A Long-Short-Term Memory Network Model for Biscuit Baking

Alberto Tonda, Nathalie Perrot
UMR 782 GMPA, Team MALICES, INRA
1 av. Lucien Brétignières, 78850, Thiverval-Grignon, France
email: {alberto.tonda,nathalie.perrot}@inra.fr

KEYWORDS

Food process modelling, Neural networks, LSTM, Biscuit baking

ABSTRACT

Long-Short-Term Memory (LSTM) networks are a relatively recent addition to the field of Artificial Neural Networks (ANNs). LSTM networks are specifically tailored for machine learning of time series, where the outputs of a system are not just a function of their inputs, but also of a internal state. The state itself can be seen as dependent on the historical series of all inputs seen by the system up to that point in time. In this paper, we present an application of LSTM networks to the modeling of biscuit baking. Starting from 16 real-world time series of biscuit baking, gathered by the United Biscuits company under different conditions, we show how the proposed LSTM network can correctly predict unseen values. Remarkably, the network is also able to reproduce a dynamic behavior up to variations that might be overlooked as noise.

INTRODUCTION

The process of baking biscuits in industrial ovens involves several biochemical and physical phenomena, including gelatinization of starch, denaturation of proteins, and Maillard reactions. Given this complexity, creating a physically accurate mathematical model of the biscuit baking process seems a daunting task. A possible alternative is to use a data-driven approach, for example a machine learning technique, to derive a black-box model of the whole process from experimental data, to then test its prediction capabilities on unseen data. Such an approach would also ease the difficulty in modeling outputs such as biscuit color, that are traditionally hard to describe mathematically. While most machine learning approaches are unable to deal with time-dependent system, a specific class of Artificial Neural Networks (ANNs), called Long-Short-Term memory (LSTM) networks, currently represent the state of the art for several applications related to time series.

This paper presents a LSTM-based approach to machine learning the biscuit baking process. Starting from a training dataset of real-world time series of biscuit baking, collected by the company United Biscuits, the proposed approach learns the dynamics of two output variables of interest, color and weight loss, and it is then tested on an unseen test dataset.

BACKGROUND

In this section, minimal information about biscuit baking and LSTM Networks are given, in order to introduce the scope of the present work.

Biscuit baking

Industrial biscuit baking aims to transform raw biological materials into a final product which satisfies multiple criteria. For example, thickness and weight of the biscuits can create issues for packaging, if they are not constrained between specific thresholds; on the other hand, the color of the product must be pleasant to the eye of the customer. The transformation process from dough to biscuit is performed in tunnel ovens, and it is the result of complex coupled biochemical and physical phenomena still not completely understood and controlled (Savoye et al. (1992)).

The principal biochemical reactions in the process are gelatinization of starch, denaturation of proteins, and Maillard reactions, that give browned food its distinctive flavor; all these phenomena are linked to temperature, humidity, and water activity inside the biscuit (Wade (1988)). Moreover, conduction, radiation and convection contribute to different degrees to baking, depending on the design of the industrial baking oven. A precise mathematical description of such heat-mass transfers is far from trivial, as the properties of the product change constantly during the process, and little information about the thermal properties of commercial doughs is available. Finally, it would be extremely useful to include the evolution of sensory characteristics of biscuits, such as loss of moisture, formation of color, and change in mass, in the mathematical description of the process; but describing a relationship between such characteristics and the control variables is not straightforward.

Given this complexity, it is not surprising that several approaches have been proposed to model and control the industrial baking process, ranging from fuzzy logic (Per-

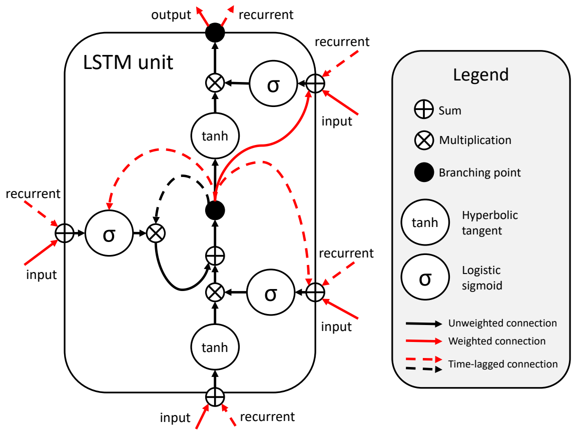

Figure 1: Basic unit of a LSTM network.

rot et al. (2000), Perrot et al. (2006)), to heat-transport models (Sablani et al. (1998), Trystram et al. (1993)), to models tackling air properties in tunnel ovens (Mirade et al. (2004)).

