

Load Forecast by Multi-Task Learning Models: Designed for a New Collaborative World



Leontina Pinto, Jacques Szczupak, and Robinson Semolini

Abstract This paper proposes a forecasting model designed for lack of data problems based on Multi-Task Learning techniques (MTL). It is especially useful for evolutionary markets and systems, where new paradigms (like renewable penetration or prosumers) significantly impact behavior and dynamics, creating unforeseen responses that would be unpredictable from past (possibly obsolete) historical data. A case study targeting the recent Brazilian load changes illustrates the approach performance: it was possible to combine data from three different distribution companies, creating a learning network, yielding reliable results where all other models failed.

Keywords Load forecast · Lack of data · Multi-tasking learning · Collaborative learning

1 Introduction

Energy demand is perhaps the market's most important pillar: all institutions, agents, and processes—from planning and operation to marketing and management—are essentially organized to serve it. However, although projecting load future evolution is crucial for an economical and secure supply, it is still one of our major challenges. The behavior of the consumer changes continuously, offering unpredictable reactions to various stimuli, as prices, economic indicators, expectations, and perceptions not always based on reality.

Brazilian load offers an interesting case study. The year 2018 experienced an anomalous increase in consumption throughout Brazil, almost always without connection to any of the classical explaining triggers: GDP experienced a sharp fall, as did income and all economic activities' indicators. We currently face a major

L. Pinto (✉) · J. Szczupak
ENGENHO, Rio de Janeiro, RJ 22793-312, Brazil
e-mail: leontina@engenho.com

R. Semolini
ELEKTRO-NEOENERGIA, Campinas, SP 13053-024, Brazil

challenge: consumer behavior has changed, old dynamics no longer represent the present and we *must predict the future without any past basis*. In fact, in this context, the longer the history, the worse is the prediction.

This behavior almost lies within the concepts proposed by [1, 2], where income raise yields a sensible behavior change, breaking the previous classical correlations between consumption and economy indicators.

However, the Brazilian case steps further: even without a significant income raise, popular expectations lead to new apparel acquisitions (specially climatization) and thus to consumption increase. Correlations are broken, and only behavioral economics can explain this anomaly.

It is necessary to develop mathematical models and computational tools as agile as the consumer, able to understand, follow and maybe anticipate its behavior, with the speed of our new times.

2 Objective

This paper describes a model able to accommodate more than just lack of data: we deal with *extreme* scarcity, where forecast needs to be performed from very few observations—for example, one year (twelve months). In this case, historical records are not even enough to allow a backtracking test (identification/prediction): it will be necessary to start from scratch.

It is necessary to “populate” the load history with valid information—and it is important to distinguish information from numbers: it would be possible to create synthetic samples from the available data, but they would contain the same poor information—anything else could even lead us to distorted results.

However, although it is not possible to extract more information from a history beyond the availability limits, it is feasible to *combine* similar experiences: observations from different agents that exhibit similar behaviors. For example, it is possible that distributors in neighboring regions share the same dynamics of consumption. In this case, it might be interesting to “blend the knowledge” of each company into a single richer, more complete history.

This is the proposal of collaborative learning (MTL) [3–5]. By joining forces, information is shared without losing individuality. The model should select the common dynamics and point specificities, leading to a more consistent and reliable projection.

The advantages of the proposed model are highlighted through a comparison between the new model and a Hilbert Space approach, previously used in many Brazilian companies, also designed for lack of data forecast problems.

3 Multi-Task Learning Approach

Considering space limitations, this article summarizes the applied collaborative learning model. More details, including alternative implementations, may be found in [3].

The proposed approach establishes a set of outputs or tasks t (in our case, the target variables, loads, or consumption). Each of these tasks is associated to a set of explanatory variables (inputs) x (in our case, economic, climatic, behavioral activities, etc.). The successful collaborative learning model requires that outputs t react similarly to inputs x .

The function that “maps” the input x to the output t is written as

$$f_t(x) = \sum_{i=1}^d a_{it}u_i(x) : \forall t \in T; a_{it} \in \mathbb{R}; x \in \mathbb{R}^d \tag{1}$$

where

x is the vector of input variables

$f_t(x)$ is the output associated to task t .

function $u_i(x)$ expresses the shared responses of all inputs x and different tasks t .

coefficients a_{it} measure the “coupling” between different tasks.

