

ПОСТРОЕНИЕ НЕЧЕТКОГО КЛАССИФИКАТОРА В ЗАДАЧЕ КОНТРОЛЯ КАЧЕСТВА НЕФТЕПРОДУКТОВ

Максимова А.Ю.

Институт прикладной математики и механики НАН Украины

Донецк

E-mail: lunaplus@mail.ru

Аннотация

Максимова А.Ю. Построение нечеткого классификатора в задаче контроля качества нефтепродуктов. Решается задача классификации жидких нефтепродуктов. Для решения задачи используется алгоритм распознавания образов на базе нечетких портретов. Результат работы алгоритма представляется в виде нечеткого множества. Для повышения качества распознавания было выполнено разбиение на кластеры некоторых классов образов, что позволило улучшить качество распознавания на 2% по сравнению с исходным алгоритмом (92% верно распознанных объектов).

Введение

В различных областях человеческой деятельности возникают задачи, сводящиеся к классу задач распознавания образов или классификации. Развитие компьютерных технологий позволило накапливать большие объемы информации, требующие обработки и преобразования с целью извлечения из них необходимых данных и знаний. В связи с этим активно разрабатываются методы обработки данных, и, в частности, методы распознавания образов. По [1] целью разведочного анализа данных (exploratory data analysis) является выявление структуры данных и описание ее статистическими моделями. Однако этот подход требует достаточно жестких модельных ограничений, вытекающих из аксиоматики теории вероятностей. Выход за рамки статистических моделей привел к формированию более общего направления интеллектуального анализа данных, в котором используются методы нечеткой логики, эволюционные и генетические алгоритмы, иммунные системы и др., а также, гибридные методы, которые в совокупности получили название «мягкие вычисления» (soft computing).

На практике исходные данные зачастую обладают сложно формализуемой, неоднородной структурой с заведомо пересекающимися классами образов. Другой проблемой является неполнота обучающей выборки и ее сильная зашумленность. В дискриминантном анализе в случаях неопределенности ответа о принадлежности образа классу образов ответ может быть получен в виде вероятности принадлежности образа каждому из классов образов. Однако описанные выше особенности данных во многих случаях не позволяют построить адекватные вероятностно-статистические модели, что обуславливает создание эмпирических методов и подходов, при разработке которых удобными становятся методы и модели из области мягких вычислений.

Целью работы является решение прикладной задачи классификации жидких нефтепродуктов (ЖНП), в частности бензинов. Эта задача возникает при контроле их качества по результатам лабораторных измерений некоторых их физико-химических характеристик. Для ее решения предлагается алгоритм распознавания образов на базе нечетких портретов, в основу которого положена идея построения интегральных характеристик классов образов в виде нечетких множеств – значений лингвистических переменных, описывающих классы образов по признакам, а принятие решения о

принадлежности рассматриваемого образца к определенному виду выполняется методами логического нечеткого вывода.

Особенности рассматриваемой задачи классификации жидких нефтепродуктов

Как отмечалось выше, в работе предлагается метод для решения задачи контроля качества нефтепродуктов, которая сводится к задаче распознавания образов. Ее суть заключается в определении производителя и вида образца ЖНП в лабораторных условиях. В лаборатории контроля качества нефтепродуктов накапливается информация об образцах топлива, поступающих на экспертизу от разных производителей и потребителей. По каждому образцу разными методами определяется ряд показателей. Такими показателями являются октановое число, содержание ароматических веществ, в том числе олефинов, ароматических бензолов, содержание серы и др. Накопленная информация по всем образцам используется в качестве эталонной для определения производителя рассматриваемого образца. Отсутствие универсальных подходов к измерению параметров ЖНП и неоднородный характер накопленной информации сдерживают развитие методов и приборов для автоматизации процессов управления в лабораториях контроля качества. Существующие разработки обладают рядом недостатков и ограничений в использовании. Например, в работе [2] предлагается нейросетевой импедансный метод идентификации ЖНП, работающий с данными, полученными с импедансометрического датчика. Эти данные обрабатываются нейросетевым алгоритмом. Однако алгоритм ориентирован на уникальный прибор, предложенный автором указанного метода.

В качестве класса образов в данной задаче рассматривается топливо определенного вида от определенного производителя. Основной особенностью исходных данных является полное или частичное совпадение некоторых классов образов. Это связано с технологическими особенностями производства ЖНП, когда разные производители реализуют топливо одного вида и с одинаковыми характеристиками. В ситуациях, когда не удается однозначно определить, к какому классу образов относиться рассматриваемый образец, важно получить ответ о степени схожести образца на каждый из предполагаемых классов образов, а, иногда, определить, что это точно не образец определенного класса образов. В связи с этим удобно ответ получать в виде нечеткого множества, элементы которого - классы образов.

Постановка задачи

Формально постановку прикладной задачи распознавания нефтепродуктов можно рассматривать как нечеткую модификацию задачи классификации в многомерном пространстве. Пусть дана обучающая выборка $Y = \{(x^{(i)}, v^{(i)}), x^{(i)} \in X, v^{(i)} \in V, i = 1, \dots, n\}$, где $x^{(i)} \in X \subset R^m$ векторы n -мерного пространства – набор информативных признаков, $V = \{v_i\}$, $i = 1, \dots, k$, $v_i \in N$ - множество номеров классов образов. Пары $(x^{(i)}, v^{(i)})$ определяют, представителем какого класса образов $v^{(i)}$ является образ $x^{(i)}$. В общем случае, необходимо определить степень принадлежности образца \bar{x} рассматриваемым классам, т. е. построить нечеткое множество $\tilde{y}(\bar{x}) = \sum_{i=1}^k \mu_i(\bar{x}) / v_i$. где $\mu_i(\bar{x})$ - степень принадлежности образа классу v_i .

Алгоритм распознавания по нечетким портретам классов образов

Для решения рассматриваемой задачи классификации нефтепродуктов используется предложенный в [3] алгоритм распознавания. Основная идея алгоритма заключается в представлении исходной информации о классах в виде их нечетких портретов, которые формируются в результате анализа обучающей выборки. Такие портреты описываются совокупностью лингвистических переменных, соответствующих информативным признакам. Терм-множества этих лингвистических переменных описывают значение признака для

каждого из классов образов и строятся в результате анализа частоты встречаемости значений признака в каждом классе образов. По построенным нечетким портретам формируется база знаний нечетких продукций, а решение принимается алгоритмом нечеткого вывода. Результат работы алгоритма представляется нечетким множеством.

В алгоритме берутся в рассмотрение признаки с низкой попарной корреляцией, поэтому данные по каждому показателю рассматриваются независимо друг от друга. Первоначальный выбор информативных признаков в задаче классификации ЖНП был осуществлен экспертом.

Рассмотрим более подробно алгоритм построения нечетких портретов. Для каждого информативного признака P_i , $i=1, \dots, m$ определим лингвистическую переменную L_i : $L_i = \{eiy(P_i), T_i, U_i, G, M\}$, где U_i - область определения значений признака P_i , $T_i = \{(\mu_{i,j}(x)), j=1, \dots, k\}$ - терм-множества лингвистической переменной, $\mu_{ij}(pr_i \bar{x}) \in [0,1]$ - функция принадлежности, определяющая степень уверенности, с которой образ \bar{x} относится к классу образов v_j , причем $pr_i \bar{x}$ определяет значение признака P_i для образа \bar{x} . Синтаксическое правило G , порождающее названия переменных, в данном случае тривиально, т.к. все термы атомарные, и заключается в присвоении функции принадлежности имени класса, который она представляет. Семантическое правило M представлено в виде алгоритма формирования функций принадлежности, который основан на концепции скользящего окна и является расширением подхода, используемого при построении гистограмм в статистике. Вид функций принадлежности зависит от выбора коэффициентов α и β , которые фактически определяют ширину скользящего окна и шаг скольжения и являются настраиваемыми параметрами алгоритма распознавания.

Нечетким портретом первого порядка S_j класса v_j определяется как совокупность значений лингвистических переменных, соответствующих классу v_j : $S_j = \{\mu_{ji}\}, i=1, \dots, n$. Для каждого нечеткого портрета S_j строится правило нечеткого вывода по определенным ранее лингвистическим переменным.

ПРАВИЛО « S_1 »:

ЕСЛИ « L_1 есть v_1 » И... И L_i есть v_1 И ... И L_m апои v_1 ТО Y есть (v_1 / μ_{v_1}) ;

...

ПРАВИЛО « S_k »:

ЕСЛИ « L_1 есть v_k » И... И L_i есть v_k И ... И L_m апои v_k ТО Y есть (v_k / μ_{v_k}) .

Принятие решения осуществляется на основе механизма нечеткого вывода. Следует отметить, что в качестве операции «И» на этапе агрегирования используется m -местная функция $f(a_1, a_2, \dots, a_m) = \log_2((a_1 + 1)(a_2 + 1) \dots (a_m + 1)) / m$, $a_i \in (0,1]$. Результатом работы алгоритма является нечеткое множество \tilde{y} .

