

# Using matrix factorisation for the prediction of electrical quantities

Fabian Lecron<sup>1</sup> ✉, Zacharie De Grève<sup>1</sup>, François Vallée<sup>1</sup>, Gerard Mor<sup>2</sup>, Daniel Pérez<sup>2</sup>, Stoyan Danov<sup>2</sup>, Jordi Cipriano<sup>2</sup>

<sup>1</sup>University of Mons, (UMONS), Faculty of Engineering, Mons, Belgium

<sup>2</sup>Universitat Politècnica de Catalunya (UPC), International Centre for Numerical Methods in Engineering (CIMNE), Terrassa, Spain

✉ E-mail: fabian.lecron@umons.ac.be

**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 to operate their network in the presence of a high share of renewables, or for energy producers to maximise their profits by optimising their portfolio management. In the literature, statistical approaches are usually proposed to predict electrical quantities. In the present study, the authors present a novel method based on matrix factorisation. The authors' 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 factorisation model provides similar results than auto regressive integrated models in terms of accuracy (MAE and RMSE). The authors' approach is nevertheless highly scalable and can deal with noisy data (e.g. missing data).

## 1 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 to operate their network in the presence of a high share of renewables, or for energy producers to maximise their profits by optimising their portfolio management. Even the importance of predicting unknown data in the past has been noted in [1].

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, auto regressive moving average (ARMA) and auto regressive integrated (ARIMA) models [2]. For instance, a periodic autoregressive model was used in [3] to perform customer segmentation based on the forecast of short-term load. Wind speed data were also predicted by ARMA models in [4].

Other methods involving machine-learning techniques can be identified. In this context, artificial neural networks have been extensively used in the literature [2, 5–8]. Other specific approaches relying on generalised principal component analysis [9] or extreme learning machines [10] were also proposed.

In the present paper, we introduce a novel method based on matrix factorisation 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  $X$  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  $X$ . One of the most effective techniques to do so is matrix factorisation (MF). It has been popularised by the winners of the Netflix prize in [11]. The idea is to decompose the incomplete

matrix  $X$  into the product of two matrices  $W$  and  $H$ . Finally, the unknown values in the matrix  $X$  can be obtained by computing  $WH$ .

Our contribution is to propose a MF 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  $X$  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  $X$  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 factorising the matrix.

A very few works are dedicated to the prediction of electrical quantities based on MF. Authors in [12] used sparse MF for that matter. Very recently, temporal dependency was included into a regularised MF procedure [13]. 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 MF model provides similar results than ARIMA models in terms of accuracy (MAE and RMSE).

## 2 Methods

### 2.1 Matrix factorisation

MF allows us to decompose a given matrix into the product of two matrices.

It can be proven [14] that the rank factorisation of a matrix  $X \in \mathbb{R}^{m \times n}$  of rank  $f$  is a factorisation:

$$X = WH, \quad (1)$$

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

Based on that, we can define the low-rank matrix approximation of a matrix  $X \in \mathbb{R}^{m \times n}$ . The idea is to find the best approximation  $\hat{X} \simeq X$  such that  $\text{rank}(\hat{X}) \leq f$ . This problem can be written as [15]

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

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

Since  $\text{rank}(\hat{X}) \leq f$ ,  $\hat{X}$  can be factorised following (1). The problem defined in (2) becomes

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

where  $\hat{X} = WH$ .

## 2.2 MF for missing data

When data are missing in the matrix  $X$ , it is possible to approximate  $X$  and then recover the missing data. In the collaborative filtering community, MF 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 characterised 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  $R$  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  $R$  with three users and three movies. User 1 has seen movie 2, user 2 has seen movie 3, and user 3 has seen movie 1. The matrix  $R$  would be

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

where “?” are unknown ratings.

In [11], Koren *et al.* present a basic MF model based on (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  $W$  and  $H$  described in (3), we can write that

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

where  $w_u$  is a row of  $W$  representing user  $u$  and  $h_i$  is a column of  $H$  representing item  $i$ .

The challenge is then to compute all the  $w_u$  and  $h_i$ . For that matter, it is described in [11] that  $w_u$  and  $h_i$  can be obtained by modelling directly the known ratings only, while avoiding over fitting thanks to a regularised model. As a consequence, the system proposed

in [11] will perform a MF by only considering known ratings and the objective function to be minimised is written as

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

where  $\kappa$  is the set of the  $(u, i)$  pairs for which  $r_{ui}$  is known,  $\lambda$  is a parameter controlling the regularisation to avoid over fitting.

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

Problem defined in (6) can be solved with stochastic gradient descent. Further details on how to implement it are provided in [11].

## 2.3 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  $u$  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 three rows (the last row is corresponding to the current week) and seven 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  $R$  is therefore written as

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

where “?” is the unknown electrical load.

Actually, matrix in (7) has a similar structure to matrix in (4) which characterises 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  $R$  in (7) can be found with the factorisation of  $R$ , following (6).

