Computers in Biology and Medicine 87 (2017) 347–357

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Assessment of temporal predictive models for health care using a formal method☆

Ward van Breda [a,*], Mark Hoogendoorn [a], A.E. Eiben [a], Matthias Berking [b]

[a] VU University Amsterdam, Department of Computer Science, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
[b] Friedrich-Alexander University, Erlangen-Nuremberg, Germany

ARTICLE INFO

Keywords:
Predictive modeling
Mathematical modeling
Regression
Model evaluation
Multi-objective optimization
Health care
Depression

ABSTRACT

Recent developments in the field of sensor devices provide new possibilities to measure a variety of health related aspects in a precise and fine-grained manner. Subsequently, more empirical data will be generated than ever before. While this greatly improves the opportunities for creating accurate predictive models, other types of models besides the more traditional machine learning approaches can provide insights into temporal relationships in the data. Models that express temporal relationships between states in a mathematical manner are examples of such models. However, the evaluation methods traditionally used in the field of predictive modeling are not appropriate for those models, making it difficult to distinguish them in terms of validity. Appropriate assessment methodology is therefore necessary to drive the research of mathematical modeling forward. In this paper we investigate the applicability of such a formalized method. The method takes into account important model aspects, namely descriptive and predictive capability, parameter sensitivity and model complexity. As a case study the method is applied to a mathematical model in the domain of mental health, showing that the method generates useful insights into the behavior of the model.

1. Introduction

Nowadays, more and more measurement devices surround us that are able to record a range of aspects of our health state. Such measurements range from physiological recordings via wrist bands (e.g. heart rate), activity levels recorded through mobile phones to measurements of ones weight via Internet connected scales. This information can be complemented by user input, where the user is prompted regularly to rate physical or mental health aspects. All this data provides a huge richness of highly fine-grained information, which gives ample opportunities for the development of predictive models. Such predictive models for example make use of certain user patterns to predict therapeutic outcomes or to predict disease development over months or years. These models can therefore be drivers for decision making processes, such as therapy selection, or early interventions. While such models are certainly useful, the fine-grained data also paves the road for models that predict the development of health states in terms of hours, or days. These predictive models can be a driver for personalized therapies that support patients on an hourly or daily level.

Although many types of models can be used for this purpose, mathematical models can contribute both to making predictions and gaining insight into the temporal relationships between relevant concepts (see e.g. Ref. [2]). Mathematical models are composed of states represented by continuous values and temporal relationships (dynamics) between states expressed via mathematical equations. In Ref. [3] such models are discussed favourably as they:

1. Are an effective way for theorists to translate verbal hypotheses into precise and unambiguous models.
2. Can be compared in terms of their strengths and their weaknesses, and subsequently their theoretical assumptions and applied implications.
3. Can be used as a common framework to express models, that unite diverse behavioral phenomena.
4. Are a natural way of expressing temporal relationships between different states while still providing insight into the nature of the relationship (opposed to e.g. a recurrent neural network)
5. Can contain parameters that enable tailoring of models (e.g. towards individual patients).

☆ This paper is a significantly extended and improved version of [1].
* Corresponding author.
E-mail address: w.r.j.van.breda@vu.nl (W. van Breda).

http://dx.doi.org/10.1016/j.compbiomed.2017.06.014
Received 22 November 2016; Received in revised form 6 June 2017; Accepted 14 June 2017

A variety of studies have been reported that focus on mathematical modeling for health care, see e.g. Refs. [4–8]. For the reported studies it is unclear how well the proposed models actually perform. Mostly the models output simulated data, that show certain prototypical behavior. Furthermore, no optimization function is used, which would generate the best approximations of the model to describe real life data. How well the models would perform predicting multiple objectives is therefore not known either, as well as how different variants of the models with added complexities perform in comparison, and how the parameters specifically influence the performance of the models. Such an evaluation is far from trivial: typically, the models aim at predicting multiple states (attributes, objectives) over time, and hence the performance on all these states need to be taken into account, and trade-offs between the performance on the different states might be required.

To tackle this problem we propose to use a formal methodology that aims to generate valuable insights in the quality of a model (or multiple model variants) that can be used to the advantage of the modeler, which are often human domain experts. Earlier we proposed the formal methodology in Ref. [1]. In this paper the previous work is extended significantly by adding framework extensions and extended information for each metric. Furthermore, we apply the methodology to a case study to illustrate its applicability.

