

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

Лизунова Елизавета Дмитриевна, студентка 1 курса бакалавриата института экономики и управления АПК, ФГБОУ ВО РГАУ – МСХА имени К. А. Тимирязева, lizunova_ed@bk.ru

Научный руководитель – Храмов Дмитрий Эдуардович, ассистент кафедры статистики и кибернетики, ФГБОУ ВО РГАУ – МСХА имени К. А. Тимирязева, khramovde@rgau-msha.ru

Аннотация. В статье представлена краткая информация об алгоритмах машинного обучения и их типах. Представлена информация об онтологии как подходе к созданию преимущественно исследовательских или узко направленных продуктов и подсистем.

Ключевые слова: алгоритмы машинного обучения, онтология, оптимизация анализа данных.

MACHINE LEARNING ALGORITHMS

Lizunova Elizaveta Dmitrievna, 1st year undergraduate student of the Institute of Economics and Management of the Agroindustrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, lizunova_ed@bk.ru

Scientific supervisor – Dmitry Eduardovich Khramov, Assistant at the Department of Statistics and Cybernetics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, khramovde@rgau-msha.ru

Annotation. The article provides brief information about machine learning algorithms, their concepts and categories. Information is provided on ontology as an approach to the creation of predominantly research or narrowly focused products and subsystems.

Key words: machine learning algorithms, ontology, optimization of data analysis.

Алгоритмы машинного обучения являются основными инструментами в области анализа данных. Однако извлечение знаний из наборов данных, полученных в реальной жизни, требуют сложной обработки. Получение доступных точных наборов данных и выбор подходящего алгоритма анализа являются важными задачами для аналитиков данных.

Различают всего четыре вида алгоритмов машинного обучения: контролируемое обучение, обучение без учителя, обучение с подкреплением.

Рассмотрим каждый из них более подробно, выделив отличительные характеристики.

Контролируемое обучение подразумевает обучение модели на маркированных данных, где ожидаемый результат уже известен. Например, вы сможете обучить контролируемую модель машинного обучения, которая имеет исторически помеченный набор данных, и использовать его для прогнозирования значения целевой переменной. Например, если задача предсказать цену дома на основе количества спален, площади, местности и т. д., то необходимо будет обучить модель контролируемого обучения для прогнозирования цены на дом [1]. Модели контролируемого обучения могут быть использованы даже если у вас небольшое количество учебных данных, потому что это все равно приведет «достаточно хороший» результат, который можно будет использовать.

Алгоритмы контролируемого и неконтролируемого машинного обучения имеют разные области применения и имеют свои сильные и слабые стороны. Наиболее распространенным различием между этими двумя алгоритмами является способ их обучения. Неконтролируемая модель обучения не имеет целевой переменной, она обычно используется для получения информации из данных. конечная цель в модели обучения без контроля заключается в том, чтобы извлечь некоторые идеи или закономерности из данных.

Обучение с подкреплением — метод, который именуется также «средним подходом», может уменьшить негативные последствия как контролируемого, так и неконтролируемого обучения, но он не будет столь эффективным, как оба метода с точки зрения их сильных сторон. Этот метод довольно эффективен, потому что маркировка данных — это трудоемкая операция, и большинство реальных данных не маркируются. В обучении с подкреплением система может обрабатывать частично помеченные данные, что является преимуществом. Например, сочетание логистической регрессии (контролируемого обучения) и техники кластеризации k-сред (неконтролируемое обучение) может составлять обучение с подкреплением [2]

На втором этапе процесса машинного обучения (т.е. при разработке гипотез) разработка алгоритма на основе онтологий способствует включению онтологических знаний в процесс проектирования и разработки алгоритмов машинного обучения. Деревья решений, как одни из алгоритмов машинного обучения, основанные на онтологиях, относятся к алгоритмическому подходу проектирования, который включает онтологические принципы в построении и использовании деревьев решений [3].

Методы алгоритмизации машинного обучения используются для создания прогностических моделей на основе знаний о конкретной области. Знания могут быть использованы для ускорения поиска оптимальных решений и ограничения поиска. Системы искусственного интеллекта используют знания, закодированные в онтологиях и базах данных, для улучшения своей производительности.

Онтология же в свою очередь – модель, которая описывает концепции и взаимосвязи в определенной, конкретной области. Эта система значительно упрощает восприятие больших данных и систематизирует максимально доступно для понимания. Онтология позволяет алгоритмам машинного обучения работать с более сложным контекстом, используя знания о взаимосвязях между концепциями. Это значительно расширяет возможности в таких областях, как обработка естественного языка, семантический поиск и аналитика больших данных.

В статье были рассмотрены три основных вида алгоритмов машинного обучения, их принцип, преимущества. Также онтология была представлена не как отдельная область машинного обучения, а в роли инструмента, систематизирующей работу и функционал алгоритмов. С развитием алгоритмов машинного обучения онтология также становится незаменимым компонентом систем искусственного интеллекта.

Библиографический список

1. Nathan Rosidi, Machine Learning Algorithms You Should Know for Data Science — URL: <https://www.stratascratch.com/blog/machine-learning-algorithms-you-should-know-for-data-science/>
2. Ghidalia S., Labbani Narsis O., Bertaux A., Nicolle C., COMBINING MACHINE LEARNING AND ONTOLOGY: A SYSTEMATIC LITERATURE REVIEW — URL: <https://arxiv.org/pdf/2401.07744>
3. Tianxing M., Zhukova N./ An Ontology of Machine Learning Algorithms for Human Activity Data Processing — URL: <https://fruct.org/publications/volume-22/acm22/files/Tia.pdf?ysclid=m2wfkro319281441124>
4. Korel L./ Machine learning with ontologies — URL: https://cw.fel.cvut.cz/b211/_media/courses/xp36rgm/machine_learning_with_ontologies.pdf
5. Храмов, Д. Э. Онтология алгоритмов оценки продолжительности жизненного цикла программного обеспечения / Д. Э. Храмов // Проблемы управления в социально-экономических и технических системах : Материалы XX Международной научно-практической конференции. Сборник научных статей, Саратов, 17–18 апреля 2024 года. – Саратов: Издательский центр "Наука", 2024. – С. 137-140. – EDN RQHOF7.