План статьи

**Классификация объектов на основе системы нечеткого логического вывода**

**Вступ**

Задача разделения множества наблюдений (объектов) на группы, называемые классами, на основе анализа их формального описания. При классификации каждая объект наблюдения относится к определенной группе или номинальной категории на основе некоторого качественного свойства или совокупности свойств.

**Задачам классификации посвящено много работ.**

К числу распространенных методов решения задачи классификации относятся:

нейронные сети;

логистическая и пробит-регрессия;

деревья решений;

метод ближайшего соседа;

машины опорных векторов;

дискриминантный анализ.

Традиционные подходы, основанные на аппарате математической статистики или имитационном моделировании, не позволяют строить адекватные модели в условиях ограниченности временных, вычислительных и материальных ресурсов. Поэтому при решении многих практических задач, связанных с классификацией объектов, широко применяются модели и методы искусственного интеллекта с применением технологий интеллектуального анализа данных.

Так в работе [Штовба] для решения задачи классификации применяется аппарат нечеткой логики. В работе осуществляется расширение функц. возможностей пакета Fuzzy Logic Toolbox системы MATLAB для выполнения нечеткого логического вывода для объектов с дискретным выходом. (подробно об особенностях работы)

В работе [**Романов Вадим**] исследуется применение нечетких моделей в задачах классификации на основе представления данных в виде нечетких градаций. Введены классы объектов, эталонные образцы в каждом классе. Каждый класс характеризуется распределенной областью значений нечетких критериев. Исходные данные для классификации являются противоречивыми, а объекты, эталоны и классы пересекаются, поэтому построение решения задачи в классической постановке вызывает непреодолимые сложности. Классификация рассматривается как разновидность задачи принятия решений, в которой посредством обобщения нечетких фактов, характеризующих свойства, состояния или изменение состояний объектов, осуществляется выбор наилучшего класса для каждого объекта. Проведено сравнение различных мер согласования объектов с классами и их влияние на результаты классификации.

В настоящей работе предлагается подход…..

**Постановка задачи.**

Рассматривается множество объектов некоторой предметной области. Объекты множества Х характеризуются некоторыми признаками К. Признаки могут быть нечеткими. Существует множество классов Z . Требуется определить принадлежность объектов множества Х классам Z.

Все признаки задаются нечеткими значениями. Для определения степени согласования между объектами и классами используется метрика… (Евклид, Хеминг?)

Для объектов предметной области существует обучающая выборка, которая описывает отображение а\*: Х->Z только на объектах конечной обучающей выборки. Требуется построить отображение а: Х->Z, с помощью которого возможно классифицировать произвольный объект множества Х.

**Математична модель задачі**

Для описания объектов предметной области будем использовать нечеткую модель представления знаний. Признаки объектов будем описывать при помощи лингвистических переменных.   
{\displaystyle \{x,T(x),X,G,M\}},

(имя переменной; {\displaystyle T(x)}множество значений лингвистической переменной{\displaystyle x}, каждое из которых является нечеткой переменной на множестве{\displaystyle X}; {\displaystyle G}семантическая процедура, позволяющая преобразовать новое имя, образованное процедурой{\displaystyle G}, в нечеткую переменную (вид [функции принадлежности](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F_%D0%BF%D1%80%D0%B8%D0%BD%D0%B0%D0%B4%D0%BB%D0%B5%D0%B6%D0%BD%D0%BE%D1%81%D1%82%D0%B8)), ассоциирует имя с его значением, понятием).

Для відображення чітких вхідних значень – в нечіткі множини вводяться функції приналежності  виду:

 (3)

де  - числові параметри, які визначають границі термів, набувають дійсних значень і впорядковані відношенням: .

Для определения границ термов в работе используется IQR - міжквартильний розмах IQR = Q3 − Q1.

Подробно описать зачем и как он вычисляется.

Система правил, которая описывает соответствие между признаками объектов и соответствующими класами, представляется в виде продукционных правил.

**Метод розв’язання**

Для виконання процедури нечіткого логічного використовуються операції:

фазификации, операція імплікації, agg – операція агрегування, які реалізовані операцією знаходження  та  відповідно.

Чітке значення виходу *Y\** визначається в результаті дефазифікації нечіткої множини  за методом центру тяжіння.

Для виконання ідентифікації об'єкта використовується алгоритм нечіткого логічного виведення [7], модифікований введенням додаткового механізму порівняння чіткого вихідного значення *Y\**, отриманого для об'єкта , і вихідних значень об'єктів  наявних в базі знань.

Відстань між об'єктами визначається на основі обраної метрики в просторі характеристик. Для оцінки міри близькості елементів використовується Евклідова відстань:

 (10)

 (11)

У разі, коли для об'єкта *X\** умова (11) не виконується, необхідно сформулювати нове логічне правило:

  (12)

де,– нечіткі значення змінних , які обчислюються з використанням функції приналежності (5) при значеннях .

Після виконання процедури дефазифікації обчислюється *d(Y\*,Y)* та перевіряється умова (11).

Далі отримуємо множину об’єктів  таких, що , , M – кількість елементів цієї множини.

Далее вычисляется мощность всех найденных объектов  относительно их класса принадлежности.

 (13)

, де c – клас (X – характеризується <”Назва”, вектор ознак, “Клас”>)

После чего определяется класс входного объекта, относительно наибольшей мощности класса из всех найденных объектов.

**Алгоритм**

**Анализ результатов**

Запропонований підхід до класифікації об’єктів за допомогою нечіткої логіки та елементів навчальної вибірки, був протестований для задачі класифікації Iris Data Set - Іриси Фішера.

