Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

КАФЕДРА
ТЕЛЕМАТИКА

Семинар по специальности на английском языке

(Workshop in English)

Hybrid artificial intelligence:
Distilling the Knowledge in a Neural Network

(занятие 13)

5 мая
2022 г.

Что обсуждали на прошлой лекции: "propaedeutics" of machine learning

"propaedeutics" - (пропедевтика в медицине заключение о сущности заболевания) это термин, обозначающий вводный курс, систематически изложенный в элементарной (легко объяснимой) форме. В нашем случае это вывод о том, что

- истинность результатов любых формальных методов основана на непротиворечивости используемой аксиоматики. Современная аксиоматика машинного обучения основана на решении обратных статистических задач (задач ретросинтеза), а эти задачи не имеют единственного решения

- Поэтому для достижения практически значимого результата с использованием методов машинного обучения применяются методы оптимизации, в частности, градиентные методы поиска лучших с точки зрения выбранного критерия параметров (весов) нейронных сетей – инструментальных средств поиска решения.

- Однако, у таких методов обучения обнаружено ряд «принципиальных» недостатков (в частности, катастрофическое «затухание» или «взрывном росте» градиента в процессе настройки параметров при «обратном распространении ошибки), поэтому одна из актуальных задач - сделать методы обучения robust, то есть повысить их «устойчивость» к влиянию несущественных факторов. Это можно сделать, если «пополнить» методы машинного обучения новыми средствами, которые учитывают:
  - причинно-следственные связи между входными параметрами (переменными)
  - топологические инварианты обрабатываемых структур данных

  а также
  - субъективный тезаурус системы обучения (например, размерность выходного слоя нейронов используемой ИНС), характеризующий «разрешающую способность» системы. К сожалению, в процессе работы эту точность пока увеличить не удается.

Sources describing current "machine learning" problems

The essence of the problem is that the gradient of the target function in the first layers of the neural network is the product of the gradients of all subsequent layers of the network.

- The case for hybrid artificial intelligence. By **Ben Dickson** -

March 4, 2020

- The Book of Why: Exploring the missing piece of artificial intelligence By **Ben Dickson** -

December 9, 2019

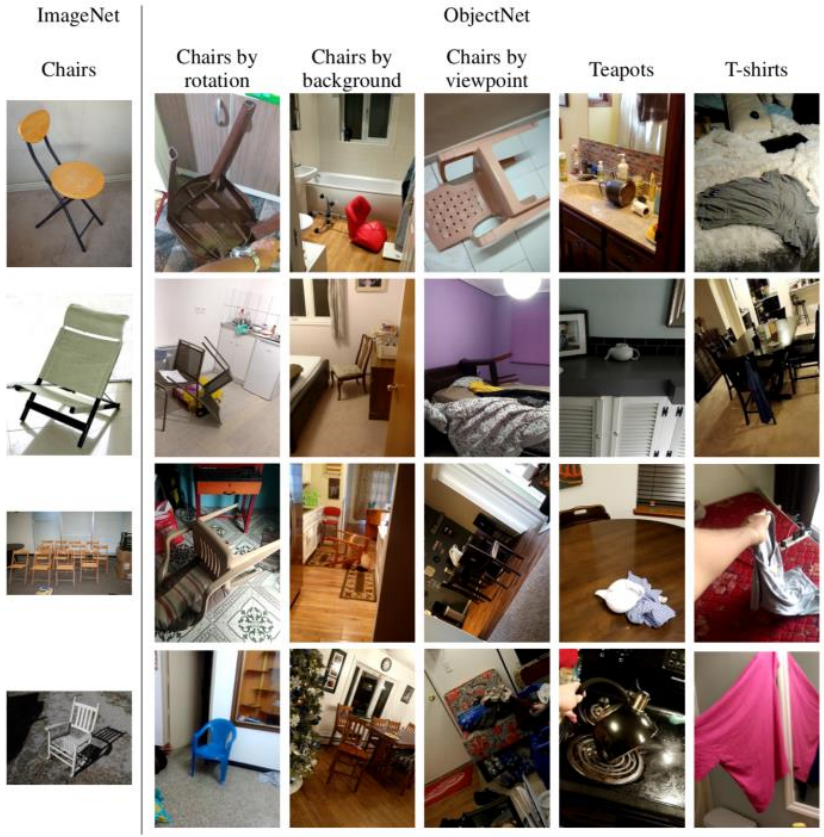
Почему машинному обучению с трудом дается причинно-следственная связь?

MaxRokatansky18 мая 2021 в 18:39

Напоминание: Robust mean - ignores irrelevant factors

- Current machine learning methods seem weak when they are required to generalize beyond the training distribution, which is what is often needed in practice.
- What can we do to take AI to the next level of robust end effectiveness ?
- Understanding – or involve an ability not only to correlate (сопоставлять) and discern (различать) patterns in complex data sets, but also has the capacity to address questions such as
  - what, why, when, and how

Фундаментальная проблема машинного обучения - «проклятие» отождествления



Objects in training datasets compared to objects in the real world can be very different. По мере того, как среда в которой находится объект становится все более сложной, практически невозможно с помощью обучающей выборки охватить все возможные ситуации только посредством добавления большего количества примеров.

