

РЕФЕРАТ

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
ОСНОВНАЯ ЧАСТЬ.....	6
1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.....	6
1.1 Графы	7
1.1.1 Основные определения	7
1.1.2 Простые алгоритмы на графах	10
1.2 Классификаторы.....	13
1.2.1 Нормализация входных данных.....	14
1.2.2 Метод k-ближайших соседей.....	16
1.2.3 Дерево решений	19
1.2.4 Случайный лес	25
1.3 Метрики качества.....	26
2 ПРАКТИЧЕСКАЯ ЧАСТЬ.....	28
2.1 Постановка задачи	28
2.1.1 Описание данных.....	28
2.2 Предварительный анализ	30
2.3 Построение графов.....	32
2.3.1 Структура программы	32
2.3.2 Описание интерфейса и возможностей.....	33
2.3.3 Анализ результатов.....	34
2.4 Проектирование признаков.....	37
2.4.1 Отбор признаков	38
2.5 Результаты	41

ЗАКЛЮЧЕНИЕ	43
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	44

ВВЕДЕНИЕ

ОСНОВНАЯ ЧАСТЬ

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Графы

1.1.1 Основные определения

В данной работе графы будут использоваться в двух местах: для визуализации и анализа, а также при построении деревьев решений.

Определение 1.1. Графом называется упорядоченная пара $G = (V(G), E(G))$, где $V(G)$ - непустое множество вершин (узлов) графа G , а $E(G)$ – множество ребер графа G .

Далее будут обозначены некоторые вспомогательные определения:

- количество вершин графа называется его порядком и обозначается $v = |V(G)|$. В данной работе будут рассматриваться только конечные графы, т.е. множества вершин и ребер принимают конечное число значений;
- количество ребер графа называется его размером и обозначается $e = |E(G)|$;
- ребро $e = \{u, v\}$ соединяет вершины, называемые концевыми (концами) u и v ;
- соседними вершинами называются такие концы, которые соединены одним и тем же ребром;
- смежными называются ребра, имеющие общую концевую вершину;
- вершина называется изолированной, если она не является конечной ни для одного из ребер;
- висячей вершиной или листом называется вершина, которая является концом ровно одного ребра.

Если не говорится об обратном, то граф считается неориентированным (пример приведен на рис 1.1), это означает, что каждое его ребро имеет два конца, порядок которых не имеет значения.

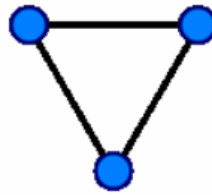


Рис. 1.1. Неориентированный граф

Пусть $(v, w) \in E$ называется дугой. Тогда вершину v называют её началом, а w — концом. Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

Определение 1.2. Граф называется ориентированным (орграфом), если $G = (V(G), E(G))$ — упорядоченная пара, где $V(G)$ — непустое множество вершин (узлов) графа G , а $E(G)$ — множество упорядоченных пар различных вершин, которые называются дугами. Пример такого графа приведен на рис. 1.2.

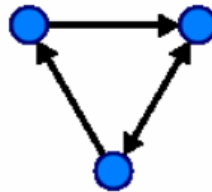


Рис. 1.2. Ориентированный граф

Вспомогательные определения для орграфов:

- маршрутом в графе называется конечная последовательность вершин, в которой каждая вершина (исключая последнюю) соединяется со следующей в последовательности вершиной ребром;
- путем в графе называют конечную последовательность вершин и дуг, в которой каждый элемент соединен с предыдущим и последующим;
- цепью называется маршрут без повторяющихся ребер;
- циклом называют цепь, в которой начальная и конечная вершины совпадают;
- цикл называют простым, если ребра в нем не повторяются;

- цикл называют элементарным, если он простой и вершины в нем не повторяются.

Определение 1.3. Граф H называется подграфом графа G , если $V(H) \subset V(G)$ и $E(H) \subset E(G)$.

Определение 1.4. 1) Вершины a и b графа G называются связанными, если в графе существует путь между ними.

2) Граф называется связным, если любые две его вершины связаны.

Существует много несколько разновидностей графов. В данной работе будет использоваться такие структуры как дерево и лес.

Определение 1.5. 1) Деревом называется связный граф без циклов.

2) Лесом называется упорядоченное множество деревьев.

Определение 1.6. Степенью вершины называется количество инцидентных ей ребер.

Определение 1.7. Двоичным (бинарным) деревом называется:

1. неориентированное дерево, степени вершин которого не превосходят 3;
2. ориентированное дерево, в котором число исходящих из каждой вершины ребер не превосходит 2.

Вершина, из которой выходят ребра называется родительской. А связанные с ней ребрами вершины называются левой и правой дочерней.

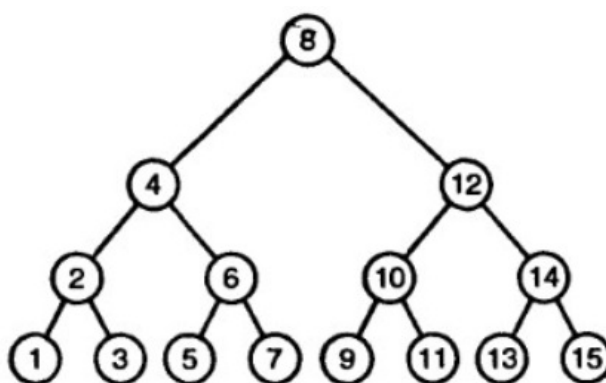


Рис. 1.3. Пример неориентированного бинарного дерева

1.1.2 Простые алгоритмы на графах

В этом разделе будет приведено несколько основных алгоритмов обхода графов. Обходя граф, мы двигаемся по ребрам и проходим все вершины. При этом накапливается довольно много информации, которая полезна для дальнейшей обработки графа.

1.1.2.1 Поиск в ширину

Поиск в ширину (breadth-first search) – один из базисных алгоритмов, составляющий основу многих других.

Пусть задан граф $G = (V(G), E(G))$ и фиксирована начальная вершина s .

Алгоритм поиска в ширину перечисляет все достижимые из s вершины, доступные при проходе по ребрам, в порядке возрастания расстояния от s . Расстоянием считается длина минимального пути из начальной вершины. Алгоритм применим как к ориентированным графам, так и к неориентированным.

Такое название объясняется тем, что в процессе поиска мы идем вширь, а не вглубь, т.е. сначала просматриваем все соседние вершины, затем соседей соседей и так далее.

Алгоритм 1.1.

0. Начинаем обход из фиксированной начальной вершины s . Пометить ее как посещенную. Добавить вершину в изначально пустую очередь.
1. Извлечь из начала очереди вершину u .
2. Если она является уже посещенной повторить шаг 1. Иначе добавить все соединенные с u не посещенные вершины в очередь и перейти к шагу 1.
3. Если очередь пустая, закончить алгоритм. В противном случае перейти к шагу 1.

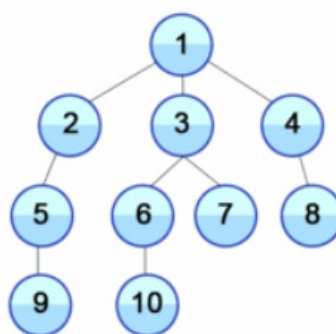


Рис. 1.4. Порядок обхода графа при поиске в ширину

На рис. 1.4. приведен пример работы алгоритма поиска в ширину. Цифры являются порядковым номером посещения вершины в процессе работы алгоритма.

1.1.2.2 Поиск в глубину

Поиск в глубину (depth-first search) наряду с вышеописанным алгоритмом также является одним из базисных методов обхода графа.

Он имеет следующую стратегию: как и в поиске в ширину, фиксируем начальную вершину s и начинаем от нее идти «вглубь», пока имеется такая возможность, т.е. пока существуют не пройденные ребра, затем возвращаться и искать иной путь, в случае, когда таких ребер не осталось. Алгоритм работает, пока не обнаружит все вершины, достижимые из исходной.

Алгоритм 1.2.

0. Начинаем обход из фиксированной начальной вершины s .
1. Пометить текущую вершину как посещенную.
2. Если есть соседние не посещенные вершины, перейти к одной из них и выполнить для нее алгоритм начиная с шага 1. Если все соседние вершины посещены, либо отсутствуют вовсе – закончить алгоритм.

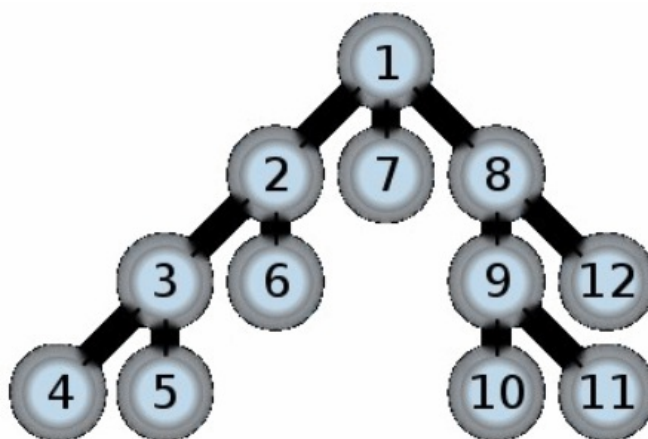


Рис. 1.5. Порядок обхода графа при поиске в глубину

На рис. 1.5. представлен пример работы алгоритма поиска в глубину. Аналогично рис. 1.4. цифры соответствуют порядковому номеру вершины при обходе графа алгоритмом 1.2.

Классификаторы

Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи. Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Классовая принадлежность остальных объектов неизвестна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Существуют несколько типов классов:

- двухклассовый, число классов равно двум;
- многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).
- непересекающиеся классы.
- пересекающиеся классы. Объект может относиться одновременно к нескольким классам.
- нечёткие классы. Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

В данной работе будет рассматриваться двухклассовый случай.

Введем некоторые базовые определения, которые будут использованы в дальнейшем.

Определение 1.8. Признаком (feature) называется результат измерения некоторой характеристики объекта. Можно сказать, что признак есть отображение $f: X \rightarrow D_f$, где D_f — множество допустимых значений признака.

В зависимости от природы этого множества, признаки делятся на нижеперечисленные типы:

- бинарный признак $D_f = \{0,1\}$;
- номинальный признак D_f — конечное множество;
- **порядковый признак: D_f — конечное упорядоченное множество;**
- количественный признак: $D_f = \mathbb{R}$.

Если все признаки имеют одинаковый тип, то исходные данные называются однородными, в ином случае – разнородными.

Определение 1.9. Пусть имеется набор признаков f_1, \dots, f_n . Признаковым описанием объекта $x \in X$ называют вектор $((f_1(x), \dots, f_n(x)))$, составленный из значений фиксированного набора признаков на данном объекте.

В задачах машинного обучения не делается различия между объектом и его признаковым описанием. Полагается, что $X = D_{f_1} \times \dots \times D_{f_n}$.

Постановка задачи классификации выглядит следующим образом. Пусть X — множество описаний (признаков) объектов. **Чем является объект, определяется спецификой предметной области.** Например, в задачах спортивного менеджмента объектами являются спортсмены.

Пусть Y — конечное множество номеров (имён, меток) классов. Существует неизвестная целевая зависимость — отображение $y^*: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. Требуется построить алгоритм $a(x): X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

1.2.1 Нормализация входных данных

Зачастую, чтобы достичь адекватности работы модели, необходимо нормализовать (масштабировать) входные данные. Как будет видно далее работоспособность некоторых моделей зависит от расстояния между объектами, вследствие чего возникает необходимость проведения данной

процедуры. Проблема заключается в разных измерениях признаков. Например, если рассматривать погоду, то такие ее признаки как температура, давление, скорость ветра и т.д. измеряются в различных физических величинах, а их числовые значения могут на порядки отличаться.

Нормализовать данные можно разными способами, вот два основных.

Минимаксная нормализация:

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}. \quad (0.1)$$

Данный метод осуществляет переход от абсолютных значений признаков к относительным. Новые переменные будут принимать значения в диапазоне от 0 до 1.

Z-нормализация:

$$x_i = \frac{x_i - \bar{x}}{s}, \quad (0.2)$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ – выборочное среднее, $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ – выборочное среднеквадратичное отклонение.

Поскольку не все признаки имеют количественные значения, может применяться создание фиктивных переменных (dummy coding). В этом случае категориальные признаки заменяются бинарными. Например, заменяется признак «пол» на два новых: «пол мужской», «пол женский» со значения 0 и 1.

1.2.2 Метод k-ближайших соседей

Алгоритмы, основанные на анализе сходства объектов, часто называют метрическими.

Метрическим классификатором (similarity-based classifier) называют алгоритм классификации, основанный на вычислении оценок сходства между объектами. Чтобы формализовать понятие сходства вводится функция расстояния между объектами $\rho(x, x')$ в пространстве объектов X . Следует заметить, что данная функция может не всегда удовлетворять всем аксиомам метрики. Например, довольно часто не выполняется неравенство треугольника.

Метрические классификаторы опираются на гипотезу компактности. Она, в свою очередь, предполагает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных. Можно сказать, что классы образуют компактно локализованные подмножества в пространстве объектов. То есть граница между классами имеет довольно простую форму.

Метод ближайшего соседа позиционируется, как один из простейших метрических классификаторов. Классифицируемый объект x относится к тому классу y_i , которому принадлежат ближайший к нему объект обучающей выборки x_i .

Метод k ближайших соседей (k-nearest neighbors algorithm, k-NN) для повышения надёжности классификации относит объект к тому классу, которому принадлежит большинство из его соседей, то есть k ближайших к нему объектов обучающей выборки x_i . В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Пусть задана обучающая выборка $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ и на множестве объектов задана функция расстояния $\rho(x, x')$. Данная функция должна быть достаточно адекватной моделью сходства объектов, для этого можно провести

процедуру нормализации, описанную в разделе 1.2.1. Чем больше значение этой функции, тем менее схожими являются два объекта x, x' .

Для произвольного объекта u расположим объекты обучающей выборки x_i в порядке возрастания расстояний до u : $\rho(u, x_{1,u}) \leq \rho(u, x_{2,u}) \leq \dots \rho(u, x_{m,u})$, где $x_{i,u}$ обозначает объект обучающей выборки, который является i -ым соседом объекта u . Аналогичное обозначение введём и для ответа на i -ом соседе: $y_{i,u}$. Таким образом, произвольный объект u порождает свою перенумерацию выборки. В наиболее общем виде алгоритм ближайших соседей выглядит так:

$$a(u) = \arg \max_{y \in Y} \sum_{i=1}^m [y(x_{i,u}) = y] w(i, u), \quad (0.3)$$

где $w(i, u)$ – заданная весовая функция, которая оценивает степень важности i -го соседа для классификации объекта u . Эта функция неотрицательна и не возрастает по i .

Различно задавая весовую функцию, получаются различные варианты методы ближайших соседей.

- $w(i, u) = [i = 1]$ – простейший метод ближайшего соседа;
- $w(i, u) = [i \leq k]$ – метод k ближайших соседей;
- $w(i, u) = [i \leq k] q^i$ – метод k экспоненциально взвешенных ближайших соседей, где предполагается $q < 1$ (обычно используется в случае 3-х и более классов).

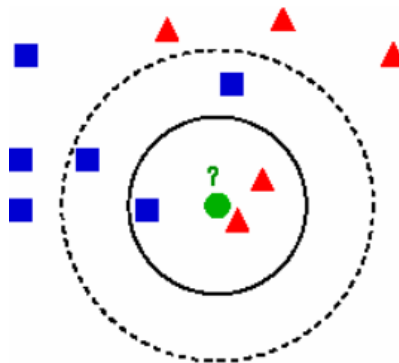


Рис. 1.6. Метод k ближайших соседей

На рис. 1.6. приведена ситуация классификации объекта, в данном случае зеленого круга. Круг должен быть классифицировать как синий квадрат, либо как красный треугольник (класс 1 и класс 2 соответственно). На иллюстрации видно два круга.

Круг, обведенный сплошной линией, показывает поведение алгоритма при $k=3$. В этом случае объект будет классифицирован как 2-ой класс, так как внутри круга находятся 2 треугольник и 1 квадрат, треугольников больше, а значит и решение принимается в сторону этого класса.

В кругу, обведенном штрихом, $k=5$. Тогда ситуация меняется, потому что количество квадратов начало превалировать над треугольниками. Соответственно и объект будет классифицирован как синий квадрат, то есть 1-й класс.

1.2.3 Дерево решений

Решающими деревьями называется семейство моделей, которые позволяют восстанавливать нелинейные зависимости произвольной сложности. Они воспроизводят логические схемы, позволяющие получить окончательное решение о классификации объекта с помощью ответов на иерархически организованную систему вопросов. Причем вопрос, задаваемый на последующем иерархическом уровне, зависит от ответа, полученного на предыдущем уровне.

Каждой из вершин дерева за исключением листьев соответствует некоторый вопрос, подразумевающий несколько вариантов ответов, соответствующих выходящим ребрам. В зависимости от выбранного варианта ответа осуществляется переход к вершине следующего уровня. Листьям поставлены в соответствие метки, указывающие на отнесение распознаваемого объекта к одному из классов.

Решающее дерево называется бинарным, если каждая внутренняя или корневая вершина инцидентна только двум выходящим рёбрам.

Определение 1.10. Рассмотрим бинарное дерево, в котором каждой внутренней вершине v приписана функция (или предикат) $\beta_v : X \rightarrow \{0,1\}$, а каждому листу v приписан прогноз $c_v \in Y$. В случае классификации листу может быть приписан вектор вероятностей. При классификации объекта $x \in X$ он проходит путь от корня дерева до некоторой конечной вершины, в соответствии с алгоритмом $a(x)$.

Теперь рассмотрим алгоритм $a(x)$, который стартует из корневой вершины v_0 и вычисляет значение функции β_{v_0} .

Алгоритм 1.3.

0. Алгоритм начинается с начальной вершины v_0 .
1. Далее вычисляется значение функции β_{v_0} .
2. Если текущая вершина является листом, вернуть класс, который приписан данной вершине и закончить алгоритм.

3. Если вычисленное значение равно нулю, то алгоритм переходит в левую дочернюю, в ином случае в правую. Для новой вершины выполняем шаг 1.

Такой алгоритм называется бинарным решающим деревом.

На практике в большинстве случаев используются одномерные предикаты β_v , которые сравнивают значение одного из признаков с порогом:

$$\beta_v(x; j, t) = [x_j < t].$$

Объект x доходит до вершины v тогда и только тогда, когда выполняется конъюнкция $K_v(x)$, составленная из всех предикатов, приписанных внутренним вершинам дерева на пути от корня v_0 до вершины v .

Пусть T — множество всех терминальных вершин дерева. Множества объектов $\Omega_v = \{x \in X : K_v(x) = 1\}$, выделяемых терминальными конъюнкциями $v \in T$, попарно не пересекаются, а их объединение совпадает со всем пространством X . Данное утверждение легко доказывается индукцией по числу вершин дерева. Отсюда следует, что решающее дерево никогда не отказывается от классификации. А также, что алгоритм классификации $a(x) : X \rightarrow Y$, реализуемый бинарным решающим деревом, можно представить в виде простого голосования конъюнкций:

$$a(x) = \arg \max_{y \in Y} \sum_{v \in T} [c_v = y] K_v(x), \quad (0.4)$$

причем для любого $x \in X$ одно и только одно слагаемое во всех этих суммах равно единице.

Легко увидеть, что для любой выборки можно реализовать решающее дерево, которое не допустит на ней ни одной ошибки. Даже с простыми одномерными предикатами можно сформировать дерево, в каждом листе которого находится ровно по одному объекту выборки. Вероятнее всего, такое дерево будет переобученным и не сможет показать хорошее качество классификации на новых данных.

1.2.3.1 Построение дерева

Опишем базовый жадный алгоритм построения бинарного решающего дерева.