Finally, we provide an illustration of a specific behaviour when we predict electrical quantities with MF. If we consider the matrix in (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 characterised 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

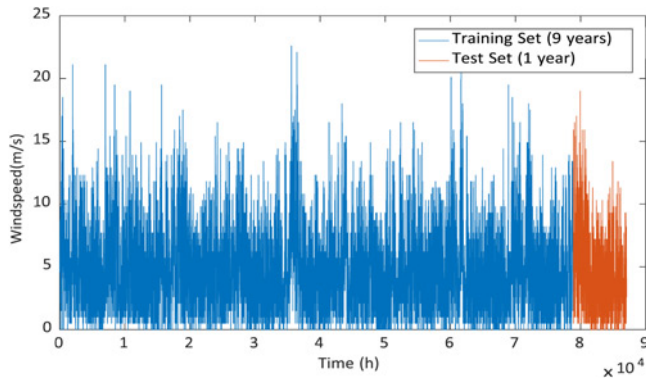
$$R = \begin{pmatrix} r_{u_1 i_2} & r_{u_1 i_3} & \dots & r_{u_1 i_7} & r_{u_2 i_1} \\ r_{u_2 i_3} & r_{u_2 i_4} & \dots & r_{u_2 i_7} & r_{u_3 i_1} \\ r_{u_3 i_4} & r_{u_3 i_5} & \dots & ? & ? \end{pmatrix}, \quad (8)$$

where “?” are the unknown electrical loads.

## 3 Experiments

### 3.1 Data and methods

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



**Fig. 1** Wind speed dataset used for the experiments

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 3652 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 (Fig. 1). The matrix for the training set ( $\mathbf{R}_{\text{training}}$ ) and the matrix for the test set ( $\mathbf{R}_{\text{test}}$ ) 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, (6) is used to factorise 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  $\mathbf{R}_{\text{test}}$ ), we need to create a new matrix by adding the first row of matrix  $\mathbf{R}_{\text{test}}$  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 (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  $\mathbf{R}_{\text{test}}$ .

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 regularisation parameter  $\lambda$  in the optimisation problem in (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

**Table 1** MAE and RMSE values for the persistence model

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

**Table 2** MAE and RMSE values for the ARIMA model

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

**Table 3** MAE and RMSE values for our approach

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

manually defined. The optimal values were chosen as the ones providing the lowest MAE resulting from the experiments.

## 4 Results

In this section, we evaluate the behaviour of our prediction approach based on MF regarding two criteria:

i. MAE

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

ii. RMSE

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

In (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:

- (i) Persistence model: it assumes that the future value to be predicted is equal to the last known value.
- (ii) ARIMA model [16]: we used the ARIMA forecast tool in MATLAB to implement this approach.

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

## 5 Conclusion

In this paper, we proposed a novel method based on MF 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 MF 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 KNMI. Results showed that our approach is competitive with ARIMA models. However, an interesting characteristic of MF 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 the precision of prediction. Another interesting property of MF is the possibility to introduce 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 MF as a promising alternative to forecast electrical quantities.

## 6 References

- 1 Hocaoglu, F., Kurban, M.: 'The effect of missing wind speed data on wind power estimation'. *Intell. Data Eng. Autom. Learn. – IDEAL 2007*, 2007, 4881, pp. 107–114
- 2 Ding, N., Benoit, C., Foggia, G., *et al.*: 'Neural network-based model design for short-term load forecast in distribution systems', *IEEE Trans. Power Syst.*, 2016, **31**, (1), pp. 72–81
- 3 Espinoza, M., Joye, C., Belmans, R., *et al.*: 'Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series', *IEEE Trans. Power Syst.*, 2005, **20**, (3), pp. 1622–1630
- 4 Erdem, E., Shi, J.: 'Arma based approaches for forecasting the tuple of wind speed and direction', *Appl. Energy*, 2011, **88**, (4), pp. 1405–1414
- 5 Senjyu, T., Takara, H., Uezato, K., *et al.*: 'One-hour-ahead load forecasting using neural network', *IEEE Trans. Power Syst.*, 2002, **17**, (1), pp. 113–118
- 6 Doucoure, B., Agbossou, K., Cardenas, A.: 'Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind speed data', *Renew. Energy*, 2016, **92**, pp. 202–211
- 7 Garcia, E., Canha, L., Pereira, P.: 'Hybrid forecasting techniques applied to distribution systems for proactive voltage control'. 23rd Int. Conf. on Electricity Distribution (CIRED 2015), 2015, pp. 1–5
- 8 Siripitayanon, P., Chen, H.-C., Jin, K.-R.: 'Estimating missing data of wind speeds using neural network'. Southeast Con, 2002. Proc. IEEE, 2002, pp. 343–348
- 9 Hu, Q., Su, P., Yu, D., *et al.*: 'Pattern-based wind speed prediction based on generalized principal component analysis', *IEEE Trans. Sustain. Energy*, 2014, **5**, (3), pp. 866–874
- 10 Ak, R., Fink, O., Zio, E.: 'Two machine learning approaches for short-term wind speed time-series prediction', *IEEE Trans. Neural Netw. Learn. Syst.*, 2016, **27**, (8), pp. 1734–1747
- 11 Koren, Y., Bell, R., Volinsky, C.: 'Matrix factorization techniques for recommender systems', *Computer*, 2009, **42**, (8), pp. 30–37
- 12 Kim, S., Giannakis, G. B.: 'Load forecasting via low rank plus sparse matrix factorization'. 2013 Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA, USA, November 3–6, 2013, 2013, pp. 1682–1686
- 13 Yu, H.-F., Rao, N., Dhillon, I. S.: 'Temporal regularized matrix factorization for high-dimensional time series prediction', in Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., Garnett, R., (Ed.): 'Advances in neural information processing systems 29' (2016), pp. 847–855
- 14 Meyer, C. D.: 'Matrix analysis and applied linear algebra' (Society for Industrial and Applied Mathematics, 2000)
- 15 Gillis, N.: 'Non-negative matrix factorization: complexity, algorithms and applications'. PhD thesis, Université catholique de Louvain, 2011
- 16 Box, G., Jenkins, G., Reinsel, G.: 'Time series analysis: forecasting and control' (Prentice-Hall, NJ, 1994, 3rd edn.)