LSTM networks

LSTM networks (Hochreiter and Schmidhuber (1997), Gers et al. (1999)) are a category of ANNs, more specifically belonging to the class of Recurrent Neural Networks (RNNs). Classical ANNs (Rosenblatt (1958)) are machine learning approaches loosely inspired by neural networks in the brain, that can work as general function approximators. ANNs are composed by a series of units called *artificial neurons* connected to each other, able to receive and send signals. Usually, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is calculated by a non-linear function of the weighted sum of its inputs. Like other machine-learning approaches, ANNs can *learn* to approximate an unknown function by tuning the weights in the artificial neurons from a dataset featuring several combinations of inputs and outputs for a target phenomenon, termed *training set*. ANNs are then usually tested on a dataset of unseen values, called *test set*, to verify whether they were able to learn a configuration of weights that generalizes well.

While ANNs are an effective approach, routinely used in applications ranging from games (Silver et al. (2016)) to image classification (Sermanet et al. (2013)), they can only model processes for which the outputs depend exclusively on the current inputs. In several real-world process, however, the outputs are also a function of an internal state, that is itself dependent on the history of inputs until that point. RNNs (Hopfield (1987)) try to address this issue, by adding connections between units to form directed cycles. Thanks to this feedback mechanisms, RNNs can exhibit dynamic temporal behavior, and are used in issues where the sequence of inputs is relevant for the outputs, such as speech recognition or natural language analysis. Among the different types of RNNs, LSTM networks are one of the most recent proposed paradigms. In a LSTM network, each unit is considerably more complex than a simple artificial neuron in an ANN (see Figure 1). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell is responsible for storing values over an arbitrary time interval, while each gate regulates the flow of values that goes through the connections of the LSTM: the input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. Thanks to the ability of storing information over variable intervals of time, LSTM currently represent the state of the art in several domains, such as speech recognition (Xiong et al. (2017)).

DATASET

Sixteen time series of biscuit cooking under different conditions have been gathered by United Biscuits, Inc.[1], in the scope of the DREAM FP7 European Project Axellos (2009-2013). The oven used during the experiments features four different zones, with different temperatures. During the cooking process, biscuits are slowly moved from one zone to the next on metal trays, while the heat flux in the oven is manually regulated by an employee. The considered input variables are: tf (heat flux measured in the top part of the oven, W/m^2), bf (heat flux measured in the bottom part of the oven, W/m^2), z_c (nominal heat flux in the current zone of the oven, W/m^2), and $z_{p1}...z_{p4}$ (nominal heat flux in the previous zones of the oven that the biscuit tray has already passed, W/m^2). The considered output variables are: c (color of the biscuits, based on the reflected light measured in lm), and wl (weight loss of the biscuits, measured in g). Each variable is measured every 5 s, with each baking process lasting 350 s, giving a total of 70 points per time series. Color is always measured on the same individual reference biscuit during the whole time series, weight loss is taken as an average on the same 3 reference biscuits during the experiment. Additionally, the initial conditions of variables c, and wl are used as inputs during the experiments.

Out of the 16 time series, several are repetitions of an experiment under the same conditions (in groups of 3, 3, 2, 3, 2, 3 time series, respectively). Table 1 summarizes the features of the dataset. Figure 2 shows an example of time series, highlighting the non-negligible differences even between repetitions under the same conditions. Another notable feature is that output variable wl presents a behavior that, at a first glance, seems ex-

[1] United Biscuits, `http://www.unitedbiscuits.com/`

tremely noisy.