Алгоритм 1.4

0. В качестве начальной возьмем всю обучающую выборку X найдем наилучшее ее разбиение на две части $R_1(j, t) = \{x | x_j < t\}$ и $R_2(j, t) = \{x | x_j \geq t\}$ с точки зрения заранее заданного функционала качества $Q(x, j, t)$.
1. Найдем наилучшие значения j и t , создадим корневую вершину дерева v_0 и примем ее за текущую.
2. В текущей вершине проверяем, не выполнилось ли некоторое условие останова. Если выполнилось, то прекращаем рекурсию и объявляем эту вершину листом. Переходим к шагу 5. Если же не выполнилось, то к шагу 3.
3. Поставим в соответствие текущей вершине предикат $[x_j < t]$. Объекты разобьются на две части — одни попадут в левое поддереву, другие в правое.
4. Для каждой из подвыборок, получившихся на шаге 3, рекурсивно повторим шаг 1, построив дочерние вершины для корневой, и так далее.
5. В построенном дереве каждому листу приписывают ответ. В случае классификации — класс, к которому больше всего относится объектов в листе или вектор вероятности. Алгоритм завершен.

Выбор конкретной функции зависит от функционала качества в исходной задаче. Таким образом, конкретный метод построения решающего дерева определяется:

- видом предикатов в вершинах;
- функционалом качества $Q(x, j, t)$;

- критерием останова.

Далее подробнее рассмотрим каждый пункт из списка выше.

При построении дерева необходимо задать функционал качества, на основе которого будет осуществляться разбиение выборки на каждом шаге. Пусть R_m — множество объектов, которые попали в вершину, разбиваемую на текущем шаге, а R_l и R_r — подмножества R_m , состоящие из объектов, попадающих при заданном предикате в левое и правое поддерево соответственно. Будет использован функционал следующего вида:

$$Q(R_m, j, s) = H(R_m) - \frac{|R_l|}{|R_m|} H(R_l) - \frac{|R_r|}{|R_m|} H(R_r), \quad (0.5)$$

где $H(R)$ есть критерий информативности (impurity criterion), который оценивает качество распределения целевой переменной среди объектов множества R . Чем меньше разнообразие целевой переменной, тем меньше должно быть значение критерия информативности — и, соответственно, мы будем пытаться минимизировать его значение. Функционал качества $Q(R_m, j, s)$ мы при этом будем максимизировать.

Как уже обсуждалось выше, в каждом листе дерево будет выдавать константу — вещественное число, вероятность или класс. Исходя из этого, можно предложить оценивать качество множества объектов R тем, насколько хорошо их целевые переменные предсказываются константой (при оптимальном выборе этой константы):

$$H(R) = \min_{c \in Y} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} L(y_i, c) \quad (0.6)$$

где $L(y_i, c)$ — некоторая функция потерь. Далее мы обсудим, какие именно критерии информативности часто используют в задачах классификации.

1.2.3.2 Критерии информативности

Обозначим через p_k долю объектов класса k ($k \in \{1, \dots, K\}$), попавших в вершину R :

$$p_k = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y_i = k], \quad (0.7)$$

Через k_* обозначим класс, представителей которого нашлось больше остальных среди объектов, попавших в данную вершину:

$$k_* = \arg \max_k p_k. \quad (0.8)$$

Приведем в пример два критерия информативности.

Первым будет ошибка классификации. Рассмотрим индикатор ошибки, как функцию потерь. Подставим в (1.6):

$$H(R) = \min_{c \in Y} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y_i \neq c], \quad (0.9)$$

Заметно, что оптимальным предсказанием тут будет наиболее популярный класс k_* . Это означает, что критерий будет равен следующей доле ошибок:

$$H(R) = \min_{c \in Y} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y_i \neq k_*] = 1 - p_{k_*}, \quad (0.10)$$

Данный критерий считается довольно грубым, так как учитывает частоту p_{k_*} лишь одного класса.

Вторым обозначим Критерий Джини. Рассмотрим ситуацию, когда в вершине выдается не один класс, а распределение по всем классам $c = (c_1, \dots, c_K)$, $\sum_{k=1}^K c_k = 1$. Качество такого распределения можно измерять с помощью критерия Бриера (Brier score):

$$H(R) = \min_{\sum_k c_k = 1} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} \sum_{k=1}^K (c_k - [y_i = k])^2, \quad (0.11)$$

Можно показать, что оптимальный вектор вероятностей состоит из долей классов p_k : $c_* = (p_1, \dots, p_K)$.

Если подставить эти вероятности в исходный критерий информативности (1.6), провести ряд преобразований, то на выходе получим критерий Джини:

$$H(R) = \sum_{k=1}^K p_k (1 - p_k). \quad (0.12)$$

1.2.3.3 Критерии останова

Можно придумать большое количество критериев останова. Перечислим некоторые ограничения и критерии:

- Ограничение максимальной глубины дерева.
- Ограничение минимального числа объектов в листе.
- Ограничение максимального количества листьев в дереве.
- Останов в случае, если все объекты в листе относятся к одному классу.
- Требование, что функционал качества при дроблении улучшался как минимум на s процентов.

С помощью грамотного выбора подобных критериев и их параметров можно существенно повлиять на качество дерева.

1.2.4 Случайный лес

В основе алгоритма случайный лес лежит идея использования ансамбля (комитета) решающих деревьев. Сами по себе входящие в ансамбль решающие деревья дают невысокий результат классификации, но за счет большого их количества, решение, принимаемое голосованием, получается достаточно хорошим.

Зададим обучающую выборку, количество образцов в которой равно N . Пусть размерность пространства признаков будет равна M . Параметр $m \approx \sqrt{M}$ (такое значение используется в задачах классификации) равен количеству неполных признаков, используемых для обучения.

Рассмотрим наиболее распространенный способ построения отдельных экземпляров деревьев для случайного леса.

Алгоритм 1.5.

1. Генерируется случайная подвыборка с возвращением размером n . Некоторые образцы попадут в выборку более одного раза, а некоторые не попадут вовсе.
2. Формируется решающее дерево в соответствии с алгоритмом 1.4. При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производится разбиение. Для выбора лучшего из этих признаков используется критерий Джини.
3. Дерево строится до срабатывания выбранного критерия останова, которые были описаны в разделе 1.2.2. Чаще всего этим критерием является полное исчерпание подвыборки.

Данный алгоритм повторяется K (количество деревьев в лесу) раз. Получившийся лес классифицирует объекты путём голосования. Процесс голосования выглядит следующим образом: каждое дерево в комитете относит объект к одному из классов. Побеждает класс, который набрал наибольшее количество голосов.

■ Метрики качества

После того как модель построена и дает результаты, возникает вопрос о качестве работы данной модели и правдоподобности результатов. Прежде чем говорить о подборе той или иной метрики для оценки качества классификатора, необходимо ввести понятие матрицы ошибок.

Пусть дана обучающая выборка, над ней произведена процедура бинарной классификации, в результате которой получен вектор ответов $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$. Поскольку выборка обучающая, то вектор ожидаемых результатов $y = (y_1, \dots, y_n)$ известен. Класс объектов равных 1, будем называть C_1 , а объектов равных 0 – C_0 .

Тогда матрица ошибок будет выглядеть следующим образом, который показан в таблице 1.1.

Таблица 1.1 Матрица ошибок

	$\hat{y}_i = 0$	$\hat{y}_i = 1$
$y_i = 0$	TN	FP
$y_i = 1$	FN	TP

Получившиеся на пересечении значения означают:

- TN – количество верно предсказанных классификатором объектов класса C_0 .
- TP – количество верно предсказанных классификатором объектов класса C_1 .
- FP – количество ошибок первого рода, которые означают, что классификатор определил класс C_0 , как класс C_1 .
- FN – количество ошибок второго рода, которые означают, что классификатор определил класс C_1 , как класс C_0 .

Первой очевидной метрикой, которая интуитивно приходит на ум является доля правильных ответов:

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP}. \quad (0.13)$$

На практике такая метрика очень редко используется, потому что является бесполезной в задачах, где количество объектов одного класса сильно превалирует на другим.

Далее введем метрики для оценки работы модели на каждом из классов по отдельности. Первой будет точность (precision):

$$precision = \frac{TP}{TP + FP}, \quad (0.14)$$

Она показывает долю объектов, определенных классификатором как класс C_1 , которые в действительности относятся к классу C_1 .

Второй метрикой будет полнота (recall):

$$recall = \frac{TP}{TP + FN}, \quad (0.15)$$

она показывает, какую долю объектов класса C_1 из всех его объектов нашел классификатор.

Объединяет последние две метрики F_1 -мера. Которая является средним гармоническим точности и полноты:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (0.16)$$

Очевидно, что F_1 -мера будет максимальной при максимальных значениях полноты и точности, и будет близка к нулю, если один из аргументов близок к нулю.

2 ПРАКТИЧЕСКАЯ ЧАСТЬ

Практическим результатом работы является обученная модель, способная определять потенциально мошеннические транзакции. Также была спроектировано прикладное программное обеспечение, позволяющее визуализировать в виде ориентированных графов взаимодействие клиентов посредством финансовых транзакций.

■ Постановка задачи

Ставится задача анализа финансовых транзакций с целью предотвращения мошеннических операций. Данная задача разбивается на две подзадачи:

1. Реализовать классификатор способный отличить мошеннические транзакции от не являющихся таковыми.
2. Построить графы на основе финансовых переводов. На графах посчитать новые параметры (расстояние до ближайшего мошенника, количество мошенников в круге радиуса n), добавить их в модель и проверить значимость.

2.1.1 Описание данных

Данные, на которых будут проводится исследования, представляют собой файл с расширением `.csv`, в котором содержится информация о 6362620 финансовых транзакциях. Мошеннические транзакции в нем помечены. Следует отметить, что этот набор данных сгенерирован синтетически (по соображениям конфиденциальности) на основе реального, предоставленного африканским сервисом мобильных переводов. Временное окно данных – 1 месяц. Для отслеживания времени введена условная переменная, равная 1 часу.

В таблице 2.1 приведены признаки, которые заданы для каждой транзакции в выборке.

Таблица 2.1 Описание признаков

Название в .csv файле	Смысл	Принимаемые значения
<i>step</i>	Аналог времени (1 step = 1 час)	1–744 (30 дней)
<i>type</i>	тип транзакции	CASH-IN, CASH-OUT, DEBIT, PAYMENT, TRANSFER
<i>amount</i>	сумма перевода	Действительное число больше нуля
<i>nameOrig</i>	ID пользователя-отправителя	ID в формате 'C165564'
<i>oldbalanceOrg</i>	баланс отправителя до транзакции	Действительное неотрицательное число
<i>newbalanceOrig</i>	баланс отправителя после транзакции	Действительное неотрицательное число
<i>nameDest</i>	ID пользователя-получателя	ID в формате 'C165564' (магазин, если первая буква ID – M)
<i>oldbalanceDest</i>	баланс получателя до транзакции	Действительное неотрицательное число
<i>newbalanceDest</i>	баланс получателя после транзакции	Действительное неотрицательное число
<i>isFraud</i>	пометка о мошеннической транзакции	1 – мошенническая 0 – в противном случае

Пример строки в файле:

184,CASH_IN,90687.53,C594792269,14620285.23,14710972.76,C14851727
63,166355.91,75668.37,0,0,16,1,1,0,0,0,0

■ Предварительный анализ

Прежде, чем перейти к построению модели, было принято решение посчитать (увидеть) некоторые очевидные вещи, которые лежат на поверхности.

В результате была получена информация, приведенная в таблице 2.2.

Таблица 2.2 Данные, полученные при предварительном анализе

Название	Значение
Всего транзакций	6362620
Мошеннических	8213
Средняя сумма перевода мошеннических транзакций	1467967.29
Средняя сумма перевода не мошеннических транзакций	178197.04
Уникальных клиентов	4777844
Количество магазинов	2151495
Уникальных магазинов	2150401
Клиенты, взаимодействующие друг с другом более 1 раза	0
Мошеннических транзакций с магазинами (из них мошеннических)	0
Количество типов CASH-IN	1399284 (0)
Количество типов CASH-OUT	2237500 (4116)
Количество типов DEBIT	41432 (0)
Количество типов PAYMENT	2151495 (0)
Количество типов TRANSFER	532909 (4097)

Анализируя полученные данные, можно сделать выводы:

- присутствует сильный дисбаланс между классом мошеннических и не мошеннических транзакций. Были приняты решение сократить выборку до ~500.000 транзакций, причем количество мошеннических оставить равным исходному. Преимущество этого шага еще и в том, что теперь для обучения модели будет требоваться меньше памяти. Факт нехватки памяти сильно тормозил (иногда и вовсе не давал посчитать результаты) процесс исследования;
- средняя сумма мошеннических транзакций на порядок выше не мошеннических. Значит сумма является важным критерием для классификации в этом наборе данных;
- количество уникальных клиентов достаточно велико. Данный факт говорит о том, что взаимосвязь между клиентами весьма слабая;
- неправдоподобным выглядит отсутствие клиентов, взаимодействующих друг с другом более 1 раза. В жизненных условиях всегда найдутся клиенты, активно переводящие друг другу средства. Скорее всего этот недостаток, обусловлен синтетической природой данной выборки;
- уникальных магазинов также очень много, почти столько же, сколько и всего транзакций, производимых с их участием. Количество магазинов совпадает с количеством транзакций с типом PAYMENT. А также мошеннические операции не проводятся в рамках контактов с магазинами;
- мошеннические транзакции производятся в пределах двух типов: CASH-OUT и TRANSFER. Наряду с суммой перевода, этот факт будет играть одну из ключевых ролей при классификации.

■ Построение графов

Для построения и визуализации графов использовалась интегрированная среда разработки Visual Studio 2017. Прикладная программа была разработана на языке C# при помощи технологии WPF (Windows Presentation Foundation).

2.3.1 Структура программы

Необходимо перевести данные из построчного представления в.csv файле в объектный вид, чтобы использовать преимущества объектно-ориентированной разработки.

Для этого были реализованы такие классы:

- *Transaction* – данный класс является объектным представлением одной строки исходного файла. В нем абсолютно те же поля, что и набор признаков в выборке. Но вместо строчного представления имени клиента в нем хранится класс *Client* как для отправителя, так и для получателя.
- *Client* – в данном классе хранится ID клиента, список его транзакций в качестве отправителя, а также в качестве получателя, пометка о том, был ли замечен этот клиент в мошеннических переводах. Также тут может содержаться дополнительная информация, например, максимальная длина цепочки, в которой замечен клиент, и расстояние до ближайшего мошенника.

При запуске программы происходит чтение данных, преобразование их к объектному виду. В результате этого формируется список клиентов, список транзакций и формируется граф на основе словаря.

Можно упрощенно описать работу программы алгоритмом 2.1.

Алгоритм 2.1.

Перед началом инициализируем пустой список транзакций и клиентов.

1. Читается каждая строка исходного файла. Из нее создаются экземпляры классов транзакции и клиента, которые добавляются в свои списки соответственно.

2. Итеративно продвигаясь по всему списку транзакций формируется словарь, в котором ключом выступает клиент, а значением – список клиентов, с которыми он взаимодействует.
3. На сформированном на шаге 2 графе, итеративно двигаясь по ключам, с помощью алгоритма 1.2 находим вспомогательные признаки (максимальная цепочка, в которой участвует клиент и расстояние до ближайшего мошенника).

2.3.2 Описание интерфейса и возможностей

После обработки данных, описанной в алгоритме 2.1 перед пользователем появляется интерфейсное окно с визуализированным графом (рис. 2.1).

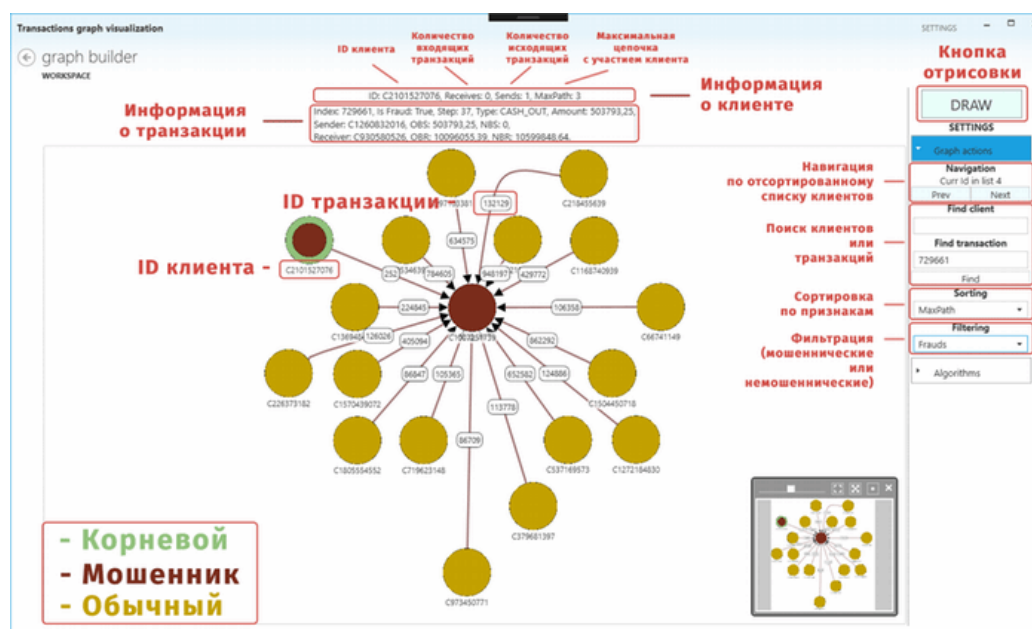


Рис. 2.1 Интерфейсное окно прикладной программы

В центральной части окна располагается поле визуализации графа. Зеленым цветом помечен клиент, с которого началось построение графа. Будем называть таких клиентов корневыми. Желтым цветом помечены обычные клиенты, не замеченные в мошеннических транзакциях. Красным помечены мошенники. Если клиент является одновременно и корневым и

мошенником, то тогда он отмечается красно-зеленым цветом. Строка под кругом – ID клиента.

Направление стрелок на визуализации указывает от отправителя к получателю. Цифры посередине стрелки означают ID транзакции, который соответствует номеру строки в исходном файле.

В верхней части окна выводится информация по выбранному клиенту и транзакции. Про клиента пишется ID, количество транзакций в роли получателя, количество транзакций в роли отправителя и максимальная цепочка, в которой он был замечен. Про транзакцию также пишется ее ID и вся информация, которая доступна из строки исходного файла, то есть все признаки.

В правой части окна располагается интерфейс взаимодействия. В нем можно изменить сортировку клиентов, отфильтровать их, найти клиента или транзакцию по ID. Все изменения вступают в силу после нажатия кнопки «DRAW».

По умолчанию клиенты отсортированы по убыванию максимальной цепочки транзакций. Можно также отсортировать по числу выступлений получателем или отправителем, а также по расстоянию до ближайшего мошенника.

Фильтрация по умолчанию отключена. Но можно отфильтровать клиентов, оставив только мошенников.

2.3.3 Анализ результатов

Построив граф на отфильтрованных данных (как было предложено в разделе 2.2). Оказалось, что длина самой длинной цепи равна четырем. Предположения о разреженности графа из раздела 2.2 подтвердились. На раннем этапе исследования на полном наборе данных максимальная длина цепи была равна 9.

Соответственно предложенные для добавления в модель признаки из постановки задачи (раздел 2.1.1) не имеют смысла. Дистанция до ближайших мошенников почти у всех клиентов отсутствует, либо равна 1 или 0 (случай, когда сам клиент мошенник). Максимальной зафиксированной дистанцией была цепь длиной в 2 транзакции. Искать количество мошенников в круге радиуса n , также не принесет особых результатов, поскольку общая картина такова: граф разбит на кучу маленьких подграфов. Один из таких подграфов представлен на рис 2.2.