How to solve the information paradox of the identification abstraction (in particular Linda's paradox)

- Парадокс Линды - ошибка, возникающая, когда предполагается, что коньюнкция конкретных условий более вероятна, чем одно общее условие.
- Используя абстракцию вероятности, которая приписывается реальных объектам, человека интуитивно может присвоить большую вероятность менее вероятному событию, **например** P(A&B) > P(A) ?, имеем P(A&B)=P(B|A)*P(A), но всегда P(A) <1, следовательно P(A&B)<P(A)….

Однако, - log P(A&B) > -log P(A),

т.е.

информации в событии A&B больше, чем в событии A **I(A&B)>I(A)**

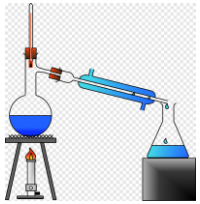
- It is the abstraction of identification that "controls" the transition from the realm of the meaningful, where relations between objective entities are formed by answering the question "why?", to the realm of the abstract, where the question "why?" itself has no meaning (the concept of "abstraction of identification" was introduced by A. A. Markov in 1954).

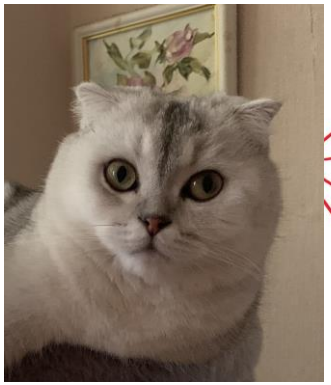
From the meaningful to the abstract - the problem of "knowledge distillation

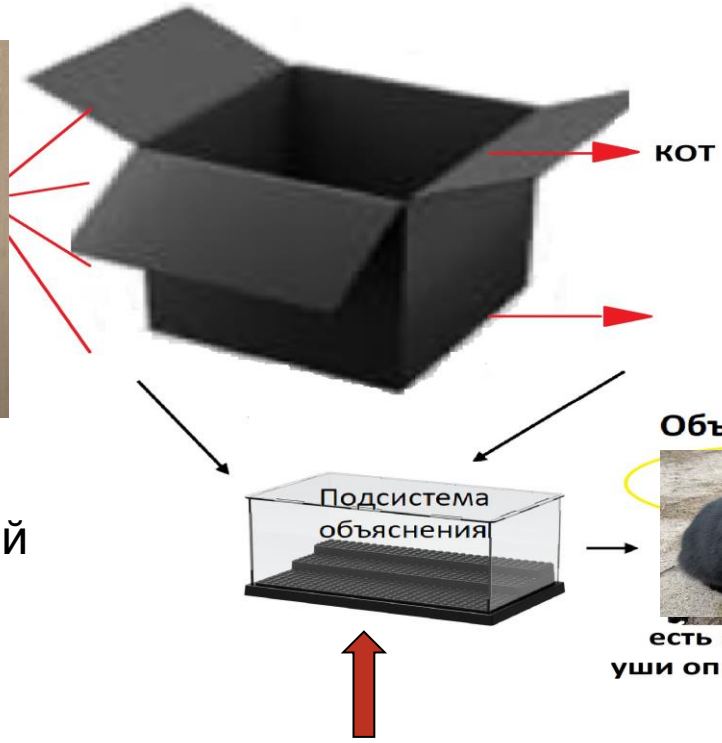
- Одним из наиболее актуальных вопросов в области машинного обучения является интерпретируемость результатов, полученных с помощью моделей на основе нейронных сетей.
- For models based on deep neural networks (DNN), it is quite difficult to get a clear rationale for decisions made with them.
- Для этого требуется алгебраизация зависимостей между признаками входных данных и рассматриваемыми целевыми классами, что может существенно прояснить интерпретацию полученных результатов. .

The Model of Explanation as a Result of the "Distillation of Knowledge

Дистилляция - технологический процесс разделения и рафинирования многокомпонентных веществ







Глубокая ИНС с m >> 100 параметрами

Распознаваемый объект как совокупность 2 миллионами параметров -2Мр пикселей, каждый из которых не характеризует объект

«суррогатная модель», содержащая n – понятий, характеризующих «существенные» свойства «объекта»

How to …. compress the knowledge

The easiest way to improve the explainability of any machine learning algorithm's model is to follow the identity abstraction, that is, to **train many different models** on the same data, and then average their predictions.

However, simple approach does not scale well

So, making predictions using a ensemble of models may be too computationally expensive, especially if the individual models are deep neural nets.

Main question: **is it possible to compress the knowledge that embedded in an ensemble of models, into a single model ?**

Analogies that appear in technology and nature

We know that many insects have a specific and very efficient mechanism that is optimized for extracting energy and information from the environment that are then use in real time mode for different requirements of traveling, reproduction, or other life activities.

In analogy, large-scale machine learning system can extract stealth structure from very large, highly redundant datasets but …. as a rule he such system in most case it does not need to operate in real time and can use a huge amount of computation.

However, the stringent latency and computational resources become important aspects for the practical use of machine learning systems.

That is why we should be willing to train simple models if that makes it easier to extract internal structure from the available data…..

Resume: Knowledge “Distillation” – specific way to transfer knowledge from teacher to student

Knowledge distillation is to train a compact neural network using distilled knowledge extrapolated from a large model or ensemble of models. Using distilled knowledge, we can efficiently train small and compact models without seriously compromising the performance of the compact model.