Для решения задачи классификации нефтепродуктов обучающая выборка состояла из 870 элементов, шести классов $\{v_1, v_2, v_3, v_4, v_5, v_6\}$ и описывалась шестью признаками, такими как содержание ароматических углеводородов (Aroma), олефинов (Olf), бензолов (ArmBnz), ксилолов (ArmKs), октановое число (RON) и массовая доля МТБЭ (МТВЕ). На рис. 1 представлены лингвистические переменные для трех из рассматриваемых признаков (Olf, Aroma, ArmBnz) и их значения для некоторых классов, полученные при значениях коэффициентов $\alpha = 0.1$ и $\beta = 4$.

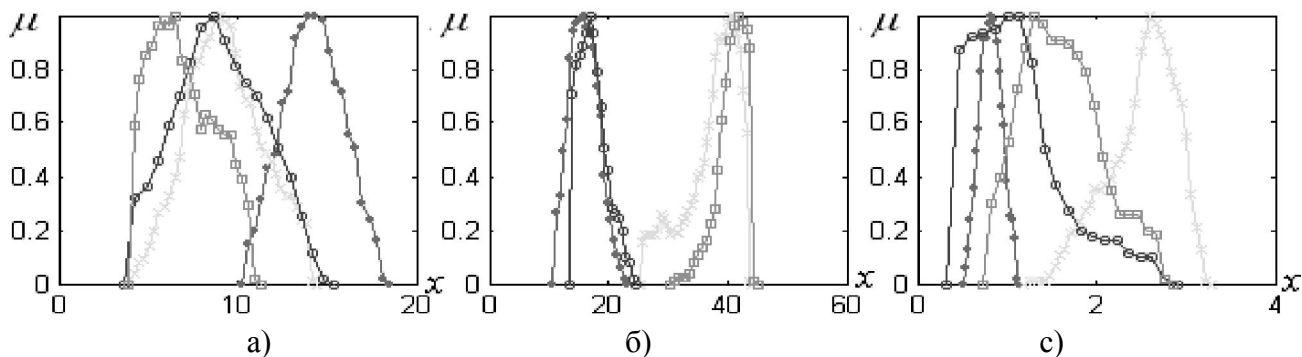


Рис. 1. Значения лингвистических переменных а) Olf, б) Aroma, с) ArBnz для классов образов v_1, v_2, v_3, v_4

На рис. 2 представлен результат работы алгоритма в виде столбиковых диаграмм для образцов $\bar{x}^{(1)} = (20, 25, 1.3, 0.93, 6)$ и $\bar{x}^{(2)} = (9, 40, 2.1, 97, 10)$. Следует отметить, что образец $\bar{x}^{(1)}$ однозначно является представителем класса образов v_1 , а класс образов, к которому относятся $\bar{x}^{(2)} = (20, 25, 1.3, 0.93, 6)$, определяется неоднозначно - v_3 или v_6 .

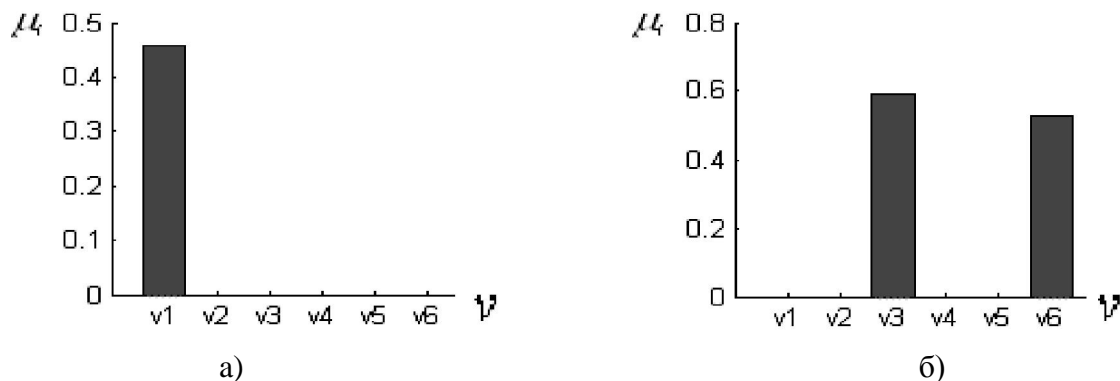


Рис. 2 Результат нечеткого вывода для образцов а) $\bar{x}^{(1)}$, б) $\bar{x}^{(2)}$

Адаптация и контроль качества алгоритма

В предлагаемом алгоритме этап адаптации осуществляется за счет выбора параметров α и β по результатам серии экспериментов. Поиск наилучшего решения затруднен отсутствием общепризнанных универсальных критериев качества решений. На практике для оптимизации небольшого числа параметров используют функционалы скользящего контроля [4]. Фактически методами скользящего контроля измеряется обобщающая способность метода обучения на заданной конечной выборке.