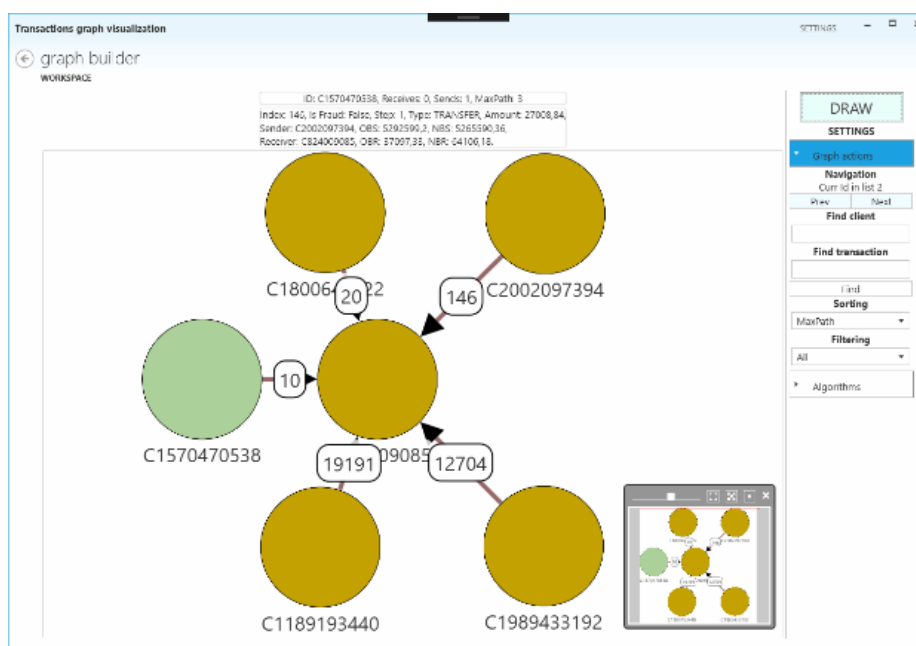


Рис. 2.2 Типичная ситуация для отдельно взятого клиента

Визуально анализируя полученные результаты, удалось установить, что типичная картина мошеннических клиентов выглядит следующим образом: клиент выступает получателем, к которому стекаются несколько транзакций, одна из которых мошенническая (рис. 2.3).

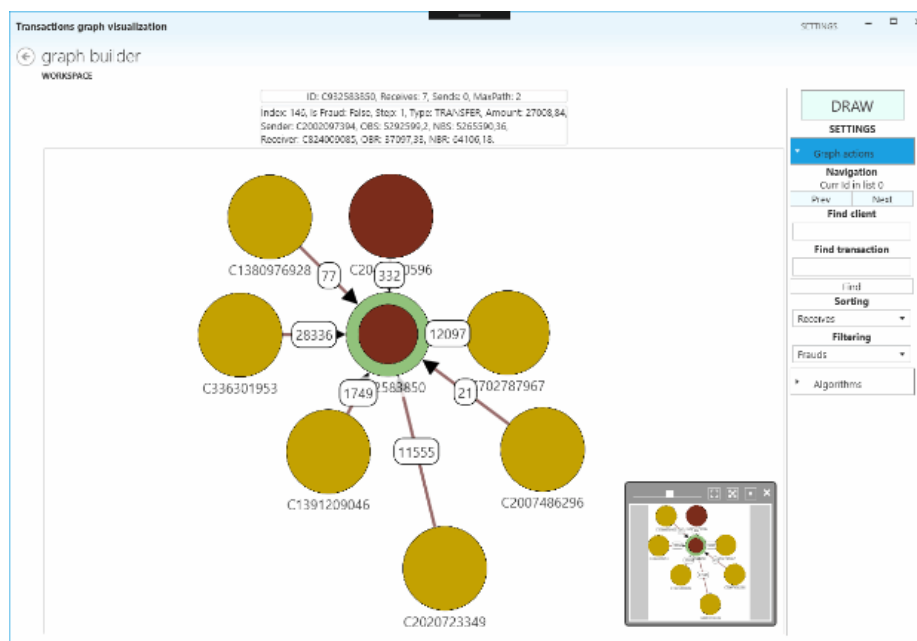


Рис. 2.3 Типичная ситуация для мошенника

Признаки, выявленные на графах, показали свою практическую неприменимость. Поэтому следует уделить большее внимание задаче классификации.

■ Проектирование признаков

Проектирование признаков (feature engineering) производится с целью повысить эффективность прогнозирования алгоритмов обучения путем выявления новых характеристик или исключения имеющихся из исходных данных. Данная процедура позволяет упростить процесс обучения.

Для сравнения в список используемых моделей (помимо описанных в разделах 1.2.2 и 1.2.4) добавляется наивный байесовский классификатор. Он основан на применении теоремы Байеса со строгими предположениями о том, что все объекты описываются независимыми признаками.

Проведение экспериментов происходит при помощи языка Python и библиотеки scikit-learn, внутри которой реализация моделей, а также методы расчета метрик, описанных в разделе 1.3.

Отправной точкой будет результат обучения модели на «сырых» данных, то есть никаких добавлений или исключений признаков. Результаты такого эксперимента приведены в таблице 2.3.

Таблица 2.3 Метрики качества на «сырых» данных

	<i>precision</i>	<i>recall</i>	F_1
Наивный байесовский	0.595	0.204	0.304
k-ближайших соседей	0.954	0.794	0.867
Случайный лес	0.994	0.831	0.905

Лучше всего себя повел случайный лес и уже показал неплохие результаты. Недавно добавленный наивный байесовский классификатор довольно плохо предсказал результат. А метод ближайших соседей повел себя уверенно, но все же уступает случайному лесу.

Далее производится отбор признаков, затем эксперимент будет еще раз проведен на новом наборе данных.

2.4.1 Отбор признаков

Для устранения сильно коррелирующих между друг другом признаков строится матрица корреляции (рис. 2.4).

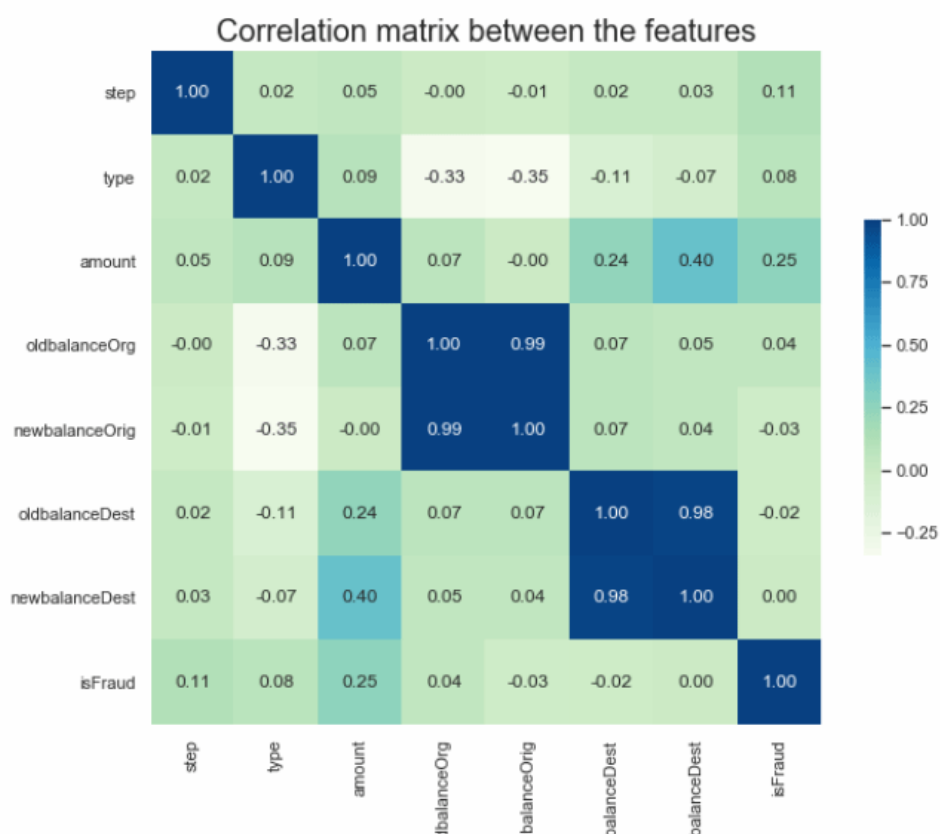


Рис. 2.4 Матрица корреляции исходных признаков

Очевидно, что балансы до и после совершения транзакции будут сильно коррелировать, поэтому можно отбросить балансы после транзакции. Информация по старому балансу получателя тоже практически ни на что не влияет, поэтому отбросим и этот признак.

Далее проектируются новые признаки на основе имеющихся. Создается новый .csv файл. Ниже перечисляются добавленные признаки (в скобках указано имя в новом файле):

- время (*hour*). Условный шаг из исходных данных переведен в 24-часовой формат. Может быть мошеннические транзакции совершаются в определенное время суток;

- новый ли отправитель (*newSender*). Пометка, показывающая выступал ли ранее текущий отправитель в такой же роли;
- новый ли получатель (*newReceiver*). Пометка, показывающая выступал ли ранее текущий получатель в такой же роли;
- является ли получатель магазином (*merchant*). Пометка, показывающая является ли принимающая сторона магазина. Поскольку с магазином не совершаются злоумышленные транзакции, этот пункт может быть полезным;
- был ли хоть один из клиентов, участвующих в текущей транзакции, ранее замечен в нарушениях (*fraudsEarly*). Пометка может быть полезна, если клиент совершает кражи регулярно;
- время, прошедшее с момента последней транзакции в качестве отправителя (*LTS*). Количество часов указывается количество часов, либо ставится -1, если первое появление.
- время, прошедшее с момента последней транзакции в качестве получателя (*LTR*). Аналогично пункту, написанному выше;
- остается ли ноль на балансе отправителя (*IZoB*). Если клиент остается с 0 на балансе, то скорее всего его обчистил мошенник.

Далее с помощью встроенного в библиотеку `scikit-learn` метода `RFE` (`recursive feature elimination`) производится ранжирование признаков по значимости. Название метода переводится как «рекурсивное отсечение признаков». Алгоритм его работы следующий: модель обучается на исходном наборе признаков, оценивает их значимость, отсекает один или несколько наименее значимых, обучается на новых признаках и повторяет эти действия, пока не останется n наиболее значимых признаков.

Также у класса случайного леса из библиотеки `scikit-learn` есть метод, возвращающий значимость каждого признака.

Результаты работы двух этих методов приведены в таблице 2.4.

Таблица 2.4 Результаты работы алгоритма RFE и значимость признаков

	RFE	Значимость признака
<i>step</i>	1	0.085
<i>type</i>	2	0.062
<i>amount</i>	1	0.243
<i>oldbalanceOrg</i>	1	0.389
<i>hour</i>	1	0.110
<i>newSender</i>	7	7.973e-06
<i>newReceiver</i>	4	0.010
<i>merchant</i>	5	0.009
<i>fraudsEarly</i>	6	0.0005
<i>LTS</i>	8	2.28e-08
<i>LTR</i>	3	0.016
<i>IZoB</i>	1	0.071

Результаты удивляют. Самым значимым признаком оказался баланс до транзакции на счету отправителя. Не совсем понятно, с чем связана такая зависимость. Возможно, так получилось из-за синтетической природы исходной выборки. Второй по значимости признак – ожидаемо, сумма. Из новых добавленных признаков наиболее значимыми оказались время в часах, индикатор нуля на балансе, время с последней транзакции. Остальные признаки не проявили себя, их можно отбросить. Также можно отбросить шаг, у 24-часового аналога значимость больше.

Отбросив менее значимые параметры, проводится эксперимент и сравниваются результаты с предыдущей итерацией. Далее отбрасываем еще мало значимые признаки, пока не достигнем максимального результата.

Результаты

Таким образом, остались следующие признаки *type*, *amount*, *oldBalanceOrg*, *hour*, *iZoB*.

Также была проведена процедура нормализации данных, описанная в разделе 1.2.1

Результат модели на конечных данных, а также его сравнение с результатом на исходных данных приведены в таблице 2.5.

Таблица 2.5 Результат работы модели на конечных данных

		Наивный байесовский	Случайный лес	k-ближайших соседей
<i>precision</i>	Исходные	0.595	0.994	0.954
	Конечные	0.547	0.985	0.955
	Разница	-0.048	-0.009	0.001
<i>recall</i>	Исходные	0.204	0.831	0.794
	Конечные	0.189	0.915	0.959
	Разница	-0.015	0.084	0.165
F_1	Исходные	0.304	0.905	0.867
	Конечные	0.281	0.949	0.957
	Разница	-0.023	0.044	0.90

Лучше всего себя проявил метод k-ближайших соседей. По сравнению с результатами на исходных данных прогресс явно заметен. Случайный лес тоже выдал хорошее качество предсказаний. **Наивный байесовский классификатор ухудшил свои результаты.**

Таким образом, можно сделать вывод, что метрический метод лучше всего подходит для решения данной задачи. Наравне с ним можно применять ансамбль из решающих деревьев. А вот метод, основанный на теореме Байеса, проявил себя хуже всего, его точно не стоит использовать для решения задач такого типа.

Для более наглядного сравнения результатов построены precision-recall кривые всех трех методов (рис. 2.5).

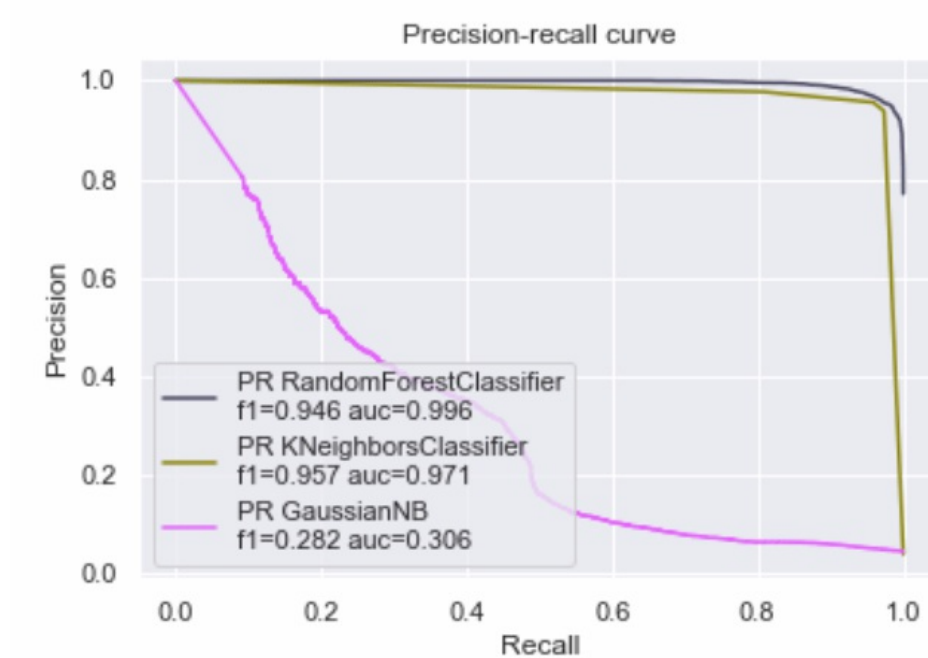


Рис. 2.5 PR-кривые использованных методов

ЗАКЛЮЧЕНИЕ

В выпускной квалификационной работе выполнены все поставленные задачи в полном объеме:

- проведен анализ финансовых транзакций с целью предотвращения мошеннических операций;
- реализована прикладная программа, визуализирующая графы на основе финансовых транзакций;
- реализована модель, способная классифицировать финансовые транзакции;
- проведена процедура проектирования признаков, которая привела к улучшению результатов работы модели;
- проведен анализ эффективности реализованной модели и ее сравнение с другими методами.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Признаковое описание - MachineLearning.ru

[http://www.machinelearning.ru/wiki/index.php?](http://www.machinelearning.ru/wiki/index.php?title=%D0%9F%D1%80%D0%B8%D0%B7%D0%BD%D0%B0%D0%BA%D0%BE%D0%B2%D0%BE%D0%B5_%D0%F)

[title=%D0%9F%D1%80%D0%B8%D0%B7%D0%BD%D0%B0%D0%BA%D0%BE%D0%B2%D0%BE%D0%B5_%D0%F](http://www.machinelearning.ru/wiki/index.php?title=%D0%9F%D1%80%D0%B8%D0%B7%D0%BD%D0%B0%D0%BA%D0%BE%D0%B2%D0%BE%D0%B5_%D0%F)
2008

Проверяемый текст

количественный признак: f_D Если все признаки имеют одинаковый тип, то исходные данные называются однородными, в ином случае – разнородными. Пусть имеется набор признаков $1, \dots, n_f$. Признаковым описанием объекта $x \in X$ называют вектор $1((f_1(x), \dots, f_{n_f}(x)))$, составленный из значений фиксированного набора признаков на данном объекте. В задачах машинного обучения не делается различия между объектом и его признаковым описанием. Полагается, что $1 \in n_f \times D$. Постановка задачи классификации выглядит следующим образом. Например, в задачах спортивного менеджмента объектами являются спортсмены.

1.2.1 Нормализация входных данных

Зачастую, чтобы достичь адекватности работы модели, необходимо нормализовать (масштабировать) входные данные. Как будет видно далее, работоспособность некоторых моделей зависит от расстояния между объектами, вследствие чего возникает необходимость проведения данной операции.

<http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0>
<http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0>
<http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC> 15 процедуры. Проблема заключается в разных измерениях признаков. Например, если рассматривать погоду, то такие ее признаки как температура, давление, скорость ветра и т.д.

Текст источника

Признаковое описание **объекта** — это **вектор, составленный из значений** $\langle \dots \rangle$. Пусть X — множество объектов (ситуаций, прецедентов). $\langle \dots \rangle$ признаком ($?feature$) называется результат измерения $\langle \dots \rangle$ бинарный **признак**: $D_f = \{0, 1\}$ $\langle \dots \rangle$. Если все признаки имеют одинаковый тип, то исходные данные $\langle \dots \rangle$ Определения · Признак · Матрица объектов $\langle \dots \rangle$

2. Классификация заёмщиков на основе тестовой выборки обезличенных данных в сфере банковского кредитования

<http://text.rucont.ru/History/Item?ServiceCallID=76491>

Петрова Ольга Владимировна

Проверяемый текст

А также, что алгоритм классификации $(\cdot) : a \in X \times Y$, реализуемый бинарным решающим деревом, можно представить в виде простого голосования конъюнкций: $(\cdot) \arg \max [(\cdot) \vee v \in Y \vee T \wedge x \in K \vee x, (0.4)]$ причем для любого $x \in X$ одно и только одно слагаемое во всех этих суммах равно единице.

Объект x доходит до вершины v тогда и только тогда, когда выполняется конъюнкция $(\cdot) \vee K \vee x$, составленная из всех предикатов, приписанных внутренним вершинам дерева на пути от корня 0_v до вершины v .

Текст источника

Отсюда следует, что алгоритм классификации $a : X \rightarrow Y$, реализуемый бинарным решающим деревом, можно записать в виде простого голосования конъюнкций: $\dots a(x) = \arg \dots$ причём для любого $x \in X$ одно и только одно слагаемое во всех этих суммах равно единице. $\langle \dots \rangle$

· Объект x доходит до вершины v тогда и только тогда, когда выполняется конъюнкция, составленная из всех предикатов, приписанных внутренним вершинам дерева на пути от корня до вершины v . $\langle \dots \rangle$

Множества объектов $\{x \in X \mid \varphi(x) = 1\}$, выделяемых терминальными конъюнкциями $\varphi \in T$, попарно не пересекаются, а их объединение совпадает со всем пространством X .

Рассмотрим бинарное дерево, в котором каждой внутренней вершине v приписана функция (или предикат) $\varphi_v : \{0, 1\}^X \rightarrow \{0, 1\}$, а каждому листу l приписан прогноз $\psi_l \in Y$.

