Neurocomputing 192 (2016) 14–28

Contents lists available at ScienceDirect

Neurocomputing

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

Can machine learning explain human learning?

Mehrnoosh Vahdat [a,b], Luca Oneto [a,*], Davide Anguita [c], Mathias Funk [b], Matthias Rauterberg [b]

[a] DITEN – University of Genova, Via Opera Pia 11a, I-16145 Genova, Italy
[b] Department of Industrial Design, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[c] DIBRIS – University of Genova, Via Opera Pia 13, I-16145 Genova, Italy

ARTICLE INFO

Article history:
Received 10 July 2015
Received in revised form
9 October 2015
Accepted 17 November 2015
Available online 8 March 2016

Keywords:
Machine learning
Human learning
Rademacher Complexity
Algorithmic stability
Exploratory experiments on students

ABSTRACT

Learning Analytics (LA) has a major interest in exploring and understanding the learning process of humans and, for this purpose, benefits from both Cognitive Science, which studies how humans learn, and Machine Learning, which studies how algorithms learn from data. Usually, Machine Learning is exploited as a tool for analyzing data coming from experimental studies, but it has been recently applied to humans as if they were algorithms that learn from data. One example is the application of Rademacher Complexity, which measures the capacity of a learning machine, to human learning, which led to the formulation of Human Rademacher Complexity (HRC). In this line of research, we propose here a more powerful measure of complexity, the Human Algorithmic Stability (HAS), as a tool to better understand the learning process of humans. The experimental results from three different empirical studies, on more than 600 engineering students from the University of Genoa, showed that HAS (i) can be measured without the assumptions required by HRC, (ii) depends not only on the knowledge domain, as HRC, but also on the complexity of the problem, and (iii) can be exploited for better understanding of the human learning process.

1. Introduction

Since the emergence of Technology-Enhanced Learning (TEL) systems and automatic analysis of educational data, many efforts have been carried out to enhance the learning experience [1,2]. For this reason, Learning Analytics (LA) and Educational Data Mining have recently gained a lot of attention as one of their major interest is to explore the way humans learn [3–5]. New advances in LA enable measuring, collecting and analyzing data about learners and their contexts and allow exploring the behavior of people while learning (e.g. through Machine Learning models), opening the door towards optimized and personalized education [6–9]. LA is a multi-disciplinary field which is tightly connected to Statistics and ML on one side and to Cognitive Science (COGS) and Pedagogy on the other side [10,11].

Machine Learning (ML) is a field of research which develops and studies algorithms that can learn from and make predictions on data [12]. Such algorithms, used as tools in LA, build models from data in order to make data-driven predictions or decisions. ML offers tools for solving many real world problems [13–16]: classification, regression, clustering, online learning, semi-supervised learning, reinforcement learning, etc. [12,17–19]. According to [20] there are three ways of using models of educational processes: the first one includes models used as scientific tools to understand an educational situation, such as using models to predict the student academic success [21]. In the second one, models are used as a component of educational artefacts such as student modeling and its application in a TEL environment [22,23] or integrating the model of student problem-solving into a TEL system with the aim to personalize and adapt educational materials to their needs [24]. Finally, the third one includes models used as basis for design of TEL systems [25].

In addition to proposing new algorithms and tools, ML develops different methods for measuring the effectiveness of a learning process. In particular, ML studies the learning ability of an algorithm in order to avoid data memorization and to improve its generalization performance, which is the ability to learn the targeted concept effectively [26]. Examples of techniques for assessing the performance of a learning algorithm are: Hypothesis Space-based methods [27] (based on the VC-Dimension [26], Rademacher Complexity (RC) [28–30], and PAC Bayes Theory [31,32]) and Algorithm-based methods [33] (based on Compression Bounds [34], and Algorithmic Stability (AS) Theory [35,36]). Thanks to these approaches, many valuable parameters that describe how a particular machine learns can be quantified. For

* Corresponding author.
E-mail addresses: mehrnoosh.vahdat@edu.unige.it (M. Vahdat), luca.oneto@unige.it (L. Oneto), davide.anguita@unige.it (D. Anguita), m.funk@tue.nl (M. Funk), g.w.m.rauterberg@tue.nl (M. Rauterberg).

http://dx.doi.org/10.1016/j.neucom.2015.11.100
0925-2312/

example, it is possible to rigorously measure the generalization performance of a learned model.

While ML studies learning algorithms, COGS studies and analyses how learning takes place in humans [37–39]. In this context, humans can be considered as information processing systems (as suggested in [40]) with a high learning potential and learning is a permanent process that is regulated by optimizing the complexity of the learning context, based on actions and mental schemata of humans [41]. Concept Learning (CL) is the area of COGS that explores how concepts are attained in humans (Human Learning – HL). Various approaches exist in categorising concepts and how they are attained [42,43]. One approach is to consider concepts as mental representations which help to identify and separate objects, events, and relationships. Another approach considers that concepts are learned inductively even from sparse and noisy evidences. In addition, concepts can be formed by combining other simpler concepts, and their meanings are derived from the ones of their constituents. Various theories integrate different approaches of CL: for instance, Exemplar Theory [44] suggests that the categorization takes place by the proximity of the new stimulus to the members of the category that one has observed, and by comparison of similarities, the label is assigned to the stimulus. Another theory, called Prototype Theory [45], explains that categorization takes place in a similar way as Exemplar Theory, while the comparison is carried out to the average of category members not to a specific member. In this case, at first, the attributes of members of a category are derived (named prototypes), then categorization is done by considering the similarity to the generated prototypes. In addition to these theories, researchers discovered that rule-based theories are important in the initial formation of categories [37,46]: first the distinguishing attributes of new items are extracted from the category, then Examplar or Prototype theory, for categorizing distinct items, is applied. In this approach, concepts are constructed by combination [47]. In particular, a concept is represented by some rule that determines whether a stimulus belongs to a category [48]. Thus, humans try to find a rule (learn a model) when being confronted with a new example.