База знаний представлена в таблице. Информация о цветке представлена пятеркой чисел — первые четыре числа соответствуют значениям признаков, а пятое — классу ириса.

Вихідна змінна характеризує класи об'єктів: Клас (I, II, III).

Для кожної вхідної та вихідної змінних вводяться функції приналежності виду (3), границі термів лінгвістичних змінних обчислюються за формулами (4).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Переменные**  **системы** |  | **Лингвистические**  **переменные** | **Лингвистические значения** | **Нечеткие**  **интервалы** |
| **входные**  **переменные** | **x1** | Скорость передачи  информации  [10 ; 1000] кбит/с | Низкая | [0 ; 300] |
| Средняя | [200 ; 800] |
| Высокая | [700 ; 1000] |
| **x2** | Задержка  [10 ; 250] мс | Маленькая | [0 ; 50] |
| Средняя | [50 ; 100] |
| Большая | [100 ; 250] |
| **x3** | Потери пакетов  [10-10 ; 10-1] | Приемлемые | [10-10 ; 10-4] |
| Неприемлемые | [10-4 ; 10-1] |
| **Выходные пе-ременные** | **y** | Качество предоставления услуги “Аудио и видео  системы, в том числе для оперативного  вмешательства”  [0 ; 100] % | Очень низкое | [0 ; 20] |
| Низкое | [20 ; 40] |
| Среднее | [40 ; 60] |
| Высокое | [60 ; 80] |
| Очень Высокое | [80 ; 100] |

Формується система правил, передумови яких складені з усіх можливих комбінацій значень нечітких вхідних змінних (всього 81 правило).

Правила нечеткого логического вывода должны удовлетворять требованию слабой согласованности, которое предполагает, что произвольные малые приращения входных параметров не должны приводить к скачкообразному изменению выходного параметра.

Навчальна вибірка складається з пар об’єкт – клас, кожен об’єкт зображується вектором параметрів.

Представление результатов

Показать результаты для всех объектов выборки (подумать как)

Сравнить с результатми Штовбы, чтобы показать преимущества учета особенностей твоїй модели

Это просто кусок, может быть для обзора подойдет

(ИАД). ИАД находится на пересечении нескольких областей, главные из которых – системы баз данных, статистика и искусственный интеллект; в ходе исследований анализируются не только цифровые данные, но и текст (лингвистические данные), изображения, звук и т.д. Новая и быстро растущая часть ИАД – это анализ связей между данными (link analysis), которая имеет приложения в таких разных областях, как биоинформатика, цифровые библиотеки и защита от терроризма [2, 3]. Указанные особенности технических систем как объектов диагностики определили необходимость создания методов диагностики с применением элементов теории нечетких множеств (НМ), позволяющей проводить анализ в условиях неопределенности на основе нечетких рассуждений и правил логического вывода. Нечеткие модели описывают сложные объекты на естественном языке при помощи лингвистических переменных, а сам механизм нечеткого логического вывода понятен человеку. Эти преимущества обусловили широкое применение методов нечеткой логики для решения задач диагностики, управления, поддержки принятия решений в различных предметных областях человеческой деятельности. Теория нечетких множеств появилась в результате обобщения и переосмысления достижений в многозначной логике, теории вероятностей и математической статистики, дискретной математике, теории матриц, теории графов, теории грамматики и т.д. и начала развиваться после публикации в 1965 году основополагающей работы Л.Заде. Дальнейшее развитие теории нечетких множеств связано с созданием строгих и гибких математических методов исследования нечетко определенных объектов [3, 4, 5, 6]. Можно выделить следующие основные классификационные признаки способов формализации нечеткости [4]: 1) по виду представления нечеткой субъективной оценки величины (нечеткого множества); 2) по виду области значений функции принадлежности; 3) по виду области определения функций принадлежности; 4) по виду соответствия между областью определения и областью значений (однозначное, многозначное); 5) по признаку однородности или неоднородности области значений функции принадлежности. Учет фактора неопределенности при решении задач во многом изменяет методы принятия решения: меняется принцип представления исходных данных и параметров модели, становятся неоднозначными понятия решения задачи и оптимальности решения. Наличие неопределенности может быть учтено непосредственно в моделях соответствующего типа с представлением недетерминированных параметров как случайных величин с известными вероятностными характеристиками, в виде нечетких величин с заданными функциями принадлежности или интервальных величин с фиксированными интервалами изменения. Решение задачи осуществляется соответственно с помощью методов стохастического, нечеткого или интервального программирования. В целом алгоритмы на базе методов нечетких множеств хорошо зарекомендовали себя на практике при решении самого разнообразного круга задач. Успешным является и применение теории нечетких множеств в стохастических системах. Это связано с тем, что для многих систем трудно получить точные значения вероятностных характеристик (например, вероятности отказов компонентов). Преимущества методов нечеткой логики по сравнению с другими состоят в возможности: - оперировать в анализе качественными переменными как входных данных, так и получаемых выходных данных (результатов); - оперировать нечеткими входными данными, например, непрерывно изменяющимися во времени значениями (динамические задачи), значениями, которые невозможно задать однозначно (результаты статистических опросов, рекламные компании и т.д.); - оперировать лингвистическими критериями, обеспечивающими возможность нечеткой формализации критериев оценки и сравнения (оперирование критериями «большинство», «возможно», «преимущественно» и т.д.); - быстро моделировать сложные динамические системы и сравнивать их с заданной степенью точности; - преодолевать недостатки и ограничения существующих методов оценки проектных рисков. Таким образом, оценивание поведения системы методами нечеткой логики, во-первых, позволяет не тратить много времени на выяснение точных значений переменных и составление описывающих уравнений, а во-вторых, позволяет оценить разные варианты выходных значений.