The methodology is intended to be used for mathematical models that are multi-objective and temporal in nature, and targeted towards the domain of health. Although the method is also applicable in other domains, the domain of health is especially suited because of the mentioned increase of fine-grained data in this field and the importance of understanding the relationship between concepts to deliver better care. The method takes into account specific aspects of the model: its descriptive and predictive capabilities, the contribution of the model parameters, possible collinearity issues between model parameters, and the model complexity. To illustrate the applicability of the method, we evaluate a fine-grained predictive model in the domain of mental health, which we compare with a simpler naive model. The evaluated model is designed to simulate states related to depression, which poses huge challenges worldwide due to its incidence and the huge risk of relapse [9]. Specifically, the model aims to predict the course of the mood and the positivity of thoughts over time.

This paper is organized as follows. In Section 2 we discuss our motivation to propose and explore the applicability of such a method in the context of (1) a number of studies that use mathematical models in the domain of health care, and (2) basic metrics and strategies to measure model performance. Next, we formally describe the assessment method in Section 3. Then, in Section 4 we describe the mood and coping model we choose to evaluate and its domain. In Section 5 the experimental setup is described, in which also the comparison model and the dataset are discussed. In Section 6 the outcome of the model assessments is described, after which we discuss the method and its specific criteria in context of the case study in Section 7.

2. Related work

In this section we present examples of studies where mathematical models are applied in the health domain, that cover important topics to study, but do not apply thorough assessment methodology. Then, we discuss related work that describe different ways to evaluate mathematical models.

2.1. Example studies using mathematical models

Many studies have been conducted that model clinical, cognitive or affective processes using mathematical modeling in the domain of health care. For example, in Ref. [4] mathematical dynamic models are proposed that attempt to describe the (interaction of) mood swings of individuals with bipolar II disorder. In Ref. [5] mathematical modeling is used to model major depressive disorder (MDD), specifically the dynamics of mood level, and certain disease states in major depressive disorder. In Ref. [6] a model is developed that conceptualizes personality as a cognitive-affective processing system using dynamic system modeling. In Ref. [7] a basic model of chronic disease prevention is developed using system dynamic modeling. And in Ref. [8] a discrete-event model is proposed that models risk in complex health care settings. While all are interesting studies, they lack an in-depth analysis of the performance of the models and its parameters.

2.2. Current evaluation methods

The assessment of temporal predictive models usually relates to the extent a model is able to fit empirical data. Note that we focus on mathematical models that aim to solve a regression problem, i.e continuous values. To evaluate such models a metric that measures the difference between the model output and the empirical data for each time point is typically used. Such a metric is often referred to as objective function, or loss-, risk- or cost function. Examples of methods that can be used as objective function are Mean Squared Error (MSE), Root-Mean-Square Error (RMSE), or Mean Absolute Error (MAE). These evaluation metrics are often applied to single-objective optimization, e.g. in Ref. [10] for calculating the descriptive performance of models. If a models focusses on multiple objectives the weighted sum of the various states that are subject to evaluation is often taken into account (see e.g. Ref. [11]).

Alternative performance metrics for multi-objective models exist that separately consider the error score (e.g. the MSE) on the different objectives and the trade-off between them. This is done by means of the Pareto efficiency which gives insight in the interrelated performance per objective. Given the Pareto efficiency, one can chose a certain solution that best fits the problem at hand, such as is done in Ref. [12] where optimization is applied towards two objectives, namely numerical accuracy and order of non-linearity. The evaluation of a Pareto efficiency can be performed using concepts such as attainment surface and hypervolume. For an overview of such evaluation methods see e.g. Ref. [13].

Different interesting model evaluation approaches originate from the domain of genetic programming. In e.g. Ref. [10] genetic programming is used to generate mathematical models with parameters that best describe the empirical data. However, the predictive performance is not part of the fitness function, and therefore the predictive accuracy of the models on unseen data are unknown. Nonetheless, the descriptive performance is an important evaluation criterion and therefore it always should be part of a model assessment methodology.

An example where complexity is considered is in Ref. [12]. The models are evaluated on predictive accuracy and minimal complexity. It is shown that these criteria together are important and result in better performing models compared to the simpler evaluation methods as e.g. in Ref. [10]. For more examples on applying model complexity measures within a genetic programming fitness function, see e.g. Refs. [14,15], or [16]. Next to the obvious necessity to include the predictive capability in a model assessment methodology, we consider model complexity to be of great importance as well. A model complexity criterion helps to promote simple, explainable models, that have equal model performance, and superior application performance, compared to more complex alternatives.

Examples where the parameter sensitivity is taken into account can be found in e.g. Ref. [17], where mathematical models are generated using genetic programming that predict changes in metabolic regulation. Next to optimizing for accuracy, sensitivity analysis was conducted to study the effects of parameter values on the output of the system, i.e. which parameters had high influence. Such analyses can provide interesting information about the dynamics of the model. For more examples of studies using parameter sensitivity measures as model performance measures see Refs. [18,19], or [20]. Such evaluations generate insight in model behavior, and therefore should be part of a model assessment methodology.