Чем больше значение этой функции, тем менее схожими являются два объекта x, x' .

Наивный байесовский классификатор ухудшил свои результаты.

Вершина, из которой выходят ребра называется родительской.

Пусть T — множество всех терминальных вершин дерева.

Множества объектов $\{x \in X \mid \varphi(x) = 1\}$, выделяемых терминальными конъюнкциями $\varphi \in T$, попарно не пересекаются, а их объединение совпадает со всем пространством X .

Бинарное решающее дерево — это алгоритм классификации, задающийся **бинарным деревом**, в котором каждой внутренней вершине v приписан предикат $\varphi_v : \{0, 1\}^X \rightarrow \{0, 1\}$, каждой терминальной вершине l приписано имя класса $\psi_l \in Y$.

Чем больше значение этой функции, тем менее схожими являются два объекта.

Наивный байесовский классификатор.

Вершина, не имеющая выходящих рёбер, называется терминальной или листом.

Пусть T — множество всех терминальных вершин дерева.

3. Распознавание режимов работы авиационного ГТД с использованием технологии нейронных сетей

http://knowledge.allbest.ru/programming/2c0a65635a3ad68a5c43a88521316d26_0.html

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм $\varphi : X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Пусть Y — конечное множество номеров (имён, меток) классов.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Анализ существующих методик распознавания режимов работы газотурбинного двигателя **Задача классификации** Классификация — один из разделов машинного обучения, посвященный решению следующей задачи.

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Классифицировать объект — значит, указать номер (или наименование класса), к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Пусть — множество описаний объектов, — конечное множество номеров (имён, меток) классов.

Классовая принадлежность остальных объектов не известна.

4. Методы замкнутых описаний в задаче классификации данных со сложной структурой

http://www.frccsc.ru/sites/default/files/docs/ds/002-073-05/diss/12-kashnitskiy/12-kashnitskiy_main.pdf?933

Кашницкий Юрий Савельевич

Проверяемый текст	Текст источника
Они воспроизводят логические схемы, позволяющие получить окончательное решение о классификации объекта с помощью ответов на иерархически организованную систему вопросов.	Деревья решений Дерево решений воспроизводит логические правила, позволяющие получить окончательное решение о классификации объекта с помощью ответов на иерархически организованную систему вопросов. <...>
Существует неизвестная целевая зависимость — отображение $*: y \in X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $1, 1, \{ (x_1, y_1), \dots, (x_n, y_n) \}$.	Существует неизвестная целевая зависимость $y : X \rightarrow Y$ — отображение, значения которого известны только на объектах конечной обучающей выборки $X = \{ (x_1, y_1), \dots, (x_n, y_n) \}$. <...>
Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.	Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества, то есть, указать номер (или наименование класса) , к которому относится данный объект . <...>
Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.	Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества , то есть, указать номер (или наименование класса) , к которому относится данный объект . <...>
Требуется построить алгоритм $(\cdot) : x \in X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.	Требуется построить алгоритм $a : X \rightarrow Y$, способный приближать целевую функцию $y : (x)$ для произвольного объекта $x \in X$. <...>
Задано конечное множество объектов, для которых известно, к каким классам они относятся.	Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>
Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.	Имеется множество объектов (ситуаций), описанное с помощью некоторого множества признаков и разделённое некоторым образом на классы. <...>
Пусть Y — конечное множество номеров (имён, меток) классов.	Формализация постановки задачи [Вью13]: Пусть X — множество объектов , Y — конечное множество ответов (меток, имён классов) . <...>
Классовая принадлежность остальных объектов неизвестна.	Принадлежность остальных объектов к классам неизвестна. <...>
$* \arg \max_k k$	$Tk = \{P \mid P \in T, P = k\}$. <...>

5. Интеллектуальная информационная технология

<http://mydocx.ru/11-57155.html>

2016

Проверяемый текст	Текст источника
Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.	Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту. <...>
13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.	Классификация — один из разделов машинного обучения, посвященный решению следующей задачи [13]. <...>
Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.	Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1. <...>

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм (\cdot) : $a \in X \times Y$, способный классифицировать произвольный объект $x \in X$.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Классовая принадлежность остальных объектов неизвестна.

Классифицировать объект — значит, указать номер (или наименование класса), к которому относится данный объект.. <...>

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.. <...>

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

Классовая принадлежность остальных объектов не известна. <...>

6. Способы повышения эффективности классификации документов для конечного множества языков.

<http://elibrary.ru/item.asp?id=22791183>

2009, Зайцев В.Г., Лан Чуньлинь

Проверяемый текст

Для произвольного объекта u расположим объекты обучающей выборки $i \in X$ в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots; m$ и u и x и x , где i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

В наиболее общем виде алгоритм ближайших соседей выглядит так: $1 \in \arg \max [(\cdot) (\cdot) m i u y Y i a u y x u w i u, (0.3)]$ где $(\cdot) w i u$ — заданная весовая функция, которая оценивает степень важности i -го соседа для классификации объекта u .

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Классифицируемый объект x относится к тому классу $i \in Y$, которому принадлежит ближайший к нему объект обучающей выборки $i \in X$.

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

Чем больше значение этой функции, тем менее схожими являются два объекта x, x' .

Эта функция неотрицательна и не возрастает по i .

Текст источника

Для произвольного объекта u расположим объекты обучающей выборки $x \in X$ в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots; m$ и u и x и x , где i обозначает объект обучающей выборки, который является i -м соседом объекта u . <...>

$Y i = 1$ где $(\cdot) w i u$ — заданная весовая функция, которая оценивает степень важности i -го соседа для классификации объекта u . <...>

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам. <...>

Классифицируемый объект x относится к тому классу $i \in Y$, которому принадлежит ближайший объект обучающей выборки $i \in X$. <...>

Таким образом, произвольный объект u порождает свою перенумерацию выборки. <...>

Чем больше значение этой функции, тем менее схожими являются два объекта x, x' . <...>

Естественно полагать, что эта функция неотрицательна и не возрастает по i . <...>

7. Распознавание режимов работы авиационного ГТД с использованием технологии нейронных сетей

<http://referat.bookap.info/work/233952/Raspoznavanie-rezhimov-raboty-aviacionnogo>

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Текст источника

Классификация объекта -- номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.. <...>

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Пусть Y — конечное множество номеров (имён, меток) классов.

Классовая принадлежность остальных объектов неизвестна.

Классификация -- один из разделов машинного обучения, посвященный решению следующей задачи. <...>

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.. <...>

Классифицировать объект -- значит, указать номер (или наименование класса), к которому относится данный объект.. <...>

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

Пусть -- множество описаний объектов, -- конечное множество номеров (имён, меток) классов. <...>

Классовая принадлежность остальных объектов не известна. <...>

8. ВКР

<http://text.rucont.ru/History/Item?ServiceCallID=149860>

Муханова Дарья Дмитриевна

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $*: y \in X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$.

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм $(\cdot): x \in X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Существует неизвестная целевая зависимость $y = X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. <...>

Классификация - один из разделов машинного обучения, посвящённый решению следующей задачи. <...>

Классифицировать объект - значит, указать номер (или наименование класса), к которому относится данный объект. <...>

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества. <...>

Требуется построить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

Классовая принадлежность остальных объектов не известна. <...>

9. Задача классификации

<https://ru.wikipedia.org/wiki?curid=524472>

Проверяемый текст

Текст источника

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм $f: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Классовая принадлежность остальных объектов неизвестна.

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту. <...>

Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1. <...>

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) произвольный объект из исходного множества. <...>

Требуется построить алгоритм f , способный классифицировать произвольный объект x . <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Задача классификации — формализованная задача, в которой имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

Классовая принадлежность остальных объектов неизвестна. <...>

10. №3

<http://rucont.ru/efd/155416>

2011

Проверяемый текст

Для произвольного объекта u расположим объекты обучающей выборки $i \in X$ в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots$, где i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Классифицируемый объект x относится к тому классу y , которому принадлежит ближайший к нему объект обучающей выборки i .

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

* $\arg \max_k p_k$.

Текст источника

Для произвольного объекта $u \in X$ расположим объекты обучающей выборки x_1, x_2, \dots, x_s в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots$, где i обозначается тот объект обучающей выборки, который является i -м соседом объекта u . <...>

Поскольку в работе рассматривается задача классификации с двумя классами, то число соседей берётся нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежит разным классам. <...>

Классифицируемый объект относится к тому классу, которому принадлежит большинство из его соседей — к ближайших к нему из объектов обучающей выборки. <...>

Требуется построить алгоритм, способный классифицировать произвольный тестовый объект из исходного множества. <...>

Имеется исходное множество объектов, разделённых экспертом на классы (обучение с учителем). <...>

Таким образом, произвольный объект $u \in X$ порождает свою перенумерацию <...>

$P = \{p_k\}_{k=1}^K$ <...>

11. Алгоритмическое обеспечение ранней диагностики заболеваний статистическими методами классификации на примере биомаркерных иммуносигнатур: магистерская диссертация по направлению подготовки: 01.04.02 -

Прикладная математика и информатика

<http://vital.lib.tsu.ru/vital/access/services/Download/vital:6862/SOURCE01>

2018, Андрющенко, Владимир Сергеевич

Проверяемый текст

Метрическим классификатором (similarity-based classifier) называют алгоритм классификации, основанный на вычислении оценок сходства между объектами.

При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производиться разбиение.

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Она, в свою очередь, предполагает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Наивный байесовский классификатор ухудшил свои результаты.

Текст источника

2.2.2 Метрическая классификация Метрический классификатор — алгоритм классификации, основанный на вычислении оценок сходства между объектами. <...>

В ходе создания очередного узла дерева выбирается набор признаков, на основе которых производится разбиение. <...>

15 2 КЛАССИФИКАЦИЯ ДАННЫХ Классификация — один из основных разделов машинного обучения, посвященный решению следующей задачи. <...>

21 Метрические классификаторы опираются на гипотезу компактности, которая предполагает, что схожие объекты чаще лежат в одном классе, чем в разных. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Имеется множество объектов, разделённых некоторым образом на классы. <...>

Наивный байесовский классификатор [Электронный ресурс]. <...>

12. Алгоритмы преобразования и классификации трафика для обнаружения вторжений в компьютерные сети

<http://rsl.ru/rsl01005390227.txt>

2011, Большев, Александр Константинович

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $y: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $1, \dots, m$ $\{(x_1, y_1), \dots, (x_m, y_m)\}$.

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Требуется построить алгоритм $f: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Текст источника

Существует неизвестная целевая зависимость - отображение $y: X \rightarrow Y$, значения которой известны только на объектах обучающей выборки $X_t = \{(x_i, y_i), \dots, (x_m, y_m)\}$. <...>

3.1 Применение классификации с обучением на одном классе к обнаружению вторжений
Классификация - один из разделов машинного обучения, посвященный решению следующей задачи. <...>

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества. <...>

Требуется построить алгоритм $f: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$ [67]. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

13. Лекции по дискретной математике. Теория графов
http://biblioclub.ru/index.php?page=book_red&id=226800
2016, Зарипова Э. Р., Кокотчикова М. Г.

Проверяемый текст

Граф называется ориентированным (орграфом), если (V, E) — упорядоченная пара, где V — непустое множество вершин (узлов) графа G , а E — множество упорядоченных пар различных вершин, которые называются дугами.

Графом называется упорядоченная пара (V, E) , где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

2) Граф называется связным, если любые две его вершины связаны.

Степенью вершины называется количество инцидентных ей ребер.

Граф H называется подграфом графа G , если $V_H \subseteq V_G$ и $E_H \subseteq E_G$.

Обходя граф, мы двигаемся по ребрам и проходим все вершины.

Неориентированный граф Пусть $(v, w) \in E$ называется дугой.

1) Деревом называется связный граф без циклов.

Текст источника

Ориентированным графом (или орграфом) называется пара $G = (V, E)$, где V - непустое множество вершин, а $E \subseteq V \times V$ - множество дуг. <...>

Дан граф $G(V, E)$: Определить: 1) **Множество вершин V и множество ребер E .** <...>

Граф (без петель) называется полным, если любые две его вершины смежны. <...>

Нечетная степень вершины означает, что число ребер, инцидентных данной вершине, нечетно. <...>

Граф $G = (V, E)$ называется подграфом графа $G = (V, E)$ при соблюдении следующих условий: 1) $V \subseteq V$. <...>

Получаем, что мы посетили **все вершины** и прошли **все ребра графа.** <...>

Неориентированным графом (или просто графом) называется пара $G = (V, E)$, где $E \subseteq V \times V$. <...>

Дерево - связный граф, не содержащий циклов. <...>

14. Методика разработки программного продукта для поиска причин в изменениях трендов в данных
http://knowledge.allbest.ru/programming/2c0b65625b3bd79a5c43b88421316d27_0.html

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Наивный байесовский классификатор ухудшил свои результаты.

Текст источника

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту. <...>

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) произвольный объект из исходного множества. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

1.1 Задача классификации Задача классификации — формализованная задача, в которой **имеется множество объектов (ситуаций), разделённых некоторым образом на классы.** <...>

Наивный байесовский классификатор объединяет модель с правилом решения. <...>

Классовая принадлежность остальных объектов неизвестна.

Классовая принадлежность остальных объектов неизвестна. <...>

15. Реферат Объем работы 123 страницы 9 иллюстраций 4 приложения 1 таблица Ключевые слова алгоритм классифика

<http://samzan.ru/77429>

2016

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Наивный байесовский классификатор ухудшил свои результаты.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Классификация объекта ? номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.. <...>

Классифицировать объект ? значит, указать номер (или наименование) класса, к которому относится данный объект.. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) произвольный объект из исходного множества.. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Задача классификации ? формализованная задача, в которой **имеется множество объектов (ситуаций), разделённых некоторым образом на классы.** <...>

Наивный байесовский классификатор объединяет модель с правилом решения. <...>

Классовая принадлежность остальных объектов неизвестна. <...>

16. Методика разработки программного продукта для поиска причин в изменениях трендов в данных

<http://bibliofond.ru/view.aspx?id=785464>

2015

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Наивный байесовский классификатор ухудшил свои результаты.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Классификация объекта - номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.. <...>

Классифицировать объект - значит, указать номер (или наименование) класса, к которому относится данный объект.. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) произвольный объект из исходного множества.. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Задача классификации - формализованная задача, в которой **имеется множество объектов (ситуаций), разделённых некоторым образом на классы.** <...>

2.2 Наивный Байесовский Классификатор. <...>

Классовая принадлежность остальных объектов неизвестна. <...>

17. Методика разработки программного продукта для поиска причин в изменениях трендов в данных

<http://referat.bookap.info/work/172718/Metodika-razrabotki-programmnogo-produkta>

Проверяемый текст

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Наивный байесовский классификатор ухудшил свои результаты.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Классификация объекта -- номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.. <...>

Классифицировать объект -- значит, указать номер (или наименование) класса, к которому относится данный объект.. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) произвольный объект из исходного множества.. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Задача классификации -- формализованная задача, в которой имеется множество объектов (ситуаций), разделённых некоторым образом на классы. <...>

2.2 Наивный Байесовский Классификатор. <...>

Классовая принадлежность остальных объектов неизвестна. <...>

18. Выбор метрики расстояния для k-NN классификатора

<http://diplomba.ru/work/24309>

Проверяемый текст

Для произвольного объекта u расположим объекты обучающей выборки ix в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots; m$ и u x_1, x_2, \dots, x_m , где i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

Чтобы формализовать понятие сходства вводится функция расстояния между объектами (x, y) в пространстве объектов X .

Она, в свою очередь, предполагает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

Пусть Y — конечное множество номеров (имён, меток) классов.

Эта функция неотрицательна и не возрастает по i .

Текст источника

X расположим элементы обучающей выборки x_1, \dots, x_m в порядке возрастания расстояний до u : $?(u, x_1, u) ? ?(u, x_2, u) ? \dots ? ?(u, x_i, u)(2.1)$. где через x_i обозначается i -й сосед объекта u . <...>

Для формализации понятия «сходства» вводится функция расстояния или метрика $?(x, y)$ в пространстве объектов X . <...>

В задачах классификации это означает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных. <...>

X порождает свою перенумерацию выборки x_1, u, \dots, x_m, u . <...>

Имеется пространство объектов X и конечное множество имён классов Y . <...>

Естественно полагать, что эта функция неотрицательна и не возрастает по i . <...>

19. Выбор метрики расстояния для k-NN классификатора

http://knowledge.allbest.ru/emodel/2c0a65635a3bd79b5c53a88521206c27_0.html

Проверяемый текст

Текст источника

Для произвольного объекта u расположим объекты обучающей выборки x_i в порядке возрастания расстояний до u : $1; 2; \dots; (u, x_1, u) \dots (u, x_m, u) \dots (u, x_n, u)$, где x_i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

Чтобы формализовать понятие сходства вводится функция расстояния между объектами (x, y) в пространстве объектов X .

Она, в свою очередь, предполагает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

Пусть Y — конечное множество номеров (имён, меток) классов.

Эта функция неотрицательна и не возрастает по i .

2.1 Метод ближайших **соседей** и его обобщения Для произвольного объекта $u \in X$ расположим элементы обучающей выборки x_1, \dots, x_n в порядке возрастания расстояний до u : $d(u, x_1, u) \leq d(u, x_2, u) \leq \dots \leq d(u, x_n, u)$ (2.1) где через x_i, u обозначается i -й сосед объекта u . $\langle \dots \rangle$

Для формализации понятия «сходства» вводится функция расстояния или метрика $d(x, y)$ в пространстве объектов X . $\langle \dots \rangle$

В задачах классификации это означает, что **схожие объекты гораздо чаще лежат в одном классе**, чем в разных. $\langle \dots \rangle$

Каждый объект $u \in X$ порождает свою перенумерацию выборки x_1, u, \dots, x_n, u . $\langle \dots \rangle$

Имеется пространство объектов X и **конечное множество имён классов Y** . $\langle \dots \rangle$

Естественно полагать, что **эта функция неотрицательна и не возрастает по i** . $\langle \dots \rangle$

20. Дискретная математика

http://biblioclub.ru/index.php?page=book_red&id=68366

2016, Макоха А. Н., Сахнюк П. А., Червяков Н. И.

Проверяемый текст

Такое название объясняется тем, что в процессе поиска мы идем **вширь**, а не **вглубь**, т.е. сначала просматриваем все соседние вершины, затем соседей соседей и так далее.

Алгоритм поиска в ширину перечисляет все достижимые из s вершины, доступные при проходе по ребрам, в порядке возрастания расстояния от s .

Расстоянием считается длина минимального пути из начальной вершины.

Пусть задан граф (V, E) и фиксирована начальная вершина s .

Степенью вершины называется количество инцидентных ей ребер.

Существует много несколько разновидностей графов.

Пример такого графа приведен на рис.

Текст источника

Название алгоритма объясняется тем, что в процессе поиска мы продвигаемся по графу **вширь**, а не **вглубь** (сначала просматриваются все соседние вершины, затем соседи соседей и т.д.) $\langle \dots \rangle$

Алгоритм поиска в ширину перечисляет все достижимые из вершины v_0 (если идти по ребрам) **вершины в порядке возрастания расстояния от этой вершины** (расстоянием считается длина, т.е. количество **ребер** кратчайшего пути). $\langle \dots \rangle$

Здесь под метками подразумевается некоторое **расстояние (длина пути) от начальной вершины s до других вершин X** . $\langle \dots \rangle$

1.49 **задан ориентированный граф $G = (V, E)$ с начальной вершиной $s = 1$** . $\langle \dots \rangle$

Степенью вершины называется число дуг, инцидентных ей. $\langle \dots \rangle$

Существует несколько способов такого рода представления графов. $\langle \dots \rangle$

Примеры разных частей графа приведены на рис. $\langle \dots \rangle$

21. Классификация зашумленных последовательностей, порожденных близкими скрытыми марковскими моделями.