For the sake of simplicity, this work assumes linear functions (non-linear extensions are possible and relatively straightforward). In this case, function $f(t)$ corresponds to a vector product which may be written as

$$w_t = \sum_{i=1}^d a_{it}u_i \tag{2}$$

and therefore

$$f_t(x) = w_t(x) : \forall t \in T; x \in \mathbb{R}^d \tag{3}$$

where $w_t(x)$ combines the individual task coefficients a to the shared u .

Finally, for concision

$$W = UA : W \in \mathbb{R}^{d \times T} \tag{4}$$

These coefficients are obtained from the historical observations among all agents (even if scarce). Among other methods, the most intuitive is the well-known technique of function fitting to the available history

$$\min \left\{ \sum_{i=1}^m L(y_{it}, \langle a_{it}, U^T x_{it} \rangle) \right\} : a_t \in \mathbb{R}^d \tag{5}$$

where $L(.,.)$ measures the empirical deviation between the model outputs and the available data.

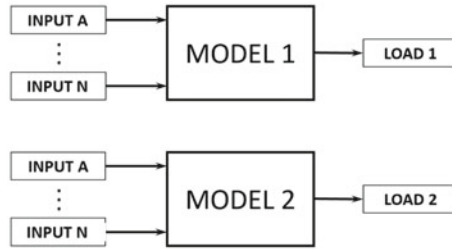


Fig. 1 Classic, individual approach

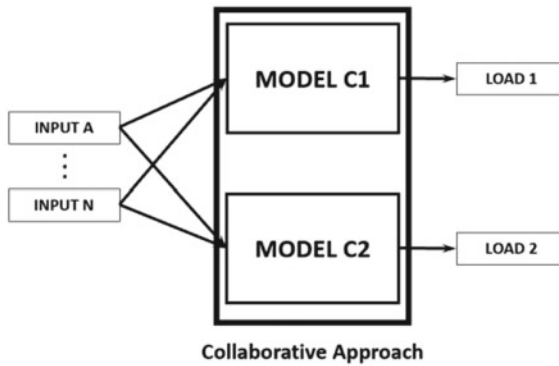


Fig. 2 Collaborative approach

4 Architecture Differences

Figures 1 and 2 illustrate the conceptual difference between the classical and collaborative approaches. While the classical approaches use each set of observations independently, collaborative approach combines all observations, creating a common pattern without losing each agent’s uniqueness.

5 The Classical Hilbert Space Approach

The classical Hilbert approach was previously designed to handle the lack of data, aiming to adapt to the ever-changing Brazilian consumer’s behavior is described in [6, 7] and will be summarized here.

5.1 Projection Theorem

Functional Analysis has been extensively applied to optimization processes [8]. It might be used on a statistical basis, as it is often found in communications, or on a deterministic point of view, the latter usually associated to Hilbert Spaces.

Hilbert Space elements may be seen as vectors, or, in our computerized world, data sequences representing loads, temperatures, economy index, etc. The Hilbert Space is a complete metric space [9], being able to approximate any given vector, always satisfying the Projection Theorem and the Orthogonality Condition [10].

This is shown in Fig. 3, where a given load vector is approximated by the vector sum of three “explaining variable” vectors, V_{e1} , V_{e2} , and V_{e3} (for instance, GDP, income, and temperature).

Figure 4 illustrates the decomposition process for just one “explaining variable”. The original vector is projected (using the Projection Theorem) over the “explaining variable” (say, V_{e1}), yielding the “explained component”. The remaining orthogonal vector corresponds to the unexplained component, or the error vector.

The unexplained component (error) will then be projected over the second explaining vector (say, V_{e2}) and the process will continue until the final error is considered negligible.

