
ON THE PROBLEM OF REPEATED SUPERVISED LEARNING

A PREPRINT

Andrey S. Veprikov
Department of Intelligent Systems
MIPT
Dolgoprudny, Russia
veprikov.as@phystech.edu

Anton S. Kritanlov
HSE University
Moscow, Russia
akhritankov@hse.ru

Alexander P. Afanasyev
IITP
Moscow, Russia

ABSTRACT

In this paper, we explore some aspects of continuous learning artificial intelligence systems as they interact with and influence their environment. For the process of repeated and multiple learning, prediction and updating of the sample, we built a mathematical model and investigated the properties of this process based on the constructed model. We consider this problem in a novel (as far as we know) sense: from the point of view of functional analysis and probability theory. Based on our results we will make several synthetic experiments.

Keywords Machine learning · Continuous machine learning · Repeated learning

1 Introduction

In this paper we consider the problem of repeated supervised learning (многократное машинное обучение), in which the training sample is not fixed, but is updated depending on the predictions of the trained model on the test sample [8, 6, 3]. This problem is widespread in many areas of machine learning and artificial intelligence, but there is still no complete theoretical description of this method of model learning, and we have tried to correct this in this paper. For example repeated supervised learning appears in recommendation systems [6, 14], healthcare [1] and predictive policing [2]. Our main goal was to consider the problem of multiple machine learning from a theoretical point of view, using probability theory, statistics and functional analysis.

The object of our research will be the set \mathbf{R} of distribution density functions

$$\mathbf{R} := \left\{ f : \mathbb{R}^n \rightarrow \mathbb{R}_+ \text{ and } \int_{\mathbb{R}^n} f(x) dx = 1 \right\}, \quad (1)$$

and a mapping D , a feedback loop mapping, that includes training a model, forming predictions on a test sample and mixing predictions into a training sample, as a result of which the distribution of features changes.

A range of problems, that we are considering in this paper are conditions under which D translates \mathbf{R} into \mathbf{R} , conditions when D is a contraction mapping and its fixed point existence conditions.

Main contributions of this paper are as follows. In Section 2 we will make a literature review of our problem (which will not be particularly large, since the problem we are considering is quite new). In Section 3 we will build a mathematical model for the process of multiple learning, prediction and updating of the sample and outline the main questions, the answers to which we will explore in Section 4. In Section 4 we will formulate several theorems (sufficient conditions) for the mapping D , which are novel to the literature (as far as we know). In Section 5, based on the results from Section 4 we will provide some synthetic experiments.

2 Related work

The problem we study is somewhat related to the feedback loops [5, 6] – an observable change in the distribution of input data that occurs over time, because of user interaction with the system. According to Conjecture 1 from [5] the positive feedback loop in a system $D : R \rightarrow R$ exists if D is a contraction mapping, but this conjecture was not proved.

Also our problem is connected with dynamical systems, which are studied in [4, 9], but all these books consider continuous time, and in our work it should be discrete.

Some authors analyzed Markov processes from the point of view of dynamical systems [15, 17], but in Markov chain the future of the process does not depend on the past, but in repeated supervised learning that's not so. Other authors [16, 10] studied stochastic dynamics systems in general.

All theoretical results on probability theory were taken from the book [13], on functional analysis the main sources were [7, 11], the results on the theory of measure and the Lebesgue integral were taken from [12].

3 Problem statement

We consider \mathbf{R} (1) – the set of distribution density functions and a mapping D representing an algorithm for training a model, forming predictions on a test sample and mixing predictions into a training sample, as a result of which the distribution of features changes.

Let's consider an autonomous (time-independent explicitly) discrete dynamical system (there is a dedicated variable - the step number, which increases), at step t and $t + 1$ of which the ratio is fulfilled

$$f_{t+1}(x) = D(f_t)(x), \quad \text{for } \forall x \in \mathbb{R}^n$$

where D becomes an evolution operator on the space of the specified functions f and the initial function $f_0(x)$ is known. Generally speaking, D can be an arbitrary mapping, not necessarily smooth or continuous.

In this paper we consider the following research questions:

1. What are the conditions so that the mapping D becomes a transformation?
2. Under what conditions the mapping D is a contraction mapping?
3. Under what conditions the mapping D has a fixed point?

4 Main results

5 Experiments

References

- [1] George Alexandru Adam, Chun-Hao Kingsley Chang, Benjamin Haibe-Kains, and Anna Goldenberg. Hidden risks of machine learning applied to healthcare: unintended feedback loops between models and future data causing model degradation. In *Machine Learning for Healthcare Conference*, pages 710–731. PMLR, 2020.
- [2] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway feedback loops in predictive policing. In *Conference on fairness, accountability and transparency*, pages 160–171. PMLR, 2018.
- [3] Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 383–390, 2019.
- [4] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*. Number 54. Cambridge university press, 1995.
- [5] Anton Khritankov. Hidden feedback loops in machine learning systems: A simulation model and preliminary results. In *Software Quality: Future Perspectives on Software Engineering Quality: 13th International Conference, SWQD 2021, Vienna, Austria, January 19–21, 2021, Proceedings 13*, pages 54–65. Springer, 2021.

- [6] Anton Khritankov and Anton Pilkevich. Existence conditions for hidden feedback loops in online recommender systems. In *Web Information Systems Engineering—WISE 2021: 22nd International Conference on Web Information Systems Engineering, WISE 2021, Melbourne, VIC, Australia, October 26–29, 2021, Proceedings, Part II 22*, pages 267–274. Springer, 2021.
- [7] Andreĭ Nikolaevich Kolmogorov and Sergeĭ Vasil’evich Fomin. *Elements of the theory of functions and functional analysis*, volume 1. Courier Corporation, 1957.
- [8] Chao Ma, Lei Wu, et al. Machine learning from a continuous viewpoint, i. *Science China Mathematics*, 63(11):2233–2266, 2020.
- [9] Viktor Vladimirovich Nemytskii. *Qualitative theory of differential equations*, volume 2083. Princeton University Press, 2015.
- [10] E Pap, O Hadžić, and R Mesiar. A fixed point theorem in probabilistic metric spaces and an application. *Journal of Mathematical Analysis and Applications*, 202(2):433–449, 1996.
- [11] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.
- [12] W. Rudin. *Real and Complex Analysis*. McGraw-Hill series in higher mathematics. Tata McGraw-Hill, 2006.
- [13] Albert N Shiryaev. *Probability-1*, volume 95. Springer, 2016.
- [14] Ayan Sinha, David F Gleich, and Karthik Ramani. Deconvolving feedback loops in recommender systems. *Advances in neural information processing systems*, 29, 2016.
- [15] Dawid Tarłowski. Global convergence of discrete-time inhomogeneous markov processes from dynamical systems perspective. *Journal of Mathematical Analysis and Applications*, 448(2):1489–1512, 2017.
- [16] Maylis Varvenne. Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory. 2019.
- [17] Anatolii Moiseevich Vershik. What does a typical markov operator look like? *Algebra i Analiz*, 17(5):91–104, 2005.