EXPERIMENTAL RESULTS

The 16 time series are split into a training set (12 time series) and a test set (4 time series), that will be unseen by the LSTM network during the training phase. The test set has been selected among the repetitions of experiments in conditions already present in the training set. All variables have been normalized by subtracting the mean and scaling to unit variance, on the basis of the values contained in the training set. After a few tentative runs, the parameters of the network are configured as follows: 8 inputs (all previously described input variables plus the initial conditions for the 2 output variables), 50 units in a single hidden layer, 2 outputs (all output variables); `tanh` activation function, 3000 training epochs, RMSprop gradient descent optimizer (Hinton et al. (2014)). All the code of the machine learning algorithm is implemented in the Keras (Chollet et al. (2015)) and `scikit-learn` (Pedregosa et al. (2011)) Python libraries.

The final model has excellent fitting on the test set, with mean squared error $MSE = 0.015$ and $R^2 = 0.9863$. An interesting result is that, visually, the model is able to reproduce trends in unseen data that at a first glance might be mistaken for noise: for example, in Figure 3, the model is able to closely predict the behavior of wl, showing that the signal-to-noise ratio is better than what a human expert could have considered from a superficial analysis of the data.

CONCLUSIONS

In this paper, a new data-driven approach to modeling biscuit cooking is presented. Exploiting a class of artificial neural networks known as long-short-term memory networks, the approach is tested on a real-world dataset collected by the company United Biscuits. The proposed methodology is proved capable of accurately predicting the dynamics of biscuit cooking, even for features that are classically hard to predict, such as the color of the biscuits. Future works will focus on testing the approach under different conditions, focusing first on the variation of humidity.

REFERENCES

Axellos M., 2009-2013. *DREAM - Design and development of REAlistic food Models with well-characterised micro- and macro-structure and composition.*

Chollet F. et al., 2015. *Keras.* `https://github.com/fchollet/keras`.

Gers F.A.; Schmidhuber J.; and Cummins F., 1999. *Learning to forget: Continual prediction with LSTM.*

Hinton G. et al., 2014. *Neural Networks for Machine Learning.* `http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`.

Hochreiter S. and Schmidhuber J., 1997. *Long short-term memory. Neural computation*, 9, no. 8, 1735–1780.

Hopfield J.J., 1987. *Neural networks and physical systems with emergent collective computational abilities.* In *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, World Scientific. 411–415.

Mirade P.; Daudin J.; Ducept F.; Trystram G.; and Clement J., 2004. *Characterization and CFD modelling of air temperature and velocity profiles in an industrial biscuit baking tunnel oven. Food research international*, 37, no. 10, 1031–1039.

Pedregosa F.; Varoquaux G.; Gramfort A.; Michel V.; Thirion B.; Grisel O.; Blondel M.; Prettenhofer P.; Weiss R.; Dubourg V.; Vanderplas J.; Passos A.; Cournapeau D.; Brucher M.; Perrot M.; and Duchesnay E., 2011. *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, 12, 2825–2830.

Perrot N.; Ioannou I.; Allais I.; Curt C.; Hossenlopp J.; and Trystram G., 2006. *Fuzzy concepts applied to food product quality control: A review. Fuzzy sets and systems*, 157, no. 9, 1145–1154.

Perrot N.; Trystram G.; Guely F.; Chevrie F.; Schoesetters N.; and Dugre E., 2000. *Feed-back Quality Control in the Baking Industry Using Fuzzy Sets. Journal of food process engineering*, 23, no. 4, 249–279.

Rosenblatt F., 1958. *The perceptron: A probabilistic model for information storage and organization in the brain. Psychological review*, 65, no. 6, 386.

Sablani S.; Marcotte M.; Baik O.; and Castaigne F., 1998. *Modeling of simultaneous heat and water transport in the baking process. LWT-Food Science and Technology*, 31, no. 3, 201–209.

Savoye I.; Trystram G.; Duquenoy A.; Brunet P.; and Marchin F., 1992. *Heat and mass transfer dynamic modelling of an indirect biscuit baking tunnel-oven. Part I: Modelling principles. Journal of food engineering*, 16, no. 3, 173–196.

Sermanet P.; Eigen D.; Zhang X.; Mathieu M.; Fergus R.; and LeCun Y., 2013. *Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:13126229.*

Silver D.; Huang A.; Maddison C.J.; Guez A.; Sifre L.; Van Den Driessche G.; Schrittwieser J.; Antonoglou

Table 1: Summary of the 16 time series on biscuit cooking gathered by United Biscuits. During the experiments, the temperature in different zones of the oven is changed, in order to explore several possible behaviors.