Fig. 3 Hilbert space decomposition

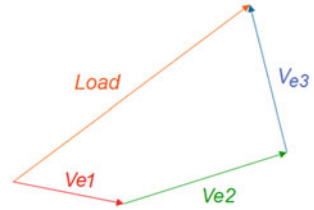
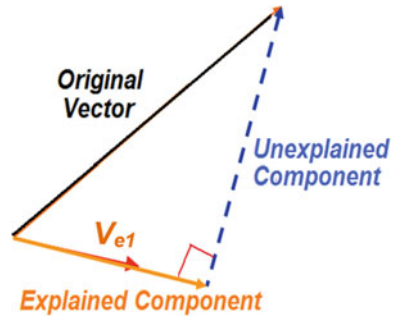


Fig. 4 Original vector decomposition over a first “explaining vector”



5.2 Parallel Processing Implementation

Let C be the desired vector to be decomposed by the set of “explaining variables-vectors” $\underline{S}_1, \underline{S}_2, \dots, \underline{S}_N$. Therefore, one should look for the optimum combination of these “basis” vectors

$$\underline{C} \cong \underline{S} \alpha = [\underline{S}_1, \underline{S}_2, \dots, \underline{S}_N] \alpha \tag{6}$$

such as to minimize the error norm

$$\min_{\alpha} \underbrace{\| \underline{C} - \underline{S} \alpha \|}_{\alpha \| \varepsilon \|} \tag{7}$$

The Projection Theorem states the optimum approximation error is orthogonal to the space of “explaining vectors” and, therefore, to any of its elements, such as

$$\varepsilon^t \underline{S}_i = \underline{C}^t \underline{S}_i - \alpha^t \underline{S}^t \underline{S}_i = 0 \text{ for } i = 1, 2, \dots, N \tag{8}$$

or, for all “explaining vectors”

$$\underline{C}^t \underbrace{[\underline{S}_1, \underline{S}_2, \dots, \underline{S}_N]}_{\underline{S}} = \alpha^t \underbrace{[\underline{S}^t]}_{\underline{S}} \underbrace{[\underline{S}_1, \underline{S}_2, \dots, \underline{S}_N]}_{\underline{S}} \tag{9}$$

leading finally to the unique [9] optimum set of coefficients

$$\alpha = \left(\underline{S}^t \underline{S} \right)^{-1} \underline{S}^t \underline{C} \tag{10}$$

The method is now able to work with large sets of “explaining vectors” in a very efficient way. Moreover, it solves the “co-integration” problem, automatically accommodating inter-correlated explaining variables, finding the best fit while eliminating possible “double counting” effects due to the interdependencies.

Finally, Hilbert Decomposition does not require a large historical period. Although, of course, more reliable information yields a more precise result, it will work at its best within a constrained history, and it suited to a lack of data framework. It has been successfully used in many Brazilian companies, and was able—until now—to yield a reliable forecast based on a mere 5-year history (60 monthly observations).

6 Case Study

6.1 The Challenge

The necessity of a new model, able to deal with lack of data, is shown in Fig. 5. After three years of stagnation, the load finally experienced a steep—and unexpected—rise.

The explanation to this phenomenon, however, was unclear. Figures 6, 7, and 8 show the classical model forecast results for a backtracking process (identification and projection) applied to three neighboring distributors (COELBA, CELPE, COSERN), based on usual explaining variables (GDP, Income, Temperature). There is a sensible, abnormal step associated to 2019 summer in all companies (in fact, all Brazilian distributors exhibited the same behavior, and many different statistical models led to similar results). No available model was able to predict—even to explain this response.

