

Машинное обучение

Материал из Википедии — свободной энциклопедии

Машинное обучение (англ. *machine learning*, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

Различают два типа обучения:

1. Обучение по прецедентам, или индуктивное обучение, основано на выявлении эмпирических закономерностей в данных.
2. Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Дедуктивное обучение принято относить к области экспертных систем, поэтому термины *машинное обучение* и *обучение по прецедентам* можно считать синонимами.

Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с извлечением информации (англ. *information extraction*, *information retrieval*), интеллектуальным анализом данных (data mining).

Содержание

Общая постановка задачи обучения по прецедентам

Способы машинного обучения

Классические задачи, решаемые с помощью машинного обучения

Типы входных данных при обучении

Типы функционалов качества

Практические сферы применения

См. также

Литература

Ссылки

Общая постановка задачи обучения по прецедентам

Имеется множество *объектов* (ситуаций) и множество возможных *ответов* (откликов, реакций). Существует некоторая зависимость между ответами и объектами, но она неизвестна. Известна только конечная совокупность *прецедентов* — пар «объект, ответ», называемая *обучающей выборкой*. На основе этих данных требуется восстановить неявную зависимость, то есть построить алгоритм, способный для любого возможного входного объекта выдать достаточно точный

классифицирующий ответ. Эта зависимость не обязательно выражается аналитически, и здесь нейросети реализуют принцип эмпирически формируемого решения. Важной особенностью при этом является способность обучаемой системы к обобщению, то есть к адекватному отклику на данные, выходящие за пределы имеющейся обучающей выборки. Для измерения точности ответов вводится оценочный *функционал качества*.

Данная постановка является обобщением классических задач аппроксимации функций. В классических задачах аппроксимации объектами являются действительные числа или векторы. В реальных прикладных задачах входные данные об объектах могут быть неполными, неточными, нечисловыми, разнородными. Эти особенности приводят к большому разнообразию методов машинного обучения.

Способы машинного обучения

Раздел машинного обучения, с одной стороны, образовался в результате деления науки о нейросетях на методы обучения сетей и виды топологий их архитектуры, с другой стороны — вобрал в себя методы математической статистики. Указанные ниже способы машинного обучения основаны на применении нейросетей, хотя существуют и другие методы, основанные на обучающей выборке — например, дискриминантный анализ, оперирующий обобщённой дисперсией и ковариацией наблюдаемой статистики, или байесовские классификаторы. Базовые виды нейросетей, такие как перцептрон и многослойный перцептрон (а также их модификации), могут обучаться как с учителем, так и без учителя, с подкреплением и самоорганизацией. Но некоторые нейросети и большинство статистических методов можно отнести только к одному из способов обучения. Поэтому, если нужно классифицировать методы машинного обучения в зависимости от способа обучения, то будет некорректным относить нейросети к определённому виду, правильнее было бы типизировать алгоритмы обучения нейронных сетей.

- Обучение с учителем — для каждого прецедента задаётся пара «ситуация, требуемое решение»:

1. Искусственная нейронная сеть

1. Глубокое обучение
2. Метод коррекции ошибки
3. Метод обратного распространения ошибки
4. Метод опорных векторов

- Обучение без учителя — для каждого прецедента задаётся только «ситуация», требуется сгруппировать объекты в кластеры, используя данные о попарном сходстве объектов, и/или понизить размерность данных:

1. Альфа-система подкрепления
2. Гамма-система подкрепления
3. Метод ближайших соседей

- Обучение с подкреплением — для каждого прецедента имеется пара «ситуация, принятое решение»:

1. Генетический алгоритм.

- Активное обучение — отличается тем, что обучаемый алгоритм имеет возможность самостоятельно назначать следующую исследуемую ситуацию, на которой станет известен верный ответ:

- Обучение с частичным привлечением учителя (англ. *semi-supervised learning*) — для части прецедентов задается пара «ситуация, требуемое решение», а для части — только «ситуация»
- Трансдуктивное обучение — обучение с частичным привлечением учителя, когда прогноз предполагается делать только для прецедентов из тестовой выборки
- Многозадачное обучение (англ. *multi-task learning*) — одновременное обучение группе взаимосвязанных задач, для каждой из которых задаются свои пары «ситуация, требуемое решение»
- Многовариантное обучение (англ. *multiple-instance learning*) — обучение, когда прецеденты могут быть объединены в группы, в каждой из которых для всех прецедентов имеется «ситуация», но только для одного из них (причем, неизвестно какого) имеется пара «ситуация, требуемое решение»
- Бустинг (англ. *boosting* — улучшение) — это процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов.
- Байесовская сеть

Классические задачи, решаемые с помощью машинного обучения

- Классификация, как правило, выполняется с помощью обучения с учителем на этапе собственно обучения.
- Кластеризация, как правило, выполняется с помощью обучения без учителя
- Регрессия, как правило, выполняется с помощью обучения с учителем на этапе тестирования, является частным случаем задач прогнозирования.
- Понижение размерности данных и их визуализация выполняется с помощью обучения без учителя
- Восстановление плотности распределения вероятности по набору данных
- Одноклассовая классификация и выявление новизны
- Построение ранговых зависимостей
- Обнаружение аномалий

Типы входных данных при обучении

- Признаковое описание объектов или матрица объекты-признаки — наиболее распространённый случай. Каждый объект описывается набором признаков.
- Матрица расстояний между объектами. Каждый объект описывается расстояниями до всех остальных объектов обучающей выборки, чаще всего отношениями попарного сходства.
- Временной ряд или сигнал. Последовательность измерений во времени, которое может представляться числом, вектором, а в общем случае - признаковым описанием в данный момент времени.
- Изображение или видеоряд.