ID	**Training?**	**Heat flux** (W/m^2)				
		z1	**z2**	**z3**	**z4**	**z5**
std-1	yes	2500	3500	4000	4000	2000
std-2	yes	2500	3500	4000	4000	2000
stdval	no	2500	3500	4000	4000	2000
T1-1	yes	4000	3500	4000	4000	2000
T1-2	yes	4000	3500	4000	4000	2000
T1val	no	4000	3500	4000	4000	2000
T2-1	yes	2500	3500	4000	4000	3000
T2-2	yes	2500	3500	4000	4000	3000
T3-1	yes	2500	3500	6000	4000	2000
T3-2	yes	2500	3500	6000	4000	2000
T3val	no	2500	3500	6000	4000	2000
T4-1	yes	2500	3500	4000	6000	2000
T4-2	yes	2500	3500	4000	6000	2000
T5-1	yes	2500	5000	1000	5000	2000
T5-2	yes	2500	5000	1000	5000	2000
T5val	no	2500	5000	1000	5000	2000

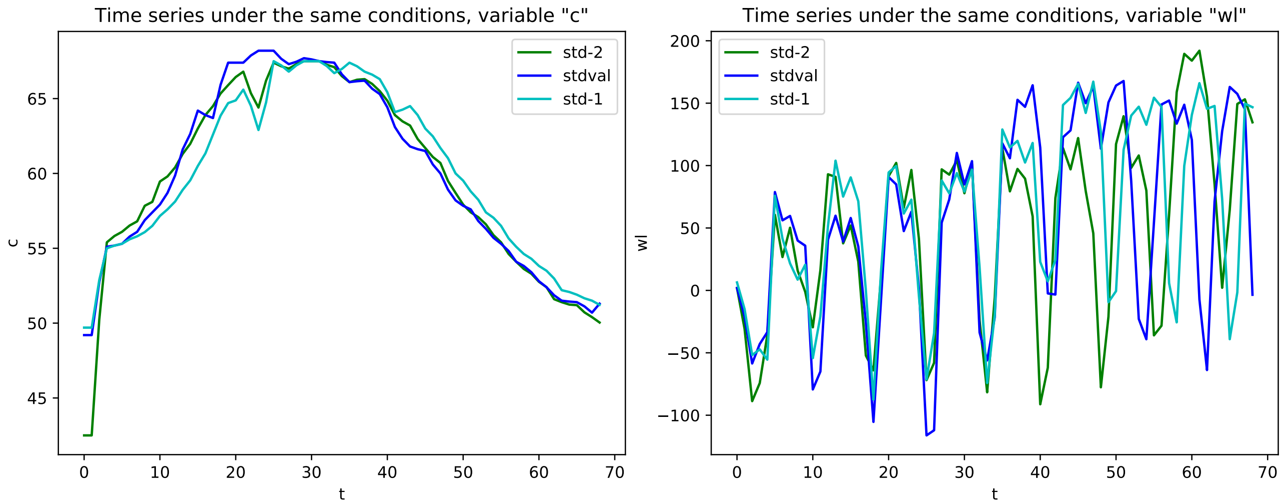

Figure 2: Comparison of three time series run under the same conditions (labeled **std-1**, **std-2**, **stdval**). It is immediately noticeable how the series differ from each other. Interestingly, the behavior of output variable wl seems to indicate the presence of a considerable amount of noise.

I.; Panneershelvam V.; Lanctot M.; et al., 2016. *Mastering the game of Go with deep neural networks and tree search. nature*, 529, no. 7587, 484–489.

Trystram G.; Fahloul D.; Duquenoy A.; and Allache M., 1993. *Dynamic modelling and simulation of the biscuit baking oven process. Computers & Chemical Engineering*, 17, S203–S208. doi: 10.1016/0098-1354(93)80230-k. URL `https://doi.org/10.1016/0098-1354(93)80230-k`.

Wade P., 1988. *Biscuit, cookies and crackers: The principles of the craft. Vol. I.*

Xiong W.; Wu L.; Alleva F.; Droppo J.; Huang X.; and Stolcke A., 2017. *The Microsoft 2017 Conversational Speech Recognition System.* Tech. rep.