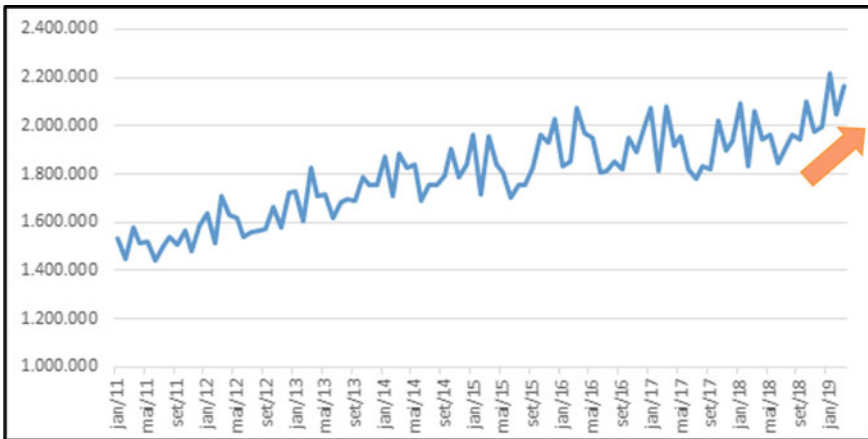


Fig. 5 Bahia (COELBA) load growth

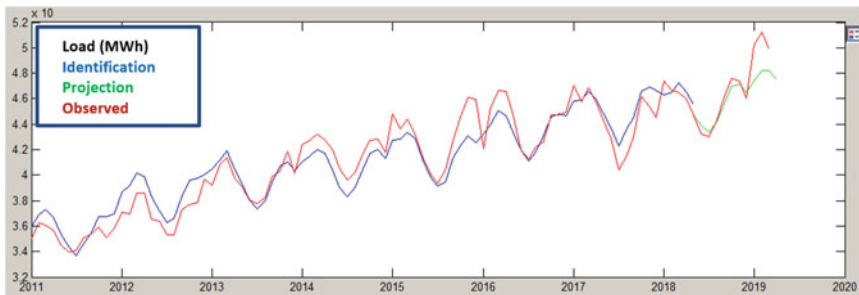


Fig. 6 Bahia (COELBA) load dynamics

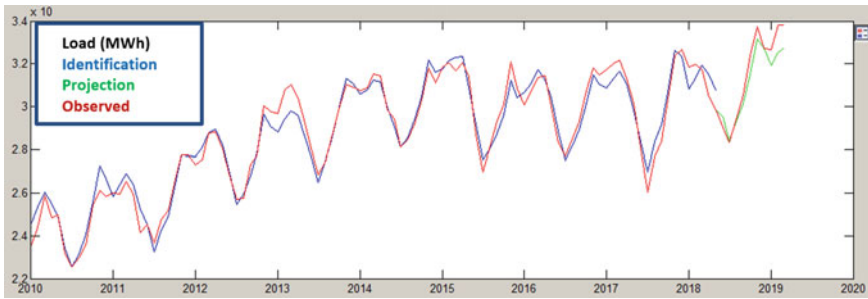


Fig. 7 Pernambuco (CELPE) load dynamics

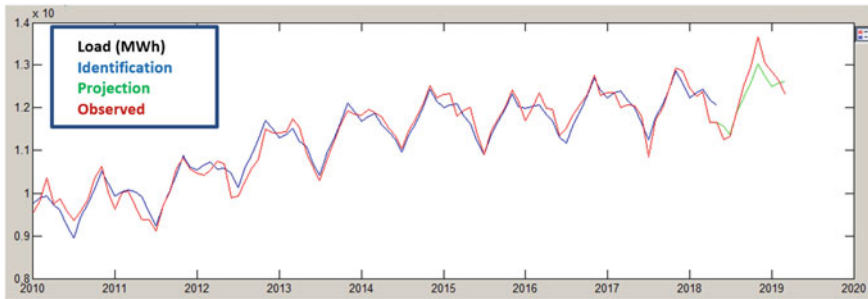


Fig. 8 Rio Grande do Norte (COSERN) load dynamics

More than absorbing the deviations, the main question is should that step be an anomaly, or should it be a change in consumer's behavior—in other words, is this a new permanent pattern? This question is, of course, related to the consumer's reactions and the answer requires a deeper—non-statistical—understanding.

Extensive field research [11], based on behavioral economics [12, 13], uncovered an interesting fact: a disputed election restored the consumer's belief on a stronger economy and a change for the better. This faith in the future, associated to an unusual warm summer, leads to the highest level of refrigeration equipment purchase observed in a decade.

It must be noticed that no economy or income growth backed up this trend: it was a matter of hope and belief. Therefore, no model based on past correlations would be able to account for this change.

As a consequence, consumers possess a new basis of installed demand, and will use it from now on. There is indeed a new standard, which will induce a new response, that must be predicted based on a few observations.

