

# Learning Cognitive-Affective Digital Twins from Social Networks

Andrea NAVARRO <sup>a</sup>, Rafael BERLANGA <sup>a</sup> and Lledó MUSEROS <sup>a</sup>

<sup>a</sup>Universitat Jaume I, Spain

**Abstract.** This paper presents an ongoing project about the implementation of digital twins (DT) for simulating cognitive-affective behaviours in social networks. Our approach relies on a pure data-driven solution, which takes existing public data from social networks to learn cognitive models according to the profile, posts and interactions of the social network users. The final aim is that the learned cognitive models can be parameterised according to existing classifications of traits and emotions so that different behaviours can be eventually simulated with the resulting DTs. In this work, we propose the use of the Transformers neural-network architectures to both interpret incoming messages according to cognitive contexts, and to generate responses to these messages. The first experiments are aimed at integrating and measuring existing approaches for emotion recognition from texts.

**Keywords.** Digital Twin, Social Networks, Emotion and Personality Recognition

## 1. Introduction

Digitization is a common and global trend that companies are investing in after a decade with higher computing capabilities, better communication infrastructures, more sensors and a large amount of available data. It is in this context where Digital Twins (DT) are starting to be deployed in companies. DT [1] can be defined as (physical and/or virtual) machines or computer-based models that are simulating, emulating, mirroring, or twinning the life of a physical entity, which may be an object, a process, a human, or a human-related feature. When the system to be replicated is very complex, learning a model from the data generated by the system is the preferred approach to emulate its behavior. At the moment, DTs are widely used in manufacturing to optimize asset performance, improve process efficiency, and minimize time and costs. Extrapolating smart capabilities from a DT approach to other sectors is challenging, not only from an implementation perspective but also from an ethical and social point of view. Thus, a *Cognitive Digital Twin* can be defined as a digital representation, augmentation, and intelligent companion of its physical twin as a whole, including its subsystems and across all of its life cycles and evolution phases”<sup>1</sup>. [2] focuses on DT for reproducing human cognitive processes in cyber-simulation. They define Cognition DT as a model that monitors, and predicts a persons cognitive status through the processing of different type of information. Finally, on the context of social media, [3] the DT paradigm has been considered to establish a

<sup>1</sup><https://www.linkedin.com/pulse/emergence-cognitive-digital-physical-twins-cdpt-21st-ahmed/>

link among social media data analysis for a virtual product. This research has attempted to use AI tools to categorize the sentiment trends and fill the gap for the relationship between user emotions and product design. Unfortunately, there are no approaches for designing DTs that takes into account both the cognitive and the affective perspectives. This paper presents a first approach towards the definition and implementation of such a cognitive DT, specially focused on exploiting the large amount of social network data by applying new deep learning methods to learn the main DT components.

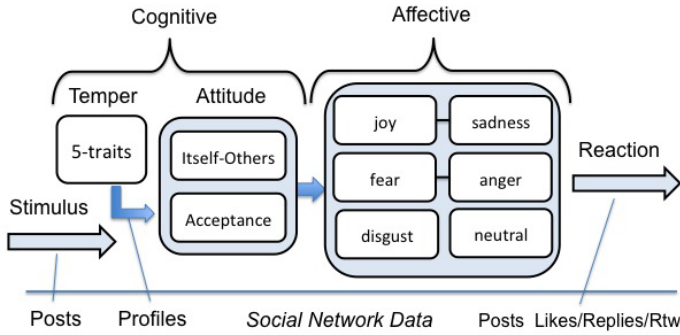
## 2. Methodology

Figure 1 sketches our proposal for designing cognitive-affective DTs. This proposal basically follows the stimulus-organism-response (SOR) scheme, which has been successfully applied to a wide range of fields from psychology to marketing. In our scenario, the scheme is adapted to social network data, as these data will guide the learning process of the different models involved in the digital twin. Thus, the stimulus consists of some message or image whose contents are interpreted according to the *temper* of the twin. This interpretation results in a particular configuration in the *attitude* model, which favours the acceptance of the stimulus as well as the direction this decision will be projected towards. Acceptance-rejection dimension is usually included in most sentiment/emotion models (e.g., polarity). To accept an idea means that it is well aligned with our cognition space, whereas to reject an idea means that it conflicts somehow with our cognition space. The second dimension of the *attitude* model is the direction towards the decision is projected. If the direction is towards oneself, then the emotion will be projected on the oneself behaviour (e.g., feeling shame, pride, etc.) If the direction is towards others, then the emotion will be projected over an object, person, community, and so on (e.g., hate, gratitude, etc.). Once the stimulus is processed in the cognitive component, the *affective* component is activated. In the *affective* component, emotions are selected according to the result of the cognitive results. The model for the *affective* component has been simplified to five major emotions and a neutral class. Emotions are organised according to two excluding class pairs: "joy vs. sadness" and "fear vs. anger". The "disgust" class has no counterpart and is not dependent of the others. From the results of the cognitive and affective components, a final reaction is generated. The reaction is guided by the selected emotions and can consists of a message or an action over the input. For example, actions like retweeting, giving a like or replying a message are possible reactions to the input stimulus.

