

# ПОСТРОЕНИЕ НЕЧЕТКОЙ СИСТЕМЫ ЛОГИЧЕСКОГО ВЫВОДА МАМДАНИ НА ОСНОВЕ АНАЛИЗА НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ ТЕСТОВОЙ ВЫБОРКИ

Силич М. П., Аксёнов С. В., Аксёнов А. В.

Кафедра автоматизации обработки информации, Томский государственный университет систем управления и радиоэлектроники

Национальный исследовательский Томский политехнический университет

Томск, Российская Федерация

E-mail: smp@muma.tusur.ru, axoenows@tpu.ru, image@tpu.ru

*Распространение систем управления, функционирующих на базе аппарата нечеