz. Neural network-based model design for short-term load forecast in distribution systems. *IEEE Transactions on Power Systems*, 31(1):72–81, 2016.
- [4] B. Doucoure, K. Agbossou, and A. Cardenas. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renewable Energy*, 92:202 – 211, 2016.
- [5] E. Erdem and J. Shi. Arma based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 88(4):1405 – 1414, 2011.
- [6] M. Espinoza, C. Joye, R. Belmans, and B. D. Moor. Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series. *IEEE Transactions on Power Systems*, 20(3):1622–1630, 2005.
- [7] E. Garcia, L. Canha, and P. Pereira. Hybrid forecasting techniques applied to distribution systems for proactive voltage control. In *23rd International Conference on Electricity Distribution (CIRED 2015)*, pages 1–5, 2015.
- [8] N. Gillis. *Nonnegative Matrix Factorization: Complexity, Algorithms and Applications*. PhD thesis, Université catholique de Louvain, 2011.
- [9] F. Hocaoglu and M. Kurban. The effect of missing wind speed data on wind power estimation. In *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, volume 4881, pages 107–114. 2007.
- [10] Q. Hu, P. Su, D. Yu, and J. Liu. Pattern-based wind speed prediction based on generalized principal component analysis. *IEEE Transactions on Sustainable Energy*, 5(3):866–874, 2014.
- [11] S. Kim and G. B. Giannakis. Load forecasting via low rank plus sparse matrix factorization. In *2013 Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, November 3-6, 2013*, pages 1682–1686, 2013.
- [12] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
- [13] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, 2000.
- [14] T. Senjyu, H. Takara, K. Uezato, and T. Funabashi. One-hour-ahead load forecasting using neural network. *IEEE Transactions on Power Systems*, 17(1):113–118, 2002.
- [15] P. Siripitayananon, H.-C. Chen, and K.-R. Jin. Estimating missing data of wind speeds using neural network. In *SoutheastCon, 2002. Proceedings IEEE*, pages 343–348, 2002.
- [16] H.-F. Yu, N. Rao, and I. S. Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 847–855. 2016.