

ОГЛАВЛЕНИЕ

Введение.....	4
1. Теоретические основы машинного обучения.....	6
2. Искусственный интеллект в цифровой экономике.....	26
3. Цифровая трансформация промышленных предприятий в условиях инновационной экономики.....	33
4. Оценка конкурентоспособности организаций в условиях цифровой экономики.....	49
5. Технологии искусственного интеллекта как фактор цифровизации экономики России и мира.....	55
6. Цифровая экономика и ее влияние на конкурентоспособность организаций.....	62
7. Использование методов и алгоритмов машинного обучения для исследования конкурентоспособности организаций.....	82
8. Риски при проведении цифровой трансформации организаций на примере технологий машинного обучения.....	87
Заключение.....	96
Список литературы.....	100

ВВЕДЕНИЕ

Направление по обработке «больших данных» входит в зону ответственности ГК «Росатом», где в рамках создания цифровой платформы «Распределенная среда обработки больших данных» предусмотрены работы по развитию компонентов систем поддержки принятия решений, программных интерфейсов доступа к данным и др. Ввиду возрастания объема данных и необходимости осуществлять поиск среди них на первый план выходит задача определения релевантности и пертинентности информации. Понятие «релевантность» означает «соответствие полученной информации информационному запросу».

Таким образом, релевантность определяется исключительно используемыми математическими моделями в конкретной информационно-поисковой системе.

Под понятием «пертинентность» понимается «соответствие полученной информации информационной потребности», т.е. пертинентность — соответствие найденных информационно-поисковой системой документов информационным потребностям пользователя независимо от того, как полно и точно эта потребность выражена в форме запроса.

Один из подходов для повышения пертинентности информационного поиска — использование методов машинного обучения как частный случай применения интеллектуального анализа данных.

Данное направление исследования лежит на пересечении целого ряда дисциплин: математики, информатики, статистики, теории вероятностей и прочих. Выделяют особый подкласс информационных систем, опирающийся на математические модели и позволяющий решать задачи по определению релевантности и пертинентности данных, который называется рекомендательными системами. Отсутствие универсальных подходов для решения задач предобработки и классификации в рекомендательных системах делает разработку методов и алгоритмов машинного обучения в этом случае особо актуальной.

Рекомендательная система представляет собой программное средство и методы, назначением которых является прогнозирование поведения пользователя в отношении объекта информационного поиска и

формирование рекомендаций для объектов, с которыми он еще не встречался. Потребность в персонификации информационных предложений также будет реализовываться в развитие интеллектуальных советников, которые будут адаптироваться под конкретного пользователя. Источником данных для рекомендательной системы выступают профили пользователей, явные и неявные.

В профиле фиксируются данные о самом пользователе, предоставленные как лично, так и через перечень действий, которые он совершил за время работы с системой.

1. Теоретические основы машинного обучения

Несмотря на тот факт, что непосредственно термин машинное обучение был введен только в 1959 году, первые попытки построения таких систем начаты еще в 1936 году с разработок AT&T Bell Labs синтезатора речи. Эпоха же нейронных сетей отсчитывается с работы Мак-Каллока и Питтса 1943 года и введенной одноименной пороговой модели биологического нейрона с функцией Хевисайда в качестве функции активации

$$H(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Такая модель является грубой, как с точки зрения описания биологического нейрона, так и ограниченной в практическом применении.

В связи с разрывностью этой функции, служившей затруднением для настройки нейросетей, построенных на такой модели нейрона, в расчетную практику был введен целый набор непрерывных функций активации, в частности, кусочно-линейная, гауссова, сигмоидальная, тангенциальная и т.д (см. рисунки 1,2,3,4)

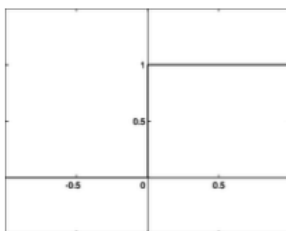


Рисунок 1 — Функция Хевисайда

ственно позволит повысить уровень сложности создаваемых алгоритмов, определяемых данными (Data-Driven подхода).

Моделирование на основе такого подхода представляется, как развитие технологии, включающей ряд взаимосвязанных этапов, а именно:

1. Постановка задачи, определение объекта моделирования;
2. Формализация задачи, формирование перечня признаков описания объекта и явления;
3. Кодирование признаков объектов;
4. Выбор нейросетевой топологии, разработка концептуальной модели, выявление её основных элементов;
5. Переход к математической модели; создание алгоритма и написание программы;
6. Планирование и проведение компьютерных экспериментов;
7. Выбор способа оценки результата, анализ и интерпретация результатов;
8. Уточнение модели при необходимости, переход к 4; 9.

Получение результата моделирования.

Схема процесса моделирования и его этапы представлены на рисунке 7.