В работе использованы функционал полного скользящего

$Q_c(\mu, X^L) = \frac{1}{N} \sum_{n=1}^N v(\mu(X_n^l), X_n^k)$ и функционал среднего отклонения частоты ошибок на

контроле от частоты ошибок на обучении $Q_d(\mu, X^L) = \frac{1}{N} \sum_{n=1}^N (v(\mu(X_n^l), X_n^k) - v(\mu(X_n^l), X_n^l))$

алгоритма μ по конечной совокупности объектов X^L , где (X_n^l, X_n^k) , $n=1, \dots, N$ -

всевозможные разбиения выборки X^L на обучающую и контрольную, $L=l+k$; $v(\mu, X^L)$ -

частота ошибок алгоритма μ на обучающей выборке X^L . В результате серии экспериментов

для задачи распознавания ЖНП в качестве оптимальных параметров были определены

$\alpha = 0.1$ и $\beta = 4$, что обеспечило 92% верно распознанных образцов на тестовой выборке.

Кластеризация и повышение уровня распознавания

Как уже отмечалось ранее, структура обучающей выборки может быть достаточно сложной. В рассматриваемой задаче в результате визуального анализа диаграмм рассеивания, функций принадлежности нечетких портретов, а также консультаций с экспертом было определено наличие кластеров в некоторых классах образов, в частности для класса v_4 . Это обусловлено технологическими особенностями производства ЖНП. Для класса v_4 были определены два кластера и в базе знаний, соответственно, было определено два правила для данного класса. В результате такой модификации было повышено качество распознавания. На рис. 3 изображена ROC-кривая, для случая, когда в качестве позитивных примеров рассматриваются элементы класса v_4 , а в качестве отрицательных – все остальные примеры. Данная кривая ошибок показывает зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров. Качество распознавания полученного алгоритма возросло за счет верно распознанных элементов класса v_4 на 2%.

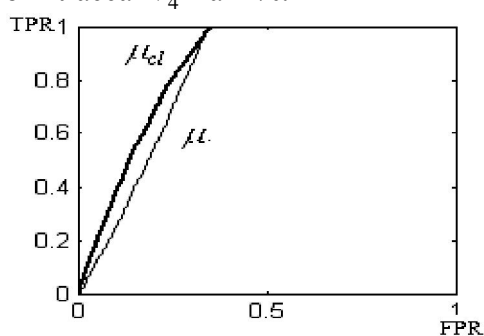


Рис. 3. ROC- кривая для класса v_4

Выводы

В работе была решена задача классификации ЖНП алгоритмом распознавания на базе нечетких портретов. На данном этапе в лаборатории контроля качества решение принимается профильными специалистами, однако предлагаемый подход позволит автоматизировать процесс и усовершенствовать работу лаборатории. Достоинством предложенного алгоритма является возможность интерпретации результата непосредственно экспертом в ситуациях, когда однозначного ответа не существует. Многие задачи в химической, пищевой промышленности могут быть сведены к описной выше модификации задачи распознавания образов.

В качестве модификации основного алгоритма было выполнено разбиение на кластеры некоторых классов образов, что позволило улучшить качество распознавания на 2% по сравнению с исходным алгоритмом (92% верно распознанных объектов).

Литература

1. Прикладная статистика: классификация и снижение размерности [Справ. изд.] / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, А. Д. Мешалкин // Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.: ил.
2. Никифоров И.К. Нейросетевой импедансный метод и устройства идентификации и определения параметров жидких нефтепродуктов: автореф. на соиск. уч. степени канд. тех. наук. – Казань, 2005. – 19 с.
3. Козловский В.А. Решение задачи распознавания по нечетким портретам классов / В.А. Козловский, А.Ю. Максимова // Искусственный интеллект. – 2010. – №4. – С. 221-228.

4. Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов. / К.В. Воронцов // Математические вопросы кибернетики. Вып. 13: Сборник статей под ред. О.Б. Лупанова. – М.: Физматлит, 2004. – С.5-36.