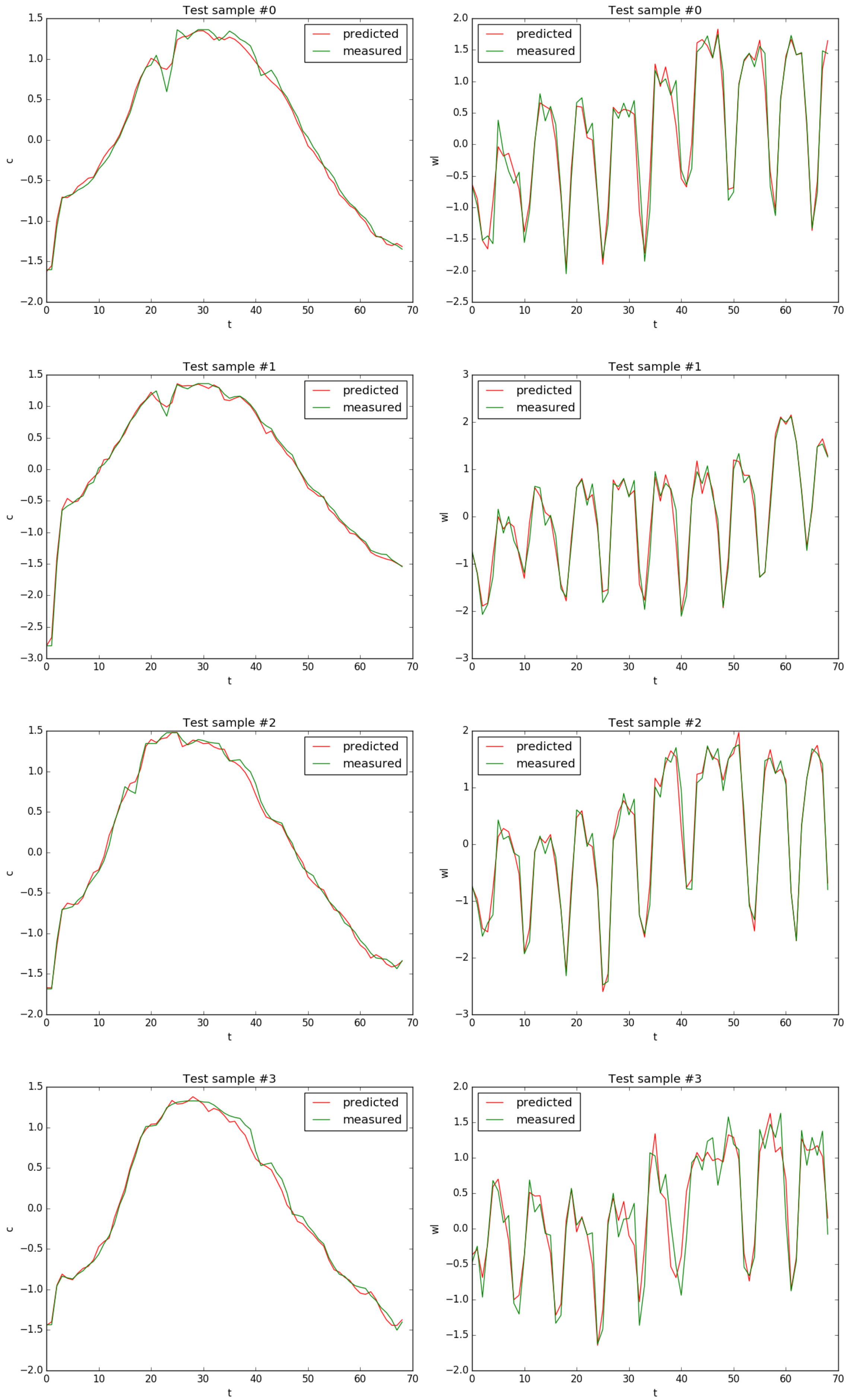

Figure 3: Predictions of the trained model on the unseen test data of the time series labeled **stdval**. The scale is different from the previous plots, as all variables have been normalized.