<http://elibrary.ru/item.asp?id=17013433>

2011, Гулятьева Т.А., Попов А.А.

Проверяемый текст

Текст источника

Для произвольного объекта u расположим объекты обучающей выборки ix в порядке возрастания расстояний до u : $1; 2; \dots; (i) u m u x u x u x$, где i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

X расположим объекты обучающей выборки x_1, x_2, \dots, x_s в порядке возрастания расстояний до u : $(1) (2) (s) u u ? u ? (u, x) ? ? (u, x) ? ? (u, x)$, где через $i (i)$ $u x$ обозначается тот объект обучающей выборки, который является i -м соседом объекта u . <...>

Поскольку в работе рассматривается задача классификации с двумя классами, то число соседей берётся нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежит разным классам. <...>

Требуется построить алгоритм, способный классифицировать произвольный тестовый объект из исходного множества. <...>

Имеется исходное множество объектов, разделённых экспертом на классы (обучение с учителем). <...>

22. Проектирование и разработка системы обратной связи с клиентом и сотрудником с применением средств машинного обучения

<http://text.rucont.ru/History/Item?ServiceCallID=106155>

Абдуллин

Проверяемый текст

При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производиться разбиение.

25 1.2.4 Случайный лес В основе алгоритма случайный лес лежит идея использования ансамбля (комитета) решающих деревьев.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Генерируется случайная подвыборка с возвращением размером n .

Пусть размерность пространства признаков будет равна M .

Текст источника

Причём в ходе создания очередного узла дерева выбирается признак, на основе которого производится разбиение, не из всех M признаков, а лишь из m случайно выбранных. <...>

Случайный лес (Random Forest) – алгоритм машинного обучения, использует ансамбль (или комитет) решающих деревьев. <...>

Требуется построить алгоритм, способный классифицировать любой объект из исходного множества. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Генерируется случайную подвыборку с повторением размером N из обучающей выборки. <...>

M – размерность пространства признаков; m – параметр, число случайно выбираемых признаков для дерева. <...>

23. Проектирование и разработка системы обратной связи с клиентом и сотрудником с применением средств машинного обучения

<http://text.rucont.ru/History/Item?ServiceCallID=106316>

Абдуллин

Проверяемый текст

При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производиться разбиение.

Текст источника

Причём в ходе создания очередного узла дерева выбирается признак, на основе которого производится разбиение, не из всех M признаков, а лишь из m случайно выбранных. <...>

25 1.2.4 Случайный лес В основе алгоритма случайный лес лежит идея использования ансамбля (комитета) решающих деревьев.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Генерируется случайная подвыборка с возвращением размером n .

Пусть размерность пространства признаков будет равна M .

Случайный лес (Random Forest) – алгоритм машинного обучения, использует ансамбль (или комитет) решающих деревьев. <...>

Требуется построить алгоритм, способный классифицировать любой объект из исходного множества[7]. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Генерируется случайную подвыборку с повторением размером N из обучающей выборки. <...>

. M – размерность пространства признаков ; n – параметр, число случайно выбираемых признаков для дерева.. <...>

24. сырые исходные данные не годятся в качестве признаков описаний но в то же время существуют эффективные

<http://samzan.ru/32165>

2016

Проверяемый текст

Для произвольного объекта u расположим объекты обучающей выборки ix в порядке возрастания расстояний до u : $1; 2; \dots; i; \dots; m$ и u в x и x , где i обозначает объект обучающей выборки, который является i -ым соседом объекта u .

Чтобы формализовать понятие сходства вводится функция расстояния между объектами (x, x) в пространстве объектов X .

Она, в свою очередь, предполагает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

Эта функция неотрицательна и не возрастает по i .

Текст источника

Для произвольного объекта $u \in X$ расположим элементы обучающей выборки x_1, \dots, x_m в порядке возрастания расстояний до u : $?(u, x_1, u) ?(u, x_2, u) ? \dots ?(u, x_i, u) (2.1)$. где через x_i обозначается i -й сосед объекта u . <...>

Для формализации понятия «сходства» вводится функция расстояния или метрика $?(x, x?)$ в пространстве объектов X . <...>

В задачах классификации это означает, что схожие объекты гораздо чаще лежат в одном классе, чем в разных. <...>

Каждый объект $u \in X$ порождает свою перенумерацию выборки x_1, u, \dots, x_m, u . <...>

Естественно полагать, что эта функция неотрицательна и не возрастает по i . <...>

25. Дискретная математика. Теория и практика решения задач по информатике учеб. пособие

<http://rucont.ru/efd/443387>

2015

Проверяемый текст

Графом называется упорядоченная пара (V, E) $G = (V, E)$, где V – непустое множество вершин (узлов) графа G , а E – множество ребер графа G .

Пусть задан граф (V, E) и фиксирована начальная вершина s .

Степенью вершины называется количество инцидентных ей ребер.

2) Граф называется связным, если любые две его вершины связаны.

Текст источника

Терминология **Граф $G = (V, E)$** состоит из двух конечных множеств — множества вершин V и множества ребер E . <...>

Вывод пути Дан ориентированный граф $G = (V, E)$, начальная и конечная вершины — $s, t \in V$. <...>

Степенью вершины графа G называется число инцидентных ей ребер и обозначается d_i или $\deg i$. <...>

Понятие **графа**, основные методы просмотра **вершин** графа **любая** из его **двух** вершин называются инцидентными. <...>

Граф H называется подграфом графа G , если $(V(H), E(H)) \subseteq (V(G), E(G))$.

Неориентированный граф Пусть (V, E) называется дугой.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

* $\arg\max_k k$ p .

Граф H называется реконструкцией графа G , если $P(G) = P(H)$. <...>

Графовый матроид Пусть $G = (V, E)$ — **неориентированный граф**. <...>

Дерево — это **связный** ациклический **граф**. <...>

Пример эйлера графа приведен на рис. <...>

$P(1, 1, \dots, 1, k-1) = (1!)^{n-k} (k-1)! <...>$

26. Дискретная математика. Теория и практика решения задач по информатике

http://biblioclub.ru/index.php?page=book_red&id=120543

2016, Окулов С. М.

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

Пусть задан граф (V, E) и фиксирована начальная вершина s .

Степенью вершины называется количество инцидентных ей ребер.

2) Граф называется связным, если любые две его вершины связаны.

Граф H называется подграфом графа G , если $(V(H), E(H)) \subseteq (V(G), E(G))$.

Неориентированный граф Пусть (V, E) называется дугой.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

Текст источника

Терминология **Граф $G = (V, E)$** состоит из двух конечных множеств — **множества вершин V** и **множества ребер E** . <...>

Вывод пути Дан ориентированный **граф $G = (V, E)$** , **начальная** и конечная **вершины** — s, t ? <...>

Степенью вершины графа G называется число **инцидентных ей ребер** и обозначается d_i или $\deg i$. <...>

Понятие **графа**, основные методы просмотра **вершин** графа **любая** из его **двух вершин** называются инцидентными. <...>

Граф H называется реконструкцией графа G , если $P(G) = P(H)$. <...>

Графовый матроид Пусть $G = (V, E)$ — **неориентированный граф**. <...>

Дерево — это **связный** ациклический **граф**. <...>

Пример эйлера графа приведен на рис. <...>

27. Применение методов машинного обучения и анализа данных для решения проблем автоматизации сбора и обработки информации

http://statref.ru/ref_jgeqasnapolmer.html

Проверяемый текст

13 Классификаторы Классификацией называют один из разделов машинного обучения, посвященный решению следующей задачи.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Текст источника

. **Классификация** — один из разделов машинного обучения, посвященный решению следующей задачи. <...>

Требуется построить алгоритм, способный классифицировать (см. ниже) **произвольный объект из исходного множества**. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

. **Задача классификации** — формализованная задача, в которой **имеется множество объектов (ситуаций), разделённых некоторым образом на классы**. <...>

Останов в случае, если все объекты в листе относятся к одному классу.

Классовая принадлежность остальных объектов неизвестна.

Самый простой критерий **останова** проверяет, **все ли объекты в вершине относятся к одному классу**. <...>

Классовая принадлежность остальных объектов неизвестна. <...>

28. Технология построения обобщенного "И/ИЛИ" дерева решения задач

<http://www.dslib.net/teor-informatika/tehnologija-postroenija-obobwennogo-i-ili-dereva-reshenija-zadach.html>

2008, Вовк Алексей Андреевич

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $*: u \rightarrow X \cup Y$, значения которой известны только на объектах конечной обучающей выборки $1, 1, \{(x_1, y_1), \dots, (x_m, y_m)\} \subset X \cup Y$.

Чтобы формализовать понятие сходства вводится функция расстояния между объектами $(x, y) \in X \times Y$ в пространстве объектов X .

Пусть задана обучающая выборка $1, 1, \{(x_1, y_1), \dots, (x_m, y_m)\} \subset X \cup Y$ и на множестве объектов задана функция расстояния $d(x, y)$.

Вершина, из которой выходят ребра называется родительской.

Текст источника

Существует неизвестная целевая зависимость $u : X \cup Y$, значения которой известны только на объектах обучающей выборки $X = \{(x_i, y_i) | i=1, \dots, m, y_i = u(x_i)\}$. <...>

Для формализации понятия сходства вводится функция расстояния или метрика $d(x, y)$ в пространстве объектов X . <...>

Задача кластеризации заключается в следующем: **Задана обучающая выборка $X = \{x_1, \dots, x_m\}$ и функция расстояния между объектами $d(x, y)$.** <...>

Если из **вершины** выходят только **ребра**, связанные отношением «И», то такая **вершина называется «И» вершиной**.. <...>

29. Применение методов машинного обучения в судебно-экспертной деятельности для целей классификации и идентификации: обзор подходов и их адаптация.

<http://elibrary.ru/item.asp?id=35018083>

2018, Боброва Н.Л., Чашин С.В.

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $*: u \rightarrow X \cup Y$, значения которой известны только на объектах конечной обучающей выборки $1, 1, \{(x_1, y_1), \dots, (x_m, y_m)\} \subset X \cup Y$.

Классифицируемый объект x относится к тому классу $i \in Y$, которому принадлежит ближайший к нему объект обучающей выборки $i \in X$.

Требуется построить алгоритм $(*) : x \rightarrow X \cup Y$, способный классифицировать произвольный объект $x \in X$.

Пусть Y — конечное множество номеров (имён, меток) классов.

Текст источника

Существует неизвестная целевая зависимость - отображение $u^*: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. <...>

В его основе лежит гипотеза о компактности: **классифицируемый объект относится к тому классу, к которому принадлежат ближайшие к нему объекты обучающей выборки.** <...>

$X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$? <...>

Пусть X - **множество** описаний объектов, Y - **конечное множество номеров (имен, меток) классов**. <...>

30. Основные алгоритмы на графах текст лекций

<http://rucont.ru/efd/237872>

2011

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Текст источника

Элементы **множества V называются вершинами графа G** , а элементы **множества E – его ребрами**. <...>

Алгоритм применим как к ориентированным графам, так и к неориентированным.

Пусть задан граф (V, E) и фиксирована начальная вершина s .

2) Граф называется связным, если любые две его вершины связаны.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

Алгоритм работает как для **ориентированных**, так и для **неориентированных графов**. <...>

Поиск в ширину Пусть задан граф $G = (V, E)$ и выделена **начальная вершина** <...>

Граф называется связным, если существует путь между **любыми двумя его вершинами**. <...>

Связный граф, в котором есть эйлеров **цикл**, называется **эйлеровым графом**. <...>

22 Рассмотрим их на **примере графа**, изображенного на **рис**. <...>

31. 1] Определение глубины дерева [0

<http://samzan.ru/24094>

2016

Проверяемый текст

Граф называется ориентированным (орграфом), если (V, E) — упорядоченная пара, где V — непустое множество вершин (узлов) графа G , а E — множество упорядоченных пар различных вершин, которые называются дугами.

Графом называется упорядоченная пара (V, E) , где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

Текст источника

Графом G называется пара множеств (V, E) , где V — конечное множество элементов, называемых вершинами графа, а E — конечное множество упорядоченных пар $e = (w, v)$, называемых дугами, где w, v — вершины. <...>

Графом G называется пара множеств (V, E) , где V — конечное множество элементов, называемых вершинами графа, а E — конечное множество упорядоченных пар $e = (w, v)$, называемых дугами, где w, v — вершины. <...>

Говорят, что **дуга e выходит из вершины w и входит в вершину v** . <...>

32. Искусство алгоритмизации

http://biblioclub.ru/index.php?page=book_red&id=130059

2016, Потопахин В. В.

Проверяемый текст

Граф называется ориентированным (орграфом), если (V, E) — упорядоченная пара, где V — непустое множество вершин (узлов) графа G , а E — множество упорядоченных пар различных вершин, которые называются дугами.

Графом называется упорядоченная пара (V, E) , где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

1) Деревом называется связный граф без циклов.

Текст источника

Ориентированным графом (орграфом) $G=(X, Y)$ называется упорядоченная пара (X, Y) , где X — непустое множество вершин орграфа (ориентированного графа), а Y — множество упорядоченных пар элементов из X , называемых дугами орграфа. <...>

Основные определения Неориентированным **графом**, или просто **графом**, $G = (X, U)$ называется **упорядоченная пара (X, U) , где X — непустое множество вершин графа, а U — множество неупорядоченных пар элементов из X , называемых ребрами графа.** <...>

Деревом называется связный граф с одним входом, в котором нет ни одного контура. <...>

33. Искусство алгоритмизации [учеб. пособие]

<http://rucont.ru/efd/199490>

2011

Проверяемый текст

Текст источника

Граф называется ориентированным (орграфом), если (V, E) — упорядоченная пара, где V — непустое множество вершин (узлов) графа G , а E — множество упорядоченных пар различных вершин, которые называются дугами.

Графом называется упорядоченная пара (V, E) , где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

1) Деревом называется связный граф без циклов.

Ориентированным графом (орграфом) $G=(X, Y)$ называется упорядоченная пара (X, Y) , где X — непустое множество вершин орграфа (ориентированного графа), а Y — множество упорядоченных пар элементов из X , называемых дугами орграфа. <...>

Основные определения Неориентированным графом, или просто графом, $G = (X, U)$ называется упорядоченная пара (X, U) , где X — непустое множество вершин графа, а U — множество неупорядоченных пар элементов из X , называемых ребрами графа. <...>

Деревом называется связный граф с одним входом, в котором нет ни одного контура. <...>

34. Интеллектуальная поддержка принятия решений с использованием интерпретируемых нечетких моделей в системах энергетического менеджмента

<http://rsl.ru/rsl01006659362.txt>

2016, Аль-Гунаид Мохаммед Амин Мохаммед

Проверяемый текст

Объект x доходит до вершины v тогда и только тогда, когда выполняется конъюнкция $\bigwedge K_x$, составленная из всех предикатов, приписанных внутренним вершинам дерева на пути от корня 0_v до вершины v .

Множества объектов $\{x \in X : \bigwedge_{v \in T} v(x) = 1\}$, выделяемых терминальными конъюнкциями $\bigwedge_{v \in T} v$, попарно не пересекаются, а их объединение совпадает со всем пространством X .

Пусть T — множество всех терминальных вершин дерева.

Текст источника

Объект X доходит до вершины V лишь в том случае, когда выполняется конъюнкция $K^V(x)$, составленная из всех предикатов, приписанных внутренним вершинам дерева на пути от корня 0_v до вершины V . <...>

Множества объектов $\Gamma_2 = \{x \in X : K^V(x) = 1\}$, выделяемых терминальными конъюнкциями $\bigwedge_{v \in T} v$, попарно не пересекаются, а их объединение совпадает со всем пространством X . <...>

Пусть T - множество всех терминальных вершин дерева. <...>

35. D_201

<http://text.rucont.ru/History/Item?ServiceCallID=199636>

Abdullin

Проверяемый текст

При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производится разбиение.

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Генерируется случайная подвыборка с возвращением размером n .

Текст источника

Причём в ходе создания очередного узла дерева выбирается признак, на основе которого производится разбиение, не из всех M признаков, а лишь из m случайно выбранных. <...>

Требуется построить алгоритм, способный классифицировать любой объект из исходного множества. <...>

Задано конечное множество объектов, для которых известно, к каким классам они относятся. <...>

Генерируется случайную подвыборку с повторением размером N из обучающей выборки. <...>

36. Методика обучения функциональному программированию будущих учителей информатики

<http://rsl.ru/rsl01000272789.txt>

2000, Головлева, Светлана Викторовна

Проверяемый текст

2) Граф называется связным, если любые две его вершины связаны.

Степенью вершины называется количество инцидентных ей ребер.

Неориентированный граф Пусть (V, E) называется дугой.

Существует много несколько разновидностей графов.

Порядок обхода графа при поиске в глубину На рис.

1) Деревом называется связный граф без циклов.

Порядок обхода графа при поиске в ширину На рис.

Текст источника

Ориенти рованный **граф называется** сильно **связным**, если сильно связны все его **вершины**. <...>

Степень вершины графа равна числу **ребер**, **инцидентных ей**. <...>

Пара (V, E) , где E - произвольное подмножество V^2 , **называется неориентированным графом**. <...>

Определите, **существуют ли** в заданном **графе** циклы? <...>

Перензомеруйте вершины неориентированного **графа в порядке обхода в глубину**. <...>

Деревом называется связный граф без циклов. <...>

Перензомеруйте вершины неориентированного **графа в порядке обхода в глубину**. <...>

37. КОМПАРИРОВАНИЕ МЕТОДОВ КЛАССИФИКАЦИИ В ЗАДАЧАХ МЕДИЦИНСКОГО ДИАГНОСТИРОВАНИЯ.

<http://elibrary.ru/item.asp?id=23604152>

2015, Артиков М.Э., Абдуллаев Ф.О., Хужаев О.К., Худайбергенов Т.А.

Проверяемый текст

25 1.2.4 Случайный лес В основе алгоритма случайный лес лежит идея использования ансамбля (комитета) решающих деревьев.

Классифицируемый объект x относится к тому классу i_y , которому принадлежит ближайший к нему объект обучающей выборки i_x .

16 1.2.2 Метод k-ближайших соседей Алгоритмы, основанные на анализе сходства объектов, часто называют метрическими.

Текст источника

Random forest (**случайный лес**) – алгоритм машинного обучения, заключающийся в **использовании ансамбля решающих деревьев**. <...>

Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки [5]. <...>

К-ближайших соседей – это метрический алгоритм классификации, **основанный** на оценивании **сходства объектов**. <...>

38. Анализ эффективности применения методов классификации.

<http://elibrary.ru/item.asp?id=29771103>

2017, Артиков М.Э.

Проверяемый текст

25 1.2.4 Случайный лес В основе алгоритма случайный лес лежит идея использования ансамбля (комитета) решающих деревьев.

Классифицируемый объект x относится к тому классу i_y , которому принадлежит ближайший к нему объект обучающей выборки i_x .

16 1.2.2 Метод k-ближайших соседей Алгоритмы, основанные на анализе сходства объектов, часто называют метрическими.

Текст источника

Random forest (**случайный лес**) — алгоритм машинного обучения, заключающийся в **использовании ансамбля решающих деревьев**. <...>

Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки [6]. <...>

К-ближайших соседей – это метрический алгоритм классификации, **основанный** на оценивании **сходства объектов**. <...>

39. Программирование в алгоритмах [учебник]

<http://rucont.ru/efd/443308>

2014

Проверяемый текст

Текст источника

Графом называется упорядоченная пара (V, E) , где V – непустое множество вершин (узлов) графа G , а E – множество ребер графа G .

Например, довольно часто не выполняется неравенство треугольника.

Неориентированный граф Пусть $(v, w) \in E$ называется дугой.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

* $\arg\max_k k \cdot p_k$.