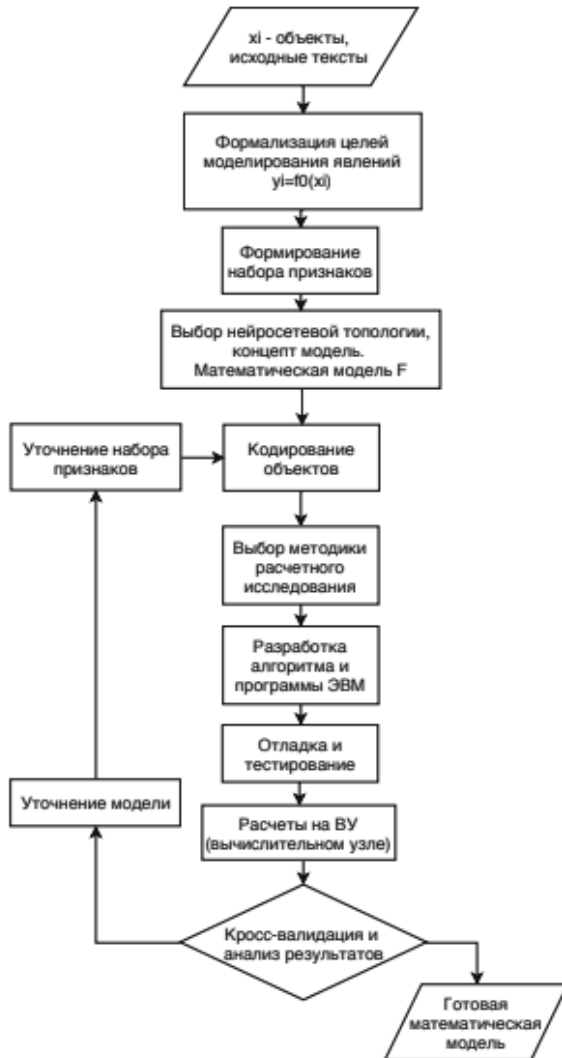


Рисунок 7 — Этапы процесса моделирования

Со времен появления первой вычислительно полезной линейно разделяющей однослойной нейронной сети на формальных математических нейронах Мак-Каллока и Питца, а именно персептрона, предложенной в 1958 году Фрэнком Розенблаттом, доказавшим теорему сходимости для обучения персептрона, парк методов машинного обучения

бурно рос быстрыми темпами. В настоящее время для решения задач с обучением на данных используются как традиционные нейронные сети коннекционистского периода (до 2000-х годов), так и более современные нейронные сети глубокого обучения.

Несмотря на тот факт, что эти методы сложились ещё до формирования концепции глубокого обучения, на практике для небольших обучающих выборок они являются базовым высокоэффективным решением, зачастую и оптимальным. Сказанное делает чрезвычайно полезной первичную обработку построения метода, основанного на данных, на этих алгоритмах с оценкой исходной точности (baseline), и в ряде случаев эта оценка становится окончательной. Ниже рассматриваются основные методы, используемые для анализа данных.

Унитарное кодирование. Унитарное (One-hot) кодирование применяется для дискретных признаков. Каждый признак кодируется двоичным вектором фиксированной длины. Длина вектора равна количеству всевозможных значений признака. Вектор содержит нули (при прямом унитарном кодировании, и единицы при обратном) во всех элементах, кроме одного, чей индекс соответствует номеру текущего значения признака в словаре его всевозможных значений.

Кодирование на основе статистической важности слов. TF-IDF (англ. Term Frequency — Inverse Document Frequency): статистическая мера отражающая некоторую важность токена для документа находящегося в некоторой коллекции (корпусе). Вектор частот токенов, характеризующий документ d , получается путем применения формулы TF-IDF для каждого токена t в коллекции документов D .

Метод опорных векторов. Машина опорных векторов (англ. Support Vector Machine, SVM) является весьма часто используемым алгоритмом для целого набора практических задач.

Метод опорных векторов является методом машинного обучения с учителем, основанным на расположении гиперплоскости в пространстве признаков наилучшим образом для разделения примеров разных классов из тренировочной выборки. Положение гиперплоскости определяется так, чтобы отдельные классы были на максимальном расстоянии с разных сторон от разделительной плоскости. Для этого с разных сторон от разделяющей плоскости строятся две параллельные гиперплоскости.

6. Цифровая экономика и ее влияние на конкурентоспособность организаций

Современные требования общества, с которыми сталкиваются компании определили необходимость искать новые средства взаимодействия, в том числе и управлении корпоративными финансами. Безусловно, что с внедрением цифровых технологий сформировались новые требования к управлению процессами в компаниях.

Четвертая промышленная революция выступает новой эрой развития общества, где стираются грани между физическими, цифровыми и биологическими технологиями.

Цифровые технологии разрушают ограничения пространства и времени, изощренные программные средства объединяют покупателей и производителей точнее, быстрее и проще, чем раньше, а бизнес-платформы, использующие технологии объединения организаций, ресурсов и населения в интерактивной экосистеме, позволяют создавать и распространять поразительный объём ценности для пользователей.

Высокий интерес к такому явлению как «цифровизация» со стороны политиков, бизнесменов, журналистов, обуславливает появление в средствах массовой информации и научных трудах большого количества синонимов данного термина. Наиболее популярными являются «цифровая экономика», «цифровизация» и «цифровая трансформация», которые по своей сути имеют разную природу. Европейская часть научного сообщества и научные труды российских экономистов в большей степени используют термин «цифровая экономика».

2018-й, образно говоря, можно назвать годом, когда в стране поднялся вопрос о смыслах цифровой экономики, о приоритетных направлениях, о том, как нужно подходить к решению приоритетных проблем через цифровые технологии, платформенные решения, и посвятили этому, по сути, целый год. При этом нужно отметить, что в 2018 году был принят ряд принципиальных решений — это, безусловно, указ Президента № 204, который сделал цифровую экономику государственной политикой. Цифровые технологии, платформенные решения проходят