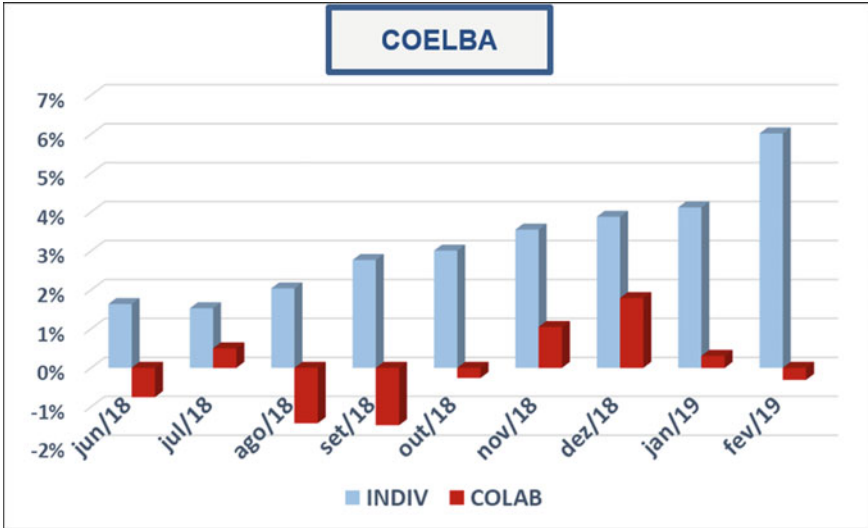


Fig. 9 Individual x collaborative learning, Bahia (COELBA)

6.2 The Proposed Solution

The anomalous behavior was detected from May 2018. It would be very difficult, if not impossible, to apply existing models to as few as 12–18 months for model identification/validation.

We proceeded to try the collaborative learning technique. As our goal was predicting 2019 summer, we based our identification phase on the period from October 2017 to May 2018—where the behavior was still establishing. Of course, more observations will improve the results and will be used as they become available.

Figure 9, 10, and 11 compare the results obtained from our best classical Hilbert Space model (individual learning) and from the collaborative learning. It is interesting to notice that (as expected) the results show slightly higher errors during springtime (as consumers were still adapting, taking decisions, buying equipment). However, projection for summer months is much better.

In any case, the proposed approach offered a clear enhancement on the overall forecast quality. All deviations are significantly lower, despite the almost non-existing information. Moreover, the “deviation trend” is broken, offering a more stable and reliable insight of the future.

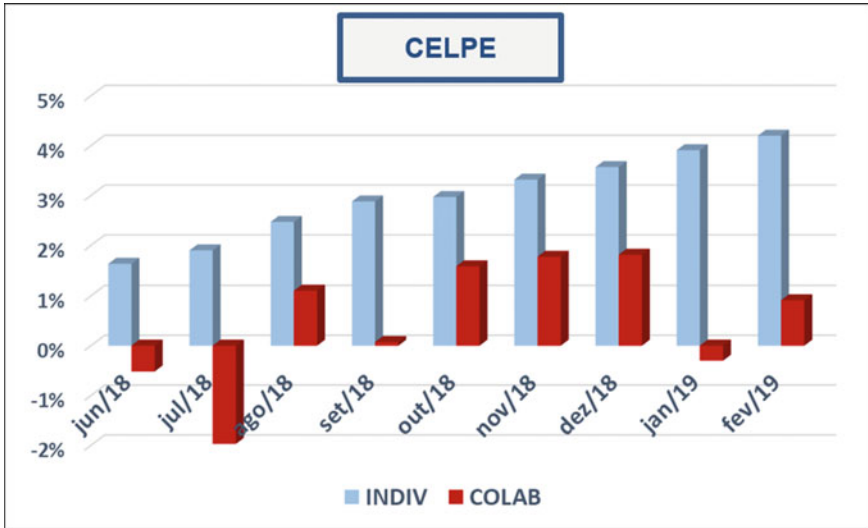


Fig. 10 Individual x collaborative learning, Pernambuco (CELPE)