Граф G называют двудольным, если его **множество вершин** можно разбить на непересекающиеся **множества** — $V = X \cup Y$, $X \cap Y = \emptyset$, причем каждое **ребро** $e \in E$ имеет вид $e = (x, y)$, где $x \in X$, $y \in Y$. <...>

Неравенство треугольника выполняется. <...>

Независимые множества Дан **неориентированный граф** $G = (V, E)$. <...>

Стягивающие **деревья** **Деревом** называют произвольный **связный** неориентированный **граф без циклов**. <...>

На **рисунке 4.39** **приведены примеры графов** и их раскраски. <...>

$\{j \mid j \in P[k]\}$. <...>

40. <http://textarchive.ru/c-1053217-pall.html>

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V – непустое множество вершин (узлов) графа G , а E – множество ребер графа G .

2) Граф называется связным, если любые две его вершины связаны.

Неориентированный граф Пусть $(v, w) \in E$ называется дугой.

Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

1) Деревом называется связный граф без циклов.

Текст источника

. Представления **графов** Из определения графа следует, что каждый **граф** $G = (V, E)$ можно задать, непосредственно перечислив его **множество вершин V** и **множество ребер E** . <...>

Неориентированный **граф называется связным**, если **любая** пара **вершин** в нем соединена путем. <...>

Пусть $G = (V, E)$ **неориентированный граф** с $|E| < |V| - 1$. <...>

Если из **вершины v** **ведет** путь в **вершину w** , то v называется предком w , а w – потомком v . <...>

Неориентированный **граф называется деревом**, если он **связный** и в нем нет **циклов**. <...>

41. Теория графов: Учебно-методическое пособие

http://window.edu.ru/resource/211/79211/files/alekseev_zakharova.pdf

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V – непустое множество вершин (узлов) графа G , а E – множество ребер графа G .

2) Граф называется связным, если любые две его вершины связаны.

Граф H называется подграфом графа G , если $(V(H), E(H)) \subseteq (V(G), E(G))$.

1) Деревом называется связный граф без циклов.

Текст источника

Дополнением (дополнительным **графом**) к **графу** $G = (V, E)$ **называется граф** $G' = (V, E')$, у которого **множество вершин** то же, что у G , а **множество ребер** является дополнением **множества E** до **множества** всех неупорядоченных пар **вершин**. <...>

Граф называется связным, если для **любых двух его вершин** имеется путь, соединяющий эти **вершины**. <...>

Подграфы **Граф** $G' = (V', E')$ **называется подграфом графа** $G = (V, E)$, если $V' \subseteq V$, $E' \subseteq E$. <...>

Важнейшие классы **графов** **Деревья** **Деревом** **называется связный граф**, не имеющий **циклов**. <...>

42. Построение и исследование полных решающих деревьев для задач классификации по прецедентам

<http://rsl.ru/rsl01006702580.txt>

2013, Генрихов, Игорь Евгеньевич

Проверяемый текст

Задано конечное множество объектов, для которых известно, к каким классам они относятся.

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Степенью вершины называется количество инцидентных ей ребер.

Вершина, из которой выходят ребра называется родительской.

Текст источника

Имеется **конечное множество объектов** (прецедентов или обучающих **объектов**) $\{^1, \dots, ^n\}$ из M , о **которых известно каким классам они принадлежат**. <...>

Множество объектов разделено на три **класса** (три вида волн). <...>

От **вершины** a_0 построим все **ребра**, которые **инцидентны ей**. <...>

Ребром в ПРД называется дуга, **выходящая** из обычной или из полной **вершины** дерева. <...>

43. Алгоритмы и структуры данных

http://biblioclub.ru/index.php?page=book_red&id=225965

2016, Гагарина Л. Г., Колдаев В. Д.

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Вершина, из которой выходят ребра называется родительской.

Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

1) Деревом называется связный граф без циклов.

Текст источника

Пусть $G = (V, E)$ — связный неориентированный **граф**, содержащий циклы, т.е. замкнутые маршруты, где V — **множество вершин**, а E — **множество ребер**. <...>

Дерево, из каждой **вершины** которого **выходит** только по два **ребра**, называется бинарным (рис. <...>

Пусть (V, W) — ребро с наименьшей стоимостью, **ведущее** из U в любую «невывбранную» **вершину** W . <...>

Связный граф без циклов называется деревом. <...>

44. СЛУЧАЙНЫЕ ЛЕСА: ОБЗОР.

<http://elibrary.ru/item.asp?id=18966671>

2013, Чистяков С.П.

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

Пусть T — множество всех терминальных вершин дерева.

1) Деревом называется связный граф без циклов.

Текст источника

Граф $G = (V, E)$ состоит из конечного **непустого множества** V , элементы которого **называются вершинами** и **множества пар вершин** E , называемых **ребрами**. <...>

Если (v, w) — некоторая **дуга**, то **вершина** v называется родителем w , а **вершина** w — **потомком вершины** v . <...>

Все такие **вершины** объявляются **терминальными**, а соответствующее **дерево** полным. <...>

Дерево представляет собой **связный граф без циклов**. <...>

45. Методики выявления потенциально вредоносных файловых объектов на основе интеллектуального анализа данных

<http://rsl.ru/rsl01006751577.txt>

2013, Комашинский, Дмитрий Владимирович

Проверяемый текст

Текст источника

Существует неизвестная целевая зависимость — отображение $y: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X_1 = \{x_1, \dots, x_n\}$.

Зададим обучающую выборку, количество образцов в которой равно N .

Пусть размерность пространства признаков будет равна M .

Существует неизвестная целевая зависимость - отображение $y: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X_1 = \{x_1, \dots, x_n\}$.

Пусть задана обучающая выборка размера n , а количество атрибутов равно t . <...>

Пространство признаков - используемый диапазон смещений и размерность. <...>

46. Сравнительное исследование алгоритмов классификации больших объемов данных

<http://ssau.ru/files/science/conferences/itnt2016/ds/1096-1099.pdf>

2016, Н.В. Ситникова, Р.А. Парингер, А.В. Куприянов

Проверяемый текст

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Классифицируемый объект x относится к тому классу y_i , которому принадлежит ближайший к нему объект обучающей выборки x_i .

Текст источника

ций неоднозначности, когда одинаковое число соседей принадлежат разным классам. <...>

Классифицируемый объект x относится к тому классу y_i , которому принадлежит первый ближайший объект обучающей выборки x_i . <...>

47. Методы и алгоритмы обработки информации в информационно-аналитических системах для анализа развития событий кризисных ситуаций

http://www.rsreu.ru/ru/component/docman/doc_download/10530-sokolova-yu-s-metody-i-algoritmy-analiza-dannykh-na-osnove-instrumentariya-intellektualnoj-obrabotki-informatsii-i-bioinspirirovannogo-modelirovaniya

Щербинин В.Н.

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

В результате была получена информация, приведенная в таблице 2.2.

Классовая принадлежность остальных объектов неизвестна.

Текст источника

Пусть O – матрица стоимости графа $G = (V, E)$ со множеством вершин V и множеством ребер E . <...>

В таблице 4.7 приведены результаты классификации, полученные для всех 18 моделей. <...>

Объекты набора V с неизвестной классовой принадлежностью помечены маркером «треугольник». <...>

48. Элементы теории графов: Учебное пособие

<http://window.edu.ru/resource/360/60360/files/dln001.pdf>

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Степенью вершины называется количество инцидентных ей ребер.

1) Деревом называется связный граф без циклов.

Текст источника

При записи различных соотношений в теории графов пользуются обозначениями V или $V(G)$ для множества вершин и E или $E(G)$ для множества ребер графа G . <...>

Число ребер, инцидентных вершине v , определяет степень вершины, которая обозначается $\deg v$. <...>

Приведем некоторые из них: 43 1) Дерево это связный граф без циклов. <...>

49. Основы дискретной математики

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Степенью вершины называется количество инцидентных ей ребер.

1) Деревом называется связный граф без циклов.

Текст источника

. Тогда **графом $G(V, A)$ называется совокупность множества вершин и множества ребер.** <...>

. **Вершина, не инцидентная никакому ребру, называется изолированной.** <...>

Деревом называется связный граф, не содержащий циклов. <...>

50. Структуры данных. Ч. II. Нелинейные динамические структуры [учеб. пособие]

<http://rucont.ru/efd/176507>

2007

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Граф H называется подграфом графа G , если $V_H \subseteq V_G$ и $E_H \subseteq E_G$.

Пример такого графа приведен на рис.

Текст источника

Основные понятия и определения **Граф $G = (V, E)$ включает конечное множество вершин V и конечное множество ребер E , соединяющих эти вершины.** <...>

Граф $G' = (V', E')$ является подграфом графа $G = (V, E)$, если $V' \subseteq V, E' \subseteq E$. <...>

Примеры неориентированных и ориентированных графов приведены на рис. <...>

51. Модели оценивания и алгоритмы управления качеством программных средств

<http://rsl.ru/rsl01005028880.txt>

2010, Бураков, Вадим Витальевич

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Граф H называется подграфом графа G , если $V_H \subseteq V_G$ и $E_H \subseteq E_G$.

Пример такого графа приведен на рис.

Текст источника

Ориентированный **граф $G = (V, E, s, t)$ состоит из двух множеств, конечного множества V , элементы которого называются вершинами, и конечного множества E , элементы которого называются ребрами.** <...>

Подграф L графа G называется вхождением L в G или совпадением для продукции p . <...>

Пример абстрактного помеченного типизированного графа. <...>

52. <http://nauchkor.ru/uploads/documents/587d363d5f1be77c40d58a4d.pdf>

Проверяемый текст

Чтобы формализовать понятие сходства вводится функция расстояния между объектами (x, y) в пространстве объектов X .

Можно сказать, что классы образуют компактно локализованные подмножества в пространстве объектов.

* $\arg\max_k k$ p .

Текст источника

Для того, чтобы **формализовать понятие «сходства» вводится функция расстояния в пространстве объектов X .** <...>

Если формализовать это высказывание: **классы образуют компактно локализованные подмножества.** <...>

($\arg\max_y y$ j=1 K <...>

53. Исследование методов установления значений синтаксических единиц естественных языков на основе

Проверяемый текст

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Пусть Y — конечное множество номеров (имён, меток) классов.

Текст источника

В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам, если число классов больше двух, то применяется взвешивание соседей. <...>

Наивный Байесовский классификатор Пусть задано множество объектов X и конечное множество имён классов (меток) Y . <...>

54. Лекции по дискретной математике. Учебное пособие

<http://www.bibliorossica.com/book.html?currBookId=12141>

2007, Дехтярь М. И.

Проверяемый текст

Граф H называется подграфом графа G , если $(V_H, E_H) \subseteq (V_G, E_G)$.

Неориентированный граф Пусть $(v, w) \in E$ называется дугой.

Можно сказать, что дуга v, w ведёт от вершины v к вершине w .

1) Деревом называется связный граф без циклов.

Текст источника

Граф $G = (V, E)$ называется подграфом графа $G = (V, E)$, если $V_H \subseteq V_G$ и $E_H \subseteq E_G$. <...>

Пусть $G = (V, E)$ — неориентированный граф. <...>

Если из вершины v ведёт путь в вершину w , то v называется предком w , а w — потомком v . <...>

Неориентированный граф называется связным, если он связный и в нём нет циклов. граф называется деревом, Определение 10.2. <...>

55. Практикум по курсу "Алгоритмизация и программирование". Часть 2: Учебное пособие

<http://window.edu.ru/resource/952/79952/files/Programming2.pdf>

Проверяемый текст

2) Граф называется связным, если любые две его вершины связаны.

Степенью вершины называется количество инцидентных ей ребер.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

Текст источника

Граф называется полным, если любые две его вершины соединены ребром. <...>

Степенью вершины — это число ребер, инцидентных ей, например, вершина «1» имеет степень 2. <...>

Деревом называется связный граф без циклов. <...>

Приведем пример матрицы инцидентности для графа, изображенного на рис. <...>

56. Программирование на языке Пролог для искусственного интеллекта (fb2)

<http://lib.rus.ec/b/372518/read>

Степанов, ...); Иван Братко

Проверяемый текст

Графом называется упорядоченная пара (V, E) $G = (V, E)$, где V — непустое множество вершин (узлов) графа G , а E — множество ребер графа G .

2) Граф называется связным, если любые две его вершины связаны.

Текст источника

Пусть $G = (V, E)$ — связный граф с множеством вершин V и множеством ребер E . <...>

Граф называется связным, если между любыми двумя его вершинами существует путь. <...>

57. Метод k-ближайших соседей

<http://mybiblioteka.su/tom2/10-85390.html>

Проверяемый текст

Круг должен быть классифицировать как синий квадрат, либо как красный треугольник (класс 1 и класс 2 соответственно).

Соответственно и объект будет классифицирован как синий квадрат, то есть 1-й класс.

Текст источника

Зеленый **круг должен быть классифицирован** как **синий квадрат (класс 1)** или как **красный треугольник (класс 2)**. <...>

Зеленый круг должен **быть классифицирован** как **синий квадрат (класс 1)** или как **красный треугольник (класс 2)**. <...>

58. Выбор и настройка параметров модели классификации спортсменов для формирования оптимальных по составу велокоманд.

<http://elibrary.ru/item.asp?id=29357979>

2015, Землянская С.Ю., Крайний Д.С.

Проверяемый текст

Классифицируемый объект x относится к тому классу i_y , которому принадлежат ближайший к нему объект обучающей выборки i_x .

Таким образом, произвольный объект u порождает свою перенумерацию выборки.

Текст источника

Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки [9]. <...>

Таким образом, произвольный объект x порождает свою перенумерацию выборки. <...>

59. Анализ базы данных траекторий автомобилей

http://nauchkor.ru/cloud_storage/documents/587d36885f1be77c40d591a4.pdf

Кокаия Мария Платоновна

Проверяемый текст

При создании очередного узла дерева случайным образом выбирается m признаков, на основе которых будет производится разбиение.

Для выбора лучшего из этих признаков используется критерий Джини.

Текст источника

Построение **деревьев** весьма специфично: в ходе **создания очередного узла дерева признак, на основе которого производится разбиение, выбирается** не из всех, а из m случайно выбранных. <...>

Чаще всего **используется критерий Джини.** <...>

60. Персональный сайт - 32. Система распознавания. Классы ...

<http://maksim-456.narod.ru/index/0-31>

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

пересекающиеся классы. Объект может относиться одновременно к нескольким классам.

Текст источника

При обучении сети предлагаются различные образцы образов с указанием того, к какому классу они относятся. <...> Многоклассовая классификация. **Когда число классов достигает многих тысяч (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно** <...>

Распознавание образов и классификация. В качестве <...> **Пересекающиеся классы. Объект может относиться одновременно к нескольким классам.**

61. Алгоритмическое и программное обеспечение web-сервиса для задач прогнозирования спроса поддержанных автомобилей: магистерская диссертация по направлению подготовки: 01.04.02 - Прикладная математика и

Проверяемый текст	Текст источника
Процесс голосования выглядит следующим образом: каждое дерево в комитете относит объект к одному из классов.	Классификация объектов проводится путём голосования: каждое дерево относит классифицируемый объект к одному из классов , и побеждает класс , за который проголосовало наибольшее число деревьев . <...>
Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.	Имеется множество объектов, разделённых некоторым образом на классы. <...>

[https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_\(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D)

Проверяемый текст	Текст источника
вершина называется изолированной, если она не является конечной ни для одного из ребер;	Граф — абстрактный математический объект, представляющий собой множество вершин <...> Две концевые вершины одного и того же ребра называются соседними. <...> Вершина называется изолированной, если она не является концом ни для одного ребра; висячей (или листом), если она является концом <...>
цикл называют простым, если ребра в нем не повторяются;	Граф — абстрактный математический объект, представляющий собой множество вершин <...> Путь (или цикл) называют простым, если рёбра в нём не повторяются ; элементарным, если он простой и вершины в нём не повторяются <...>
цепью называется маршрут без повторяющихся ребер;	Граф — абстрактный математический объект, представляющий собой множество вершин <...> Цепью называется маршрут без повторяющихся рёбер.

<http://elibrary.ru/item.asp?id=23573354>

2014, Донцова Ю.С.

Проверяемый текст	Текст источника
Генерируется случайная подвыборка с возвращением размером n .	На первом шаге из обучающей выборки генерируется случайная подвыборка с повторением размера n. <...>
Наивный байесовский классификатор ухудшил свои результаты.	Наивный байесовский классификатор ? <...>
Классовая принадлежность остальных объектов неизвестна.	Классовая принадлежность остальных объектов неизвестна. <...>

<http://referat.bookap.info/work/126090/Grafy>

Проверяемый текст	Текст источника
Графом называется упорядоченная пара (V, E) , где V – непустое множество вершин (узлов) графа G , а E – множество ребер графа G	Граф или неориентированный граф G – это упорядоченная пара $G = (V, E)$, для которой выполнены следующие условия: $V \neq \emptyset$

1) Деревом называется связный граф без циклов.

Граф G называется деревом, если он является связным и не имеет циклов. <...>

65. Практическая работа №1, Банк Рефератов

<http://www.bankreferatov.ru/referats/C325729F00717F7B43257B0B0009908A/%D0%9E%D1%81%D0%BD%D0%BE%I>

Татьяна Петровна, Банк Рефератов (с) 1998-2013

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

1) Деревом называется связный граф без циклов.

Текст источника

Тогда **графом $G(V, A)$ называется совокупность множества вершин и множества ребер. <...>**

Деревом называется связный граф, не содержащий циклов. <...>

66. Технология построения обобщенного "И/ИЛИ" дерева решения задач

<http://rsl.ru/rsl01004224143.txt>

2008, Вовк, Алексей Андреевич

Проверяемый текст

Чтобы формализовать понятие сходства вводится функция расстояния между объектами (x, y) в пространстве объектов X .

Вершина, из которой выходят ребра называется родительской.

Текст источника

Для формализации понятия сходства вводится **функция расстояния** или метрика $p(x, y)$ в пространстве объектов X . <...>

Вершина, из которой выходят 37 только «И» ребра, называется «И» вершиной. <...>

67. Управление большими системами.1

http://biblioclub.ru/index.php?page=book_red&id=77639

2016

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Пример такого графа приведен на рис.

Текст источника

Копирование (обозначается «?») выполняется над исходным **графом** (образцом или прототипом) $g = (V, E)$, где V множество вершин графа, E - множество ребер. <...>

Пример графа переходов ККА модели такого механизма представлен на рис. <...>

68. Теория и практика эволюционного моделирования

http://biblioclub.ru/index.php?page=book_red&id=82567

2016, Емельянов В. В., Курейчик В. В., Курейчик В. М.

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Пример такого графа приведен на рис.

Текст источника

Пусть задан **граф $G = (X, E, W)$** , где X представляет множество вершин графа, E - множество ребер, а W - общий суммарный вес вершин. <...>

Рассмотрим **пример для графа, приведенного на рисунке 6.48. <...>**

69. Разработка теории и основных принципов принятия решений в САПР на основе методов, инспирированных природными системами

<http://rsl.ru/rsl01005028388.txt>

Проверяемый текст

Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .

Пример такого графа приведен на рис.

Текст источника

Пусть задан **граф** $G = (X, U, W)$, где X представляет **множество вершин графа**, U - **множество ребер**, а W - **общий суммарный вес вершин**. <...>

Пример графа G Рис. <...>

70. <http://textarchive.ru/c-1997416-pall.html>

Проверяемый текст

Степенью вершины называется количество инцидентных ей ребер.

Граф H называется подграфом графа G , если $V_H \subseteq V_G$ и $E_H \subseteq E_G$.

1) Деревом называется связный граф без циклов.

Текст источника

Степенью вершины ($\deg v$) называется количество **рёбер, инцидентных** данной вершине. <...>

Граф $G_1 = (V_1, E_1)$ называется подграфом графа $G = (V, E)$, если $V_1 \subseteq V$, $E_1 \subseteq E$. <...>

Деревом называется **связный граф без циклов**. <...>

71. Введение в интеллектуальный анализ данных : учебное пособие
<http://vital.lib.tsu.ru/vital/access/services/Download/vtls:000529594/SOURCE1>
 2016, Замятин, Александр Владимирович

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $y: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X = \{x_1, \dots, x_m\}$.