The latest approach towards human rule-based learning has been a motivation for CL to benefit from the research of other fields like Artificial Intelligence, Information Theory, and ML. In this context, the cross between HL and ML [49–51] leads to development of sophisticated formal models of CL [48,52–54]. For instance, [55] measures the ability of humans in Category Learning by applying Bayesian approaches in iterative learning. In this context, a human learns a concept and produces a hypothesis on the given data, then, another human learns the previously developed hypothesis and generates a new one. This method was adopted for identifying the inductive biases in humans. In another study [56], the difficulty of concepts in relation with HL is exploited. In this context, the subjective difficulty of boolean concepts for humans is measured, and it is shown that the subjective difficulty is proportional to the complexity of boolean statements (length of the statement). Thus, by knowing the complexity of the logical structure of concepts, it is possible to predict how difficult that concept is for humans. Other examples are [57,58] where ML Theory, which helps to understand the learning ability of ML algorithms, has been used to explore HL.

Our main contribution is to build a connection between ML and HL. In particular, we apply ML methods to measure the capacity of students to find meaningful rules given various problems. Measuring the ability of a human to capture information rather than simply memorizing can be the key to optimize and improve HL. In this sense, the parallelism with ML is straightforward: for example, several approaches in the last decades dealt with the development of measures to assess the generalization ability of learning algorithms in order to minimize risks of overfitting (memorization). As a consequence, merging ML studies on the generalization ability estimation and HL has been proposed by some researchers. In particular, Zhu et al. [57] propose the application of ML approaches [59] to estimate the human capability of extracting knowledge (Human Rademacher Complexity – HRC). Unfortunately, (H)RC requires a set of models to be aprioristically defined, which includes the models to be explored by the learner (being either an algorithm or a human) [33]. While this hypothesis is not always satisfied by ML methods (e.g. k-Nearest Neighbors [60]), aprioristically defining a list of alternative models for humans is an even tougher task [61,62]. This leads to formulating further assumptions [57], which do not often hold in practice. As an alternative, we propose to exploit AS [35,33] in order to compute the Human Algorithmic Stability (HAS), which does not rely on the definition of a set of models and does not require any additional assumptions. In this study, we comparatively benchmark HRC and HAS, by designing experiments to analyze the way a group of students learns the tasks with different difficulties, and we compare the two approaches to verify which one is the most informative for getting more insights into HL. To reach this purpose, three different experiments were performed from October 2014 till May 2015 with 606 students of various engineering majors from the University of Genoa, Italy. We generated unique questionnaires for every student to measure HRC and HAS over 7 groups of students, as described in the next sections. Filled questionnaires were collected, digitized, anonymized, and analyzed. Our results show that HAS is influenced by the nature and the complexity of the problem to learn. Moreover, contrarily to HRC, HAS is also able to capture the fast-learning ability of a human when dealing with simple problems: this allows providing new perspectives with reference to the human tendency to overfit training data depending on the nature of the problem faced. These results can thus function as a bridge between ML and HL, for the measure of the propensity of the learner towards CL versus simple memorization. This work completes and extends the preliminary results reported in [58].

The paper is structured as follows: Section 2 presents the theoretical ML framework, Section 3 relates the ML framework to HL, Section 4 describes our experimental design, Section 5 reports the results of our study and finally the conclusions of the paper are drawn in Section 6.

2. Rademacher Complexity and algorithmic stability in machine learning

Let us consider the classical binary classification framework [26]. Let $\mathcal{X}$ and $\mathcal{Y} = \{\pm 1\}$ be, respectively, an input and an output space. We consider a set of labeled independent and identically distributed (i.i.d.) data $\mathcal{S}_n : \{Z_1, \dots, Z_n\}$ of size n, where $Z_{i \in \{1,\dots,n\}} = (X_i, Y_i)$, with $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y}$, sampled from an unknown distribution μ over $\mathcal{X} \times \mathcal{Y}$. We also define two modified training sets: $\mathcal{S}_n^{\setminus i}$, where the i-th element is removed and $\mathcal{S}_n^i$, where the i-th element is replaced with Z_i', which is another i.i.d. pattern sampled from μ:

$$\mathcal{S}_n^{\setminus i} : \{Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n\}, \quad \mathcal{S}_n^i : \{Z_1, \dots, Z_{i-1}, Z_i', Z_{i+1}, \dots, Z_n\}. \tag{1}$$

A learning algorithm $\mathcal{A}$ maps $\mathcal{S}_n$ into a function $f : \mathcal{A}_{\mathcal{S}_n}$ from $\mathcal{X}$ to $\mathcal{Y}$. In particular, $\mathcal{A}$ allows designing $f \in \mathcal{F}$ and defining the hypothesis space $\mathcal{F}$, which is generally unknown.

Even if often not specified [35,33], there are some properties that the algorithm $\mathcal{A}$ must satisfy in order to ensure the validity of the results of the next sections. In particular, we consider only deterministic algorithms. It is also assumed that the algorithm $\mathcal{A}$ is symmetric with respect to $\mathcal{S}_n$, i.e. it does not depend on the order of the elements in the training set.