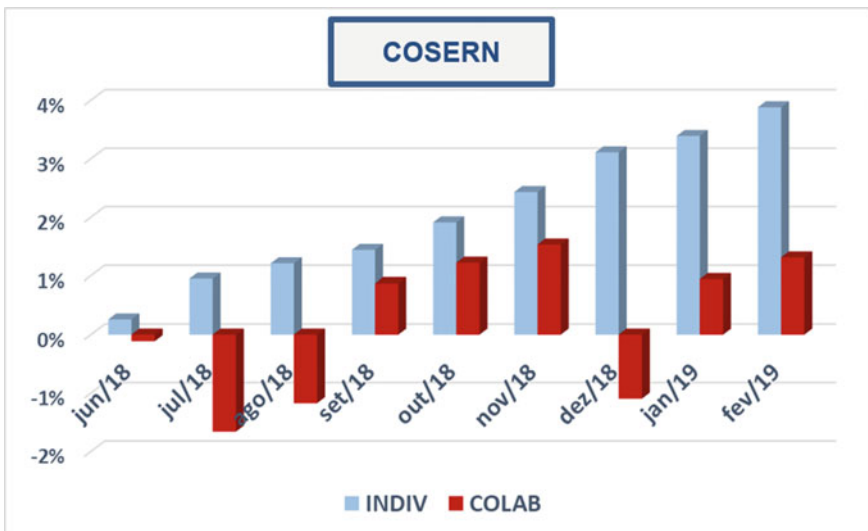


Fig. 11 Individual x collaborative learning, Rio G. do Norte (COSERN)

7 Conclusions

We live in a changing world, and consumption dynamics is not an exception. Preparedness for the future requires the forecast of the unknown. It is crucial to build models that are able to quickly detect modifications—and know the difference from anomalies. It will be necessary to adapt, adjust, absorb novelties.

In the context, classical models, that try to repeat the past, will not be able to foresee the future. The ability to collect and store a huge history may not ensure the quality of information. Number of observations will not necessarily yield precision.

We propose a model designed for this new reality: a collaborative learning technique, able to combine information from different agents, identify common and individual characteristics and build a rich history without traveling back to a distant past.

The described approach was applied to a hard challenge: the projection of the summer load for three Brazilian distributors which broke any known record. A mere 8-month observed data was able to provide much better results for all companies, paving the path to explain the (previously) unexplainable behavior.

These promising results suggest an interesting way, which will be pursued and reported in the near future.

References

1. Fuchs, A., Gertler, P., Shelef, O., Wolfram, C.: The demand for energy-using assets among the world's rising middle classes. *Am. Econ. Rev.* (2016)
2. Auffhammer, M., Wolfram, C.D.: Powering up China: income distributions and residential electricity consumption. (2014)
3. Caruana, R.: Multitask learning. *Mach. Learn.* **28**(1), 41–75 (1997)
4. Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. *Adv. Neural. Inf. Process. Syst.* **19**, 41–48 (2006)
5. Zhang, Y., Yang, Q.: A survey on multi-task learning. arxiv pre-print (2017)
6. Szczupak, J., Pinto, L., Macedo, L.H., Pascon, J., Semolini, R., Inoue, M., Almeida, C., Almeida, F.R.: Load modeling and forecast based on a Hilbert space decomposition. In: 2007 IEEE Power Engineering Society General Meeting, disponível na base de dados do repositório IEEEXPLORE. <https://ieeexplore.ieee.org/document/4275991>
7. Pinto, L., Szczupak, J., Almeida, C., Macedo, L., Inoue, M., Massaro, R., Semolini, R., Pascon, J., Albarelli, E., Tortelli, D.: Load forecast under uncertainty: accounting for the economic crisis impact. In: 2009 IEEE Bucharest PowerTech, pp. 1–5 (2009)
8. Haykin, S.: *Adaptive Filter Theory*, 4th edn, Prentice Hall (2001)
9. Debnath, L., Mikusinski, P.: *Introduction to Hilbert Spaces with Application*. Academic Press (1999)
10. Akhiezer, N.I., Glazman, I.M.: *Theory of Linear Operators in Hilbert Space*. Dover (1988)
11. ENGENHO Brazilian Load Growth Diagnostics, report, available from www.engenho.com
12. Eia, US energy information administration, *Behavioral economics applied to energy demand analysis: a foundation* (2014)
13. Thaler, R.H.: *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company (2016)