Типы функционалов качества

- При обучении с учителем — функционал качества может определяться как средняя ошибка ответов. Предполагается, что искомый алгоритм должен его минимизировать.

Для предотвращения переобучения в минимизируемый функционал качества часто в явном или неявном виде добавляют регуляризатор.

- При обучении без учителя — функционалы качества могут определяться по-разному, например, как отношение средних межкластерных и внутрикластерных расстояний.
- При обучении с подкреплением — функционалы качества определяются физической средой, показывающей качество приспособления агента.

Практические сферы применения

Целью машинного обучения является частичная или полная автоматизация решения сложных профессиональных задач в самых разных областях человеческой деятельности.

Машинное обучение имеет широкий спектр приложений:

- Распознавание речи
- Распознавание жестов
- Распознавание рукописного ввода
- Распознавание образов
- Техническая диагностика
- Медицинская диагностика
- Прогнозирование временных рядов
- Биоинформатика
- Обнаружение мошенничества
- Обнаружение спама
- Категоризация документов
- Биржевой технический анализ
- Финансовый надзор (см. также Финансовые преступления)
- Кредитный скоринг
- Прогнозирование ухода клиентов
- Хемоинформатика
- Обучение ранжированию в информационном поиске

Сфера применений машинного обучения постоянно расширяется. Повсеместная информатизация приводит к накоплению огромных объёмов данных в науке, производстве, бизнесе, транспорте, здравоохранении. Возникающие при этом задачи прогнозирования, управления и принятия решений часто сводятся к обучению по прецедентам. Раньше, когда таких данных не было, эти задачи либо вообще не ставились, либо решались совершенно другими методами.

См. также

- Глубокое обучение
- Квантовое машинное обучение
- Искусственный интеллект

Литература

- Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.

- Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985.
- Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.
- Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.
- Журавлев Ю. И., Рязанов В. В., Сенько О. В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. ISBN 5-7036-0108-8.
- Загоруйко Н. Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999. ISBN 5-86134-060-9.
- Флах П. Машинное обучение. — М.: ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7.
- Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004. ISBN 966-00-0341-2.
- Hastie, T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>). — 2nd ed. — Springer-Verlag, 2009. — 746 p. — ISBN 978-0-387-84857-0.
- Mitchell T. *Machine Learning*. — McGraw-Hill Science/Engineering/Math, 1997. ISBN 0-07-042807-7.
- Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell (1983), *Machine Learning: An Artificial Intelligence Approach*, Tioga Publishing Company, ISBN 0-935382-05-4 (*Machine Learning: An Artificial Intelligence Approach* (<http://books.google.com/books?id=TWzuUd5gsnkC&pg=PA13>) в «Книгах Google»).
- Vapnik V. N. *Statistical learning theory*. — N.Y.: John Wiley & Sons, Inc., 1998. [1] (<http://lib.mexmat.ru/books/9220>)
- Bernhard Schölkopf, Alexander J. Smola *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*. — MIT Press, Cambridge, MA, 2002 ISBN 978-0-262-19475-4 [2] (<http://www.learning-with-kernels.org/>)
- I. H. Witten, E. Frank *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. — Morgan Kaufmann, 2005 ISBN 0-12-088407-0 [3] (<http://www.cs.waikato.ac.nz/~ml/weka/book.html>)
- Liang Wang, Li Cheng, Guoying Zhao. *Machine Learning for Human Motion Analysis*. — IGI Global, 2009. — 318 p. — ISBN 978-1-60566-900-7.

Ссылки

- Weka: Data Mining Software in Java (<http://www.cs.waikato.ac.nz/~ml/weka/>)
- [www.MachineLearning.ru](http://machinelearning.ru) (<http://machinelearning.ru>) — профессиональный вики-ресурс, посвященный машинному обучению и интеллектуальному анализу данных
- ММРО — Математические методы распознавания образов (<http://mmro.ru>)
- Константин Воронцов. Курс лекций Математические методы обучения по прецедентам (<http://www.ccas.ru/voron/teaching.html>), МФТИ, 2004—2008
- Константин Воронцов. Курс «машинное обучение» (http://shad.yandex.ru/lectures/machine_learning.xml) школы анализа данных компании Яндекс.
- Игорь Кураленок. Курс «машинное обучение» (<https://www.lektorium.tv/course/22975>) Лекториум.
- Роман Шамин. Курс «Машинное обучение и искусственный интеллект в математике и приложениях» (<http://www.mathnet.ru/conf1243>). НОЦ Математического института им. В. А. Стеклова РАН

Источник — https://ru.wikipedia.org/w/index.php?title=Машинное_обучение&oldid=114150494

Эта страница в последний раз была отредактирована 12 мая 2021 в 05:28.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.