Let's describe an example to illustrate this process. This starts with a stimulus like a dark humour post in social media. After the combination of this fact with the temper traits, the information will be processed and biased to obtain its attitude, which in this hypothetical case could be located on the low acceptance and outwards projection. That attitude would generate more likely emotions, suppose: sadness and annoyance (low intensity anger). Then, a message is released as reaction: I do not get how they can make fun of suffering, people are so insensitive.

### 2.1. Learning components

The *temper* component of the twin is modelled with the big five variables (i.e., openness, agreeableness, extroversion, conscientiousness, and neuroticism). Temper is a parameter



**Figure 1.** Structure of the cognitive-affective DT.

of the digital twin so that different personalities can be simulated with this component. The learning goal of this component is to associate the big five variables to the way personality transforms the attitude space. For example, a closed-hostile-conscientious personality will likely lead to negative perceptions and an inwards projection. Some work has been carried out for analysing the big five from users activity in social networks [4]. However, this type of experiments require to previously evaluate the personality of the users through proper questionnaires. Moreover, experiments are performed in controlled scenarios so collected data is small. In our proposal, we aim at learning models from existing data in social networks. In this case, the combination of data from users profiles and their messages can be a good indicator for finding some prototypical schemes for temper. Profiling is an active area in social network analysis and its techniques could be extended to this scenario [5].

Transforming the stimulus to the *attitude* space is a challenging problem. Given the representations of both the stimulus and the temper, we need to generate the values for the two attitude dimensions. Regarding the acceptance-rejection dimension, we can find lots of approaches that classify messages polarity, most of them applied to social network analysis. State-of-the-art approaches could be adapted in order to include temper as a condition. Regarding the attitude projection, as far as we know is has not been [trhttps://www.overleaf.com/project/60a271d38920737e80cfd69e](https://www.overleaf.com/project/60a271d38920737e80cfd69e)ated in the literature yet. Transforming *attitude* to emotions has been usually done with hand-written rules. For example, in the hourglass model [6] authors directly map emotions to the *attitude* space. However, some of these mappings are dependent to the temper component and the input stimulus, and cannot be mapped a priori. Thus, it is necessary to learn a model able to associate the cognitive and *affective* components for particular inputs and tempers.

Finally, the generation of the reaction will mainly rely in text generation. It is worth mentioning that mild positive emotions will usually lead to actions like retweet or give a like to the input message. More extreme emotions will likely lead to replies or new messages to express them.

## 2.2. Experiments design

Current state-of-the-art approaches in natural language processing relies on deep learning methods like recurrent neural networks and transformers architectures. Transform-

Dataset	Size	Emotions (missing)	Metrics	Competition/Year
GoEmotion	58K	21 (0)	BERT f1=0,46	Google Research 2020
Affect in Tweets	21K	11 (0)	RNN acc=0.588	SemEval 2018
Huggingface	20K	4 (disgust)	RoBERTa acc=0.93	(not published)

**Table 1.** Large datasets for emotion classification of texts.

ers architectures like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained transformer) have become very popular thanks to their transfer-learning capabilities. This means that an initial model auto-trained with a very large dataset is then fine-tuned to perform different tasks like sentiment analysis, textual similarity and textual entailment. In the context of our proposal, we have explored existing work for detecting and classifying emotions from texts (see Table 1). Currently, we are collecting text samples from different domains (e.g., automotive, medicine, tourism and fashion) in order to train the cognitive-affective components not covered by these datasets.

### 3. Conclusions

In this position paper we describe how cognitive-affective DTs can be structured by means of a series of learning models, which interact to each other to react to stimuli. Preliminary experiments show that a few existing approaches work reasonably well for a very abstract classification of primal emotions, but fail when finer classifications are required. They also show that learning affective models needs to regard the cognitive context. This is because the interpretation of reactions and emotions are dependent on the application domain as well as on cultural aspects. Next steps are focused on building the datasets for training and validating the different components of the DT, mainly those not currently covered by the literature.

**Acknowledgments:** This project has been funded by the Ministry of Economy and Commerce with project contract TIN2016-88835-RET and by the Universitat Jaume I with project contract UJI-B2020-15.

### References

- [1] B. R. Barricelli, E. Casiraghi and D. Fogli: A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* 7: 167653-167671. (2019)
- [2] J. Lu, Z. Xiaochen, K. K. Ali Gharai, D. Kiritsis: Cognitive Twins for Supporting Decision-Makings of Internet of Things Systems. *ArXiv abs/1912.08547* (2019)
- [3] A. A. Olad and O. F. Valilai: Using of Social Media Data Analytics for Applying Digital Twins in Product Development. *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (2020): 319-323
- [4] W. Youyou, M. Kosinski, D. Stillwell: Computer-based personality judgments are more accurate than those made by humans. *PNAS* 112(4): 1036-1040. (2015)
- [5] I. Lanza-Cruz, R. Berlanga, M.J. Aramburu: Modeling Analytical Streams for Social Business Intelligence. *Informatics* 5: 33. (2018)
- [6] Y. Susanto, A. G. Livingstone, B. C. Ng and E. Cambria: The Hourglass Model Revisited. *IEEE Intelligent Systems* 35(5): 96-102. (2020)