Текст источника

Если пространство **объектов** обозначить как X и множество возможных ответов Y , то **существует неизвестная целевая зависимость** $y^*: X \rightarrow Y$, значения которой известны только на **объектах обучающей выборки** $X = \{x_i, y_i\}_{i=1}^m$, $y_i = y^*(x_i)$. <...>

72. Модификация алгоритма случайного леса для классификации нестационарных потоковых данных
<http://cyberleninka.ru/article/n/modifikatsiya-algoritma-sluchaynogo-lesa-dlya-klassifikatsii-nestatsionarnyh-potokovyh-dannyh>
 2016, ЖУКОВ АЛЕКСЕЙ ВИТАЛЬЕВИЧ, СИДОРОВ ДЕНИС НИКОЛАЕВИЧ

Проверяемый текст

Существует неизвестная целевая зависимость — отображение $y: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки $X = \{x_1, \dots, x_m\}$.

Текст источника

Существует неизвестная целевая зависимость - отображение $y^*: X \rightarrow Y$, значения которой известны только на **объектах конечной обучающей выборки** $X_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. <...>

73. Методы построения эффективных алгоритмов : учебное пособие Учебное пособие
<http://rucont.ru/efd/206661>
 2005

Проверяемый текст

2) Граф называется связным, если любые две его вершины связаны.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

* $\arg \max_k k$.

Текст источника

Граф называется связным, если существует путь между **любыми парами его вершин**. <...>

Дерево – это **связный граф без циклов**. <...>

Такой граф иногда называют лесом (рис. <...>

Примечание: предполагается, что $S^*P \geq K$. <...>

74. Математическое и программное обеспечение визуального анализа графовой информации сети

взаимодействующих объектов

<http://rsl.ru/rsl01008935986.txt>

2016, Коломейченко, Максим Игоревич

Проверяемый текст

Степенью вершины называется количество инцидентных ей ребер.

Существует много несколько разновидностей графов.

Добавить вершину в изначально пустую очередь.

Текст источника

Результат: размещение **вершины** и ее **инцидентных ребер**. <...>

Существует много промышленных продуктов для анализа **графов**. <...>

Выбирается **вершина**, с которой начинается обход, и добавляется в **изначально пустую очередь**. <...>

75. Программирование в алгоритмах

http://biblioclub.ru/index.php?page=book_red&id=236276

2016, Окулов С. М.

Проверяемый текст

Например, довольно часто не выполняется неравенство треугольника.

Неориентированный граф Пусть (V, E) называется дугой.

Пример такого графа приведен на рис.

Текст источника

Неравенство треугольника выполняется. <...>

Независимые множества Дан **неориентированный граф** $G=(V, E)$. <...>

На **рисунке 4.39** приведены **примеры графов** и их раскраски. <...>

76. Учебно-методический комплекс дисциплины "Комбинаторные алгоритмы"

http://elar.urfu.ru/bitstream/10995/1646/4/1333421_schoolbook.pdf

2008, Асанов, М. О. Расин, В. В.

Проверяемый текст

Пусть задан граф $G=(V, E)$ и фиксирована начальная вершина s .

Можно сказать, что дуга $v w$ ведёт от вершины v к вершине w .

Извлечь из начала очереди вершину u .

Текст источника

Вход: связный **граф** $G=(V, E)$ без **вершин** нечетной степени, **начальная вершина** v_0 . <...>

При каких условиях из уже помеченной **вершины** w можно пометить **вершину** v ? <...>

Более общо, пусть в **начале очереди** находится **вершина** u . <...>

77. Случайные леса: обзор

<http://cyberleninka.ru/article/n/sluchaynye-lesa-obzor>

2013, Чистяков Сергей Павлович

Проверяемый текст

Можно сказать, что дуга $v w$ ведёт от вершины v к вершине w .

Пусть T — множество всех терминальных вершин дерева.

1) Деревом называется связный граф без циклов.

Текст источника

Если M — некоторая **дуга**, то **вершина** V называется родителем w , а **вершина** w — потомком **вершины** V . <...>

Все такие **вершины** объявляются **терминальными**, а соответствующее **дерево** полным. <...>

Дерево представляет собой **связный граф** без **циклов**. <...>

78. Порождение и выбор моделей в задачах регрессии и классификации

<http://rsl.ru/rsl01005102065.txt>

Проверяемый текст

Соответственно и объект будет классифицирован как синий квадрат, то есть 1-й класс.

Например, довольно часто не выполняется неравенство треугольника.

Текст источника

Зелеными треугольниками и **синими квадратами** обозначены **объекты** разных классов. <...>

Значит, **неравенство треугольника** в этом случае выполнено. <...>

79. Решение задач классификации и кластеризации данных по экспрессии генов на основе их функциональных взаимодействий

<http://rsl.ru/rsl01004932720.txt>

2011, Бессарабова, Марина Олеговна

Проверяемый текст

Требуется построить алгоритм () : а $x \in X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Можно сказать, что дуга $v \rightarrow w$ ведёт от вершины v к вершине w .

Текст источника

Требуется построить классификатор на базе этой выборки, **способный классифицировать произвольный объект** как представителя одного из классов. <...>

Потомок для **вершины v** - любая **вершина w** , достижимая из v . <...>

80. тематики как и образованные позднее на их основе кафедры информатики и вычислительной техники не были избал

<http://samzan.ru/159136>

2016

Проверяемый текст

2) Граф называется связным, если любые две его вершины связаны.

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

Текст источника

Граф G называется связным, если каждые **две вершины** в нем являются **связанными**. <...>

Деревом будем называть неориентированный **связный граф без циклов**. <...>

Ниже **приведен пример** ориентированного **графа** и его матрицы достижимости, **рис.** <...>

81. [DOC] Моделирование и компьютерный эксперимент

http://mosmetod.ru/files/projects/urok_v_moskve/uroki/modelirovanie_s_pomoshyu_grifov/text4teacher/text4teacher1.doc

Проверяемый текст

вершина называется изолированной, если она не является конечной ни для одного из ребер;

цикл называют простым, если ребра в нем не повторяются;

Текст источника

Две концевые вершины **одного** и того же ребра называют соседними. <...> **Вершина называется изолированной, если она не является концом ни для одного** <...> Путь (цепь) в графе – конечная последовательность вершин, каждая **из** <...>

Процесс построения модели называется моделированием. Другими <...> Путь (?или **цикл**) **называют простым, если ребра в нем не повторяются**. Граф <...>

82. Структуры и алгоритмы обработки данных

http://biblioclub.ru/index.php?page=book_red&id=260413

2016, Хусаинов Б. С.

Проверяемый текст

Текст источника

<p>Графом называется упорядоченная пара (V, E), где V - непустое множество вершин (узлов) графа G, а E - множество ребер графа G.</p>	<p>Остовные деревья графа Остовным деревом для связного неориентированного графа $G = (V, E)$ с вершинами называется неориентированное дерево, содержащее все и вершин и $(n - 1)$ ребер графа.</p>
--	--

83. Математика программных систем: Межвузовский сборник научных статей (2006)
<http://window.edu.ru/resource/277/66277/files/%D0%9C%D0%9F%D0%A1-2006.pdf>

Проверяемый текст	Текст источника
Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .	Задан граф $G = (V, E)$, V - множество вершин, $n = V $, E - множество ребер, $E \subseteq V \times V$. <...>

84. Разработка инструментария анализа сложносвязных социальных данных
http://nauchkor.ru/cloud_storage/documents/587d36575f1be77c40d58d27.pdf

Лапицкая Людмила Юрьевна

Проверяемый текст	Текст источника
Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .	Рассмотрим ориентированный граф $G = (V, E)$, состоящий из множества вершин V и множества $E = V \times V$ ориентированных рёбер . <...>

85. <http://textarchive.ru/c-1220652-pall.html>

Проверяемый текст	Текст источника
Графом называется упорядоченная пара (V, E) , где V - непустое множество вершин (узлов) графа G , а E - множество ребер графа G .	Конечный граф $G = (V, E)$, содержащий p вершин и q ребер , называется (p, q) - графом . <...>

86. №6
<http://rucont.ru/efd/284582>
 2013

Проверяемый текст	Текст источника
1) Вершины a и b графа G называются связанными, если в графе существует путь между ними.	Если в дереве существует путь из вершины a к вершине b , то вершина a называется предком вершины b , а вершина b — потомком вершины a .
1) Деревом называется связный граф без циклов.	Дерево — это связный ациклический граф [5]. <...>

87. Нейросеть: дерево решений или случайный лес?
<http://elibrary.ru/item.asp?id=35160576>
 2018, Бучнева А.

Проверяемый текст	Текст источника
Он основан на применении теоремы Байеса со строгими предположениями о том, что все объекты описываются независимыми признаками.	Наивный байесовский классификатор - простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими предположениями о независимости. <...>

88. Методология поиска логических закономерностей в предметной области с нечеткой системологией

<http://rsl.ru/rsl01002852957.txt>

2005, Дюк, Вячеслав Анатольевич

Проверяемый текст

Чем является объект, определяется спецификой предметной области.

Пусть размерность пространства признаков будет равна M .

Текст источника

Специфика предметных областей с нечеткой системологией 1.2. <...>

Ранг матрицы B **равен размерности** искомого пространства отображения. <...>

89. Прогнозирование преждевременного расторжения договоров страхования жизни

http://knowledge.allbest.ru/bank/3c0b65635a3ac79b5c53a89521206c36_0.html

2016

Проверяемый текст

Генерируется случайная подвыборка с возвращением размером n .

Наивный байесовский классификатор ухудшил свои результаты.

Текст источника

Из набора примеров **генерируется подвыборка размером N** . <...>

1) **Наивный байесовский классификатор** . <...>

90. Структуры и алгоритмы обработки данных метод. указания

<http://rucont.ru/efd/202443>

2012

Проверяемый текст

Степенью вершины называется количество инцидентных ей ребер.

2) Граф называется связным, если любые две его вершины связаны.

Текст источника

Степень (валентность) вершины – это количество ребер, инцидентных этой вершине (количество смежных с ней вершин). <...>

Граф называется связанным, если для любой пары его вершин существует путь из одной вершины в другую. <...>

91. Методы и алгоритмы классификации данных на основе многомерной триангуляции Делоне

http://dekanat.bsu.edu.ru/f.php/1/disser/case/filedisser/filedisser/1049_Dissertaciya.pdf

Дорошенко Александр Юрьевич

Проверяемый текст

В результате была получена информация, приведенная в таблице 2.2.

Наивный байесовский классификатор ухудшил свои результаты.

Текст источника

Полученные результаты приведены в нижеследующих таблицах 5–7. <...>

Наивный байесовский классификатор 34 <...>

92. Граф (математика) - Дискретная математика. Теория множеств ...

<https://intellect.icu/graf-matematika-4272>

Проверяемый текст

путем в графе называют конечную последовательность вершин и дуг, в которой каждый элемент соединен с предыдущим и последующим;

Текст источника

Оценка 9/10 (2) Маршрутом в графе называют **конечную последовательность вершин**, в которой каждая вершина (кроме последней) соединена со следующей в <...> (?или путём) в орграфе **называют конечную последовательность вершин и дуг**, в которой каждый элемент инцидентен предыдущему и последующему. Циклом <...>

<http://textarchive.ru/c-1955529.html>

Текст источника

Граф называется связным, если все его вершины взаимно достижимы.. <...>

Дерево - это связный граф без циклов. <...>

http://biblioclub.ru/index.php?page=book_red&id=75502

Текст источника

Степенью вершины графа называется количество ребер, инцидентных данной вершине. <...>

Связный неориентированный ациклический граф называется деревом, множество деревьев называется лесом. <...>

<http://www.dslib.net/zaw-informacia/dinamicheskaja-iterativnaja-ocenka-riskov-informacionnoi-bezopasnosti-v.html>

Текст источника

4.4 Применение системы динамической итеративной оценки рисков **при** использовании облачных технологий 103 <...> Многоклассовая классификация? **Когда число классов достигает многих тысяч** (например, **при распознавании иероглифов или слитной речи**), задача классификации становится <...>

<https://docplayer.ru/41018206-Algorithm-opredeleniya-ceny-na-nedvizhimost.html>

Текст источника

При решении поставленной задачи были использованы такие методы интеллектуального анализа данных, предназначенные <...> Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача классификации становится <...>

<http://www.machinelearning.ru/wiki/index.php?>

Проверяемый текст

Текст источника

Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно более трудной.

103. Задача классификации — Википедия

https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BA%D0%BB%D0%B0%D1

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача? <...>

104. Задача классификации - Wikiwand

http://www.wikiwand.com/ru/%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BA%D0%BB%D0%B0%D1

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

В этих задачах разделение объектов обучающей выборки на классы не задаётся, и требуется <...> Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно более трудной.

105. Задача классификации - это... Что такое Задача классификации?

<https://dic.academic.ru/dic.nsf/ruwiki/638600>

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

В этих задачах разделение объектов обучающей выборки на классы не задаётся, и требуется <...> Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно более трудной.

106. [DOC] 1.5 Наивный Байесовский классификатор - Кубанский ...

https://www.kubsu.ru/sites/default/files/users/8723/portfolio/kursovaya1_vashchanov_2019.docx

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

2) Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача? <...>

107. Распознавание режимов работы авиационного ГТД с ...

https://knowledge.allbest.ru/programming/2c0a65635a3ad68a5c43a88521316d26_0.html

2012

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно более трудной.

108. Анализ существующих методик распознавания режимов работы ...

https://studbooks.net/2264089/informatika/analiz_suschestvuyuschih_metodik_raspoznaniya_rezhimov_raboty_gazotur

Проверяемый текст

многоклассовый, когда число классов достигает многих тысяч (при распознавании иероглифов или слитной речи).

Текст источника

Классификация объекта – номер **или** наименование класса, выдаваемый алгоритмом <...>
Многоклассовая классификация. **Когда число классов достигает многих тысяч** (например, **при распознавании иероглифов или слитной речи**), задача классификации становится существенно более трудной.

109. Алгоритмические языки и технологии программирования на языках высокого уровня [Электронный ресурс] электрон. курс лекций

<http://rucont.ru/efd/230047>

2010

Проверяемый текст

В результате была получена информация, приведенная в таблице 2.2.

Пример такого графа приведен на рис.

Текст источника

Формулы для вычисления характеристики и количество бит, необходимых для ее хранения, **приведены в таблице 2.2.** <...>

Примеры изображений графов даны на **рис.** <...>

110. Система определения музыкальных предпочтений

http://nauchkor.ru/cloud_storage/documents/587d363d5f1be77c40d58a4e.pdf

Ромашов Дмитрий Сергеевич

Проверяемый текст

Такой алгоритм называется бинарным решающим деревом.

Приведем в пример два критерия информативности.

Текст источника

В моей работе использовалось **бинарное решающее дерево.** <...>

Примеры критериев информативности: 1. <...>

111. Концептуальный анализ, принципы моделирования и оптимизация алгоритмов синтеза текстовых структур

<http://rsl.ru/rsl01008919772.txt>

2017, Суркова, Анна Сергеевна

Проверяемый текст

Например, довольно часто не выполняется неравенство треугольника.

Пример такого графа приведен на рис.

Текст источника

Заметим, что данное расстояние **не** удовлетворяет **неравенству треугольника.** <...>

6 приведен граф, соответствующий матрице с **рисунка П.** <...>

112. Модификация алгоритма случайного леса для классификации нестационарных потоковых данных.

<http://elibrary.ru/item.asp?id=27318769>

2016, Жуков А.В., Сидоров Д.Н.

Проверяемый текст

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Текст источника

Требуется построить алгоритм , способный классифицировать произвольный объект . <...>

113. Модели и методы параллельных вычислений для построения отказоустойчивых диагностических тестов в интеллектуальных системах с когнитивной компонентой

<http://rsl.ru/rsl01009492842.txt>

2017, Ямшанов, Артем Вячеславович

Проверяемый текст

пересекающиеся классы. Объект может относиться одновременно к нескольким классам.

Текст источника

Когда число классов достигает многих тысяч (например, при распознавании иероглифов или <...> **Пересекающиеся классы. Объект может относиться одновременно к нескольким классам.**

119. Классификация (задача) - это... Что такое Классификация ...

<https://dic.academic.ru/dic.nsf/ruwiki/969082>

Проверяемый текст

пересекающиеся классы. Объект может относиться одновременно к нескольким классам.

Текст источника

Когда число классов достигает многих тысяч (например, при распознавании иероглифов или <...> **Пересекающиеся классы. Объект может относиться одновременно к нескольким классам.**

120. Автоматическая расстановка поисковых тегов / Блог компании ...

<https://habr.com/ru/company/anetika/blog/222731/>

2014

Проверяемый текст

пересекающиеся классы. Объект может относиться одновременно к нескольким классам.

Текст источника

способный классифицировать произвольный **объект** $x \in X$. <...> Таким образом, у нас есть задача классификации с пересекающимися классами, т.е. **объект может относиться одновременно к нескольким классам.**

121. Признаковое пространство — Студопедия

https://studopedia.ru/3_30205_priznakovoe-prostranstvo.html

Проверяемый текст

пересекающиеся классы. Объект может относиться одновременно к нескольким классам.

Текст источника

2. **Пересекающиеся классы. Объект в этом случае может относиться одновременно к нескольким классам.** 3. Нечеткие классы. Требуется определить <...>

122. Основные определения, представление графа в эвм - StudFiles

<https://studfiles.net/preview/3719069/>

2015

Проверяемый текст

вершина называется изолированной, если она не является конечной ни для одного из ребер;

Текст источника

Многие хорошие результаты, полученные **для** конечных графов, <...> **Вершина называется изолированной, если она не является концом ни для одного ребра; висячей (или листом), если она является концом ровно <...>**

123. Базовые понятия теории графов - Bourabai Research

<http://bourabai.ru/dm/graph.htm>

Проверяемый текст

Текст источника

вершина называется изолированной, если она не является конечной ни для одного из ребер;

Многие хорошие результаты, полученные для конечных графов, неверны <...> Две концевые вершины **одного** и того же ребра называются соседними. <...> **Вершина называется изолированной, если она не является концом ни для <...>**

124. Теоретические основы развития систем автоматизации технологических процессов контурной сегментации изображений

<http://rsl.ru/rsl01005100693.txt>

2015, Колдаев, Виктор Дмитриевич

Проверяемый текст

1) Деревом называется связный граф без циклов.

Пример такого графа приведен на рис.

Текст источника

Пусть $G = (V, E)$ **связный** неориентированный граф. <...>

Для **графа, приведенного на рис.** <...>

125. Распознавание: математические методы, программная система, практические применения

http://knowledge.allbest.ru/mathematics/2c0b65635a2bc69a4d53a89521306c37_0.html

Проверяемый текст

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Текст источника

Пусть **имеется некоторое множество** допустимых **объектов**, являющееся объединением **классов**. <...>

126. Применение методов Data Mining для оценки надежности кредито заемщиков банка.

<http://elibrary.ru/item.asp?id=29808832>

2017, Сидорова Н.П.

Проверяемый текст

Имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Текст источника

Задача классификации [7] определяется следующим образом: **имеется** конечное **множество объектов (ситуаций)**, которые **разделённых** на **классы**. <...>

127. Порождение и выбор моделей в задачах регрессии и классификации

<http://www.ccas.ru/avtorefe/0016d.pdf>

Стрижова Вадима Викторовича

Проверяемый текст

Соответственно и объект будет классифицирован как синий квадрат, то есть 1-й класс.

Текст источника

Задача классификации [7] определяется следующим образом: **имеется** конечное **множество объектов (ситуаций)**, которые **разделённых** на **классы**. <...>

128. Избранные главы теории графов Wybrane rozdzialy teorii grafow

<http://rucont.ru/efd/301456>

2009

Проверяемый текст

Пусть задан граф (V, E) и фиксирована начальная вершина s .

* $\arg\max_k k$.

Текст источника

В таком **графе** $G = (W, E)$ обозначим через $E_s = E \cap \{w \in W : s \text{ -- } w \in E\}$ — множество **слепых (концевых) вершин**, а через $E^+ = E \cup \{w \in W : s \text{ -- } w \in E\}$ — множество **начальных вершин**. <...>
 $0, P \leq k \leq P$.

129. Постановка задачи

http://studopedia.ru/5_164907_postanovka-zadachi.html

2015

Проверяемый текст

Таким образом, произвольный объект порождает свою перенумерацию выборки.

Текст источника

Таким образом, произвольный объект порождает свою перенумерацию выборки. <...>

130. ФЗО_ИСТ_ИСТ_ДР_Столярова_Валентина_Владимировна

<https://text.rucont.ru/History/Item?ServiceCallID=260151>

2014

Проверяемый текст

Алгоритм работает, пока не обнаружит все вершины, достижимые из исходной.

Текст источника

В процессе работы **алгоритма** дерево разрастается, **пока не охватит все вершины исходного графа**. <...>

131. Выберите верные утверждения цикл - это цепь, в которой ...

<https://znanija.com/task/32306703>

Проверяемый текст

циклом называют цепь, в которой начальная и конечная вершины совпадают;

Текст источника

цикл - это **цепь, в которой начальная и конечная вершины совпадают** вершины неориентированного графа соединены дугами

132. Методы и алгоритмы обработки текстовых данных на основе графовых дискурсивных моделей

<http://www.frccsc.ru/sites/default/files/docs/ds/002-073-04/diss/05-ilvovsky/diss005-ilvovsky.pdf?194>

Ильвовский Дмитрий Алексеевич

Проверяемый текст

В результате была получена информация, приведенная в таблице 2.2.

Текст источника

Полученные результаты приведены в таблице. <...>

133. Модели выявления и предотвращения несанкционированных транзакций в области банковских карт в системе мягкого реального времени

<http://rsl.ru/rsl01004594625.txt>

2009, Муханов, Лев Евгеньевич

Проверяемый текст

В результате была получена информация, приведенная в таблице 2.2.

Текст источника

В результате были ото браны параметры, приведенные в таблице 2.2. <...>

134. Использование метрических признаков в решающих деревьях на примере задачи классификации типов лесных массивов

<https://cyberleninka.ru/article/n/ispolzovanie-metricheskih-priznakov-v-reshayuschih-derevyah-na-primere-zadachi-klassifikatsii-tipov-lesnyh-massivov>

2015, Китов Виктор Владимирович

Проверяемый текст

Для выбора лучшего из этих признаков используется критерий Джини.

Текст источника

В качестве **критерия** неравномерности моделей **использовался критерий Джини**. <...>

135. Информационные технологии в науке, управлении, социальной сфере и медицине. Ч. 1

http://earchive.tpu.ru/bitstream/11683/46139/1/conference_tpu-2017-C24_V1.pdf

2017, Берестнева, Ольга Григорьевна

Проверяемый текст

Текст источника

Данная задача разбивается на две подзадачи: 1. **Задача разбивается на две подзадачи:** сбора и передачи данных. <...>

* $\arg\max_k k$ k p .

K ?? p <...>

136. Прогнозирование преждевременного расторжения договоров страхования жизни с учетом параметров договоров и социально-демографических характеристик клиентов

http://statref.ru/ref_jgemerpolna.html

Проверяемый текст

Текст источника

Наивный байесовский классификатор ухудшил свои результаты.

. **Наивный байесовский классификатор** (naiveBayesclassifier) – особый случай **байесовского классификатора**. <...>

137. Разработка метода проактивного обнаружения мошенничества потребителей услуг телекоммуникационной компании.

<http://elibrary.ru/item.asp?id=28360836>

2016, Нгуен Т.А., Щербаков М.В., Чан В.Ф., Кравец А.Г.

Проверяемый текст

Текст источника

Наивный байесовский классификатор ухудшил свои результаты.

В [19, 30] использован **наивный байесовский классификатор**. <...>

138. Определение характеристик автора сообщений сетевых диалогов

http://nauchkor.ru/cloud_storage/documents/587d365e5f1be77c40d58de5.pdf

Сухочев Александр Владимирович

Проверяемый текст

Текст источника

Наивный байесовский классификатор ухудшил свои результаты.

Наивный байесовский классификатор . <...>

139. Распознавание последовательности аккордов в цифровом звуке

http://nauchkor.ru/cloud_storage/documents/587d36455f1be77c40d58b25.pdf

Шевченко Настасья Сергеевна

Проверяемый текст

Текст источника

Наивный байесовский классификатор ухудшил свои результаты.

Наивный Байесовский классификатор <...>

140. Оценка факторов риска возникновения аварийных ситуаций при сопровождении эксплуатационного бурения бурения (на примере Новопортовского нефтегазоконденсатного месторождения, Западная Сибирь)

<http://earchive.tpu.ru/bitstream/11683/41860/1/TPU421923.pdf>

2017, Григорьев, Алексей Сергеевич

Проверяемый текст

Текст источника

Наивный байесовский классификатор ухудшил свои результаты.

Наивный байесовский классификатор
.....51 <...>

141. №4

<http://rucont.ru/efd/314636>

2014

Проверяемый текст

Вершина, из которой выходят ребра называется родительской.

Текст источника

Вершины, для **которых** не существует **выходящих ребер**, называются терминальными или листьями дерева. <...>

142. Программирование на языке Пролог для искусственного интеллекта (fb2)

<http://coollib.com/b/183629/read>

Проверяемый текст

2) Граф называется связным, если любые две его вершины связаны.

Текст источника

Граф называется связным, если между **любыми двумя его вершинами** существует путь. <...>

143. [PDF] К ПРОБЛЕМЕ ПОСТРОЕНИЯ МОДЕЛИ ПОИСКОВОГО ... - СО РАН

http://elib.ict.nsc.ru/jspui/bitstream/ICT/1182/1/leonova_fedotov_lyap.pdf

Проверяемый текст

порядковый признак: fD — конечное упорядоченное множество;

Текст источника

Определение признаков – формирование поискового образа документа <...> **порядковый признак**: – **конечное упорядоченное множество** (образование, ..

144. ОБЗОР МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В КОНТЕКСТЕ РЕШЕНИЯ ЗАДАЧИ ОБНАРУЖЕНИЯ АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ.

<http://elibrary.ru/item.asp?id=32834567>

2018, Коновалов И.А., Бауман Ю.И.

Проверяемый текст

Пусть размерность пространства признаков будет равна M .

Текст источника

где - **размерность пространства признаков** объекта . <...>

145. Математическая модель и алгоритм распознавания стабильнографических сигналов при исследовании опорно-двигательного аппарата человека

<http://rsl.ru/rsl01006626637.txt>

2013, Ажигитов, Рамиль Фяритович

Проверяемый текст

Пусть размерность пространства признаков будет равна M .

Текст источника

Таким образом, общая **размерность** признакового **пространства равна 20**. <...>

146. Математическое обеспечение многоуровневых систем распознавания сигнальной информации в условиях априорной неопределенности

<http://rsl.ru/rsl01000322023.txt>

2000, Геппенер, Владимир Владимирович

Проверяемый текст

Пусть размерность пространства признаков будет равна M .

Текст источника

Размерность исходного **пространства признаков** **будет равна размерности** эталонов. <...>

147. Математическое и программное обеспечение процесса распознавания первичной информации об успеваемости в информационной системе вуза

<http://rsl.ru/rsl01004591779.txt>

2009, Грушин, Алексей Валерьевич

Проверяемый текст

Пусть размерность пространства признаков будет равна M .

Текст источника

Задачи небольшой **размерности пространства формальных признаков**. <...>

148. Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской научной конференции RCDL"2009

<http://window.edu.ru/resource/819/67819/files/rcdl2009.pdf>

Проверяемый текст

Результаты такого эксперимента приведены в таблице 2.3.

Текст источника

Результаты эксперимента приведены в таблице 2. <...>

149. Исследование и разработка методов извлечения знаний для создания интеллектуальных систем поддержки принятия решений

<http://dlib.rsl.ru/loader/view/01002900085?get=pdf>

2005, Айман Мохамед Мофтах Кхамес Йоунес Бериша

Проверяемый текст

Результаты такого эксперимента приведены в таблице 2.3.

Текст источника

Результаты экспериментов приведены в таблицах 1 и 2. <...>

150. <http://tekhnosfera.com/issledovanie-i-razrabotka-metodov-izvlecheniya-znaniy-dlya-sozdaniya-intellektualnyh-sistem-podderzhki-prinyatiya-resheni>

Проверяемый текст

Результаты такого эксперимента приведены в таблице 2.3.

Текст источника

Результаты экспериментов приведены в таблицах 1 и 2. <...>

151. КОНЦЕПТУАЛЬНЫЙ АНАЛИЗ, ПРИНЦИПЫ МОДЕЛИРОВАНИЯ И ОПТИМИЗАЦИЯ АЛГОРИТМОВ СИНТЕЗА ТЕКСТОВЫХ СТРУКТУР

http://www.nntu.ru/sites/default/files/file/dissertacii/2017/surkova_a_s.pdf

Суркова Анна Сергеевна

Проверяемый текст

Результаты такого эксперимента приведены в таблице 2.3.

Текст источника

В таблице 5.4 приведены результаты 12 экспериментов над 45 текстовыми объектами. <...>

152. Реферат: Основные определения курса Распознавание Образов

<http://www.bestreferat.ru/files/94/bestreferat-253594.docx>

Проверяемый текст

Пусть размерность пространства признаков будет равна M .

Текст источника

K (для случая если **размерность пространства признаков** = 2). <...>

153. Классификация на основе полных решающих деревьев

<http://www.mathnet.ru/links/0665b66c22b8f762cfdc640756dd5a90/zvmmf9692.pdf>

Проверяемый текст	Текст источника
Такой алгоритм называется бинарным решающим деревом.	и методом бинарного решающего дерева (см. <...>

154. КЛАССИФИКАЦИЯ НА ОСНОВЕ ПОЛНЫХ РЕШАЮЩИХ ДЕРЕВЬЕВ.

<http://elibrary.ru/item.asp?id=17680048>

2012, Генрихов И.Е., Дюкова Е.В.

Проверяемый текст	Текст источника
Такой алгоритм называется бинарным решающим деревом.	и методом бинарного решающего дерева (см. <...>

155. Адаптивные методы извлечения информации из статистических таблиц, представленных в текстовом виде

<http://rsl.ru/rsl01005409562.txt>

2011, Кудинов, Павел Юрьевич

Проверяемый текст	Текст источника
Такой алгоритм называется бинарным решающим деревом.	Алгоритм 1Т1 строит бинарное решающее дерево. <...>

156. Простой цикл - это... Что такое Простой цикл?

<https://dic.academic.ru/dic.nsf/ruwiki/1108538>

Проверяемый текст	Текст источника
цикл называют простым, если ребра в нём не повторяются;	Путь (или цикл) называют простым , если ребра в нём не повторяются ; элементарным, если он простой и вершины в нём не повторяются . Несложно <...>

157. Основы теории графов — МТ1100: Дискретная математика ...

<https://it.rfei.ru/course/~mBme/~6yq/~jcbArf>

Проверяемый текст	Текст источника
цепью называется маршрут без повторяющихся ребер;	Можно сказать, что орграф — это граф, ребрам которого присвоено направление. <...> Путем в орграфе называется маршрут без повторяющихся ребер , <...>

158. tutorial

<https://habrahabr.ru/company/ods/blog/324402/>

Проверяемый текст	Текст источника
Ограничение минимального числа объектов в листе.	Можно задать числом или процентом от общего числа объектов (по дефолту — 2) <code>min_samples_leaf</code> — минимальное число объектов в листе . <...>

159. Теоретико-групповой подход в комбинаторной теории переобучения

<http://rsl.ru/rsl01006768907.txt>

2013, Фрей, Александр Ильич

Проверяемый текст	Текст источника
-------------------	-----------------

Приведем в пример два критерия информативности. **Критерий** предсказанной **информативности**. <...>

160. Методы решения задач распознавания образов комбинированного типа

<http://rsl.ru/rsl01004152195.txt>

2008, Борисова, Ирина Артемовна

Проверяемый текст

Текст источника

Приведем в пример два критерия информативности. Сравнение **критериев информативности** 2.2.3. <...>

161. Микропроцессорные системы управления и распознавания для агротехнологий

<http://rsl.ru/rsl01003309778.txt>

2005, Ерков, Аркадий Александрович

Проверяемый текст

Текст источника

Приведем в пример два критерия информативности. Сравнительная диаграмма **критериев информативности**. <...>

162. Индивидуальное прогнозирование параметров элементов и компонентов микросборок и узлов радиоэлектронных устройств передачи информации

<http://rsl.ru/rsl01002881425.txt>

2005, Карпов, Олег Владимирович

Проверяемый текст

Текст источника

Приведем в пример два критерия информативности. **Критерии информативности** признаков 2.3. <...>

163. Алгоритмы компьютерной обработки данных

<http://rucont.ru/efd/206398>

2011

Проверяемый текст

Текст источника

Данная задача разбивается на две подзадачи: 1. Сначала **задача разбивается** на несколько **подзадач** меньшего размера. <...>

164. Современные технологии выявления мошеннических транзакций, проводимых с использованием банковских карт.

<http://elibrary.ru/item.asp?id=23465524>

2015, Климов В.В., Щукин Б.А.

Проверяемый текст

Текст источника

Данная задача разбивается на две подзадачи: 1. Таким образом, **задача разбивается** на несколько **подзадач**: 1. <...>

165. ИВТ

<https://text.rucont.ru/History/Item?ServiceCallID=1315615>

2018, ИЗ

Проверяемый текст

Текст источника

Данная задача разбивается на две подзадачи: 1. **Задача** нахождения ассоциативных правил **разбивается на две подзадачи**:. <...>

166. Практикум по методам построения алгоритмов

http://biblioclub.ru/index.php?page=book_red&id=234674

2016, Шень А. Х.

Проверяемый текст

Текст источника

Порядок обхода графа при поиске в глубину На рис.

Обход вершин **графа** в указанном **порядке** называется **поиском в глубину**. <...>

167. Математическое и программное обеспечение визуального анализа графовой информации сети взаимодействующих объектов

https://www.mirea.ru/upload/medialibrary/1b5/kolomoychenko_dissertatsiya_f.pdf

Коломейченко Максим Игоревич

Проверяемый текст

Текст источника

Добавить вершину в изначально пустую очередь.

Выбирается **вершина**, с которой начинается обход, и добавляется в **изначально пустую очередь**. <...>

168. Структуры данных и проектирование программ

<http://www.bibliorossica.com/book.html?currBookId=18653>

2014, Роберт Л. Круз

Проверяемый текст

Текст источника

1) Деревом называется связный граф без циклов. 12.2, (е) изображен **связный граф без циклов**. <...>

169. Программирование на языке Pascal

http://biblioclub.ru/index.php?page=book_red&id=232980

2016, Андреева Т. А.

Проверяемый текст

Текст источника

1) Деревом называется связный граф без циклов. /; **Дерево** - это **связный граф без циклов**. <...>

170. Теорема о включениях и исключениях

<http://lektsii.org/11-88125.html>

2016

Проверяемый текст

Текст источника

1) Деревом называется связный граф без циклов. **Дерево**(Tree). **связный граф без циклов**.. <...>

171. Программирование на языке Pascal. Учебное пособие

<http://www.bibliorossica.com/book.html?currBookId=12086>

2006, Андреева Т. А.

Проверяемый текст

Текст источника

1) Деревом называется связный граф без циклов. **1) Дерево** — это **связный граф без циклов**. <...>

172. Презентация - Тушко Т.А., Исаев С.В., Постников А.И. Дискретная математика

<http://bib.convdocs.org/v1560/?download=file>

Проверяемый текст

Текст источника

1) Деревом называется связный граф без циклов. **ДЕРЕВЬЯ** Неориентированным **деревом** **Т** будем называть **связный** неориентированный **граф без циклов**. <...>

173. №1

<http://rucont.ru/efd/285169>

2012

Проверяемый текст

1) Деревом называется связный граф без циклов.

Текст источника

М. Б. Абросимов, Д. Д. Комаров **Связный граф без циклов называется деревом.** <...>

174. Динамическое программирование

<http://rucont.ru/efd/443406>

2015

Проверяемый текст

1) Деревом называется связный граф без циклов.

Текст источника

В теории **графов связный** неориентированный **граф без циклов называется деревом.** <...>

175. Выбор оптимальной сложности класса логических решающих функций в задачах анализа разнотипных данных

<http://rsl.ru/rsl01003386504.txt>

2006, Бериков, Владимир Борисович

Проверяемый текст

1) Деревом называется связный граф без циклов.

Текст источника

Деревом, как известно, **называется связный** неориентированный **граф**, не содержащий **циклов.** <...>

176. Разработка моделей и методов повышения эффективности разработки и функционирования автоматизированных информационно-управляющих систем в социально-экономической сфере

<http://rsl.ru/rsl01003296233.txt>

2002, Зайцев, Константин Сергеевич

Проверяемый текст

Пример такого графа приведен на рис. **Такие графы** называют взвешенными. <...>

Текст источника

177. Методы и алгоритмы управления маршрутизацией в транспортных сетях на основе оперативной обработки информации в разреженных графах

<http://rsl.ru/rsl01007984566.txt>

2015, Тимеряев, Тимофей Валерьевич

Проверяемый текст

Пример такого графа приведен на рис.

Текст источника

Пример разборки неориентированного нагруженного **графа.** <...>

178. Применение вероятностно-статистических методов и теории графов в электроэнергетике : учебное пособие

http://elar.urfu.ru/bitstream/10995/43624/3/978-5-321-02507-9_2016.pdf

2016, Обоскалов, В. П. Кокин, С. Е. Кирпикова, И. Л.

Проверяемый текст

Для новой вершины выполняем шаг 1.). **Вершине Шаг 1.** <...>

Текст источника

179. Модели и методы построения корпоративных интеллектуальных систем поддержки принятия решений

<http://rsl.ru/rsl01002635503.txt>

2004, Швецов, Анатолий Николаевич

Проверяемый текст

Текст источника

Для новой вершины выполняем шаг 1. Снова **выполняем** просмотр достигнутых **вершин**, аналогично **шагу 1**. <...>

180. nntu@nntu.ru

http://www.nntu.ru/sites/default/files/file/dissertacii/2018/ponomorenko_a_a.pdf

Пономаренко Александр Александрович

Проверяемый текст

Текст источника

* $\arg\max_k k$ k p . K., 1991], [Yianilos P. <...>

181. Анализ и проектирование маршрутов передачи данных в корпоративных сетях [монография]

<http://rucont.ru/efd/202831>

2012

Проверяемый текст

Текст источника

* $\arg\max_k k$ k p . p ? k; p ? j , k; p ? j. <...>

182. №1

<http://rucont.ru/efd/285170>

2013

Проверяемый текст

Текст источника

* $\arg\max_k k$ k p . P. and Sakallah K. <...>

183. Методы и модели эконометрики. Ч. 1. Анализ данных учеб. пособие

<http://rucont.ru/efd/468886>

2015

Проверяемый текст

Текст источника

* $\arg\max_k k$ k p . Пусть k ? 2 , p ? 2 , ? 1 ? ? 2 , <...>

184. Методы структурного обучения в задачах совместной разметки диссертация

<https://istina.msu.ru/download/8936822/1g5c8g:dHOCLsHgREXV4A1NpqC5s69jjxk/>

Шаповалов Роман Викторович

Проверяемый текст

Текст источника

* $\arg\max_k k$ k p . (x) ? $\arg\max P(? y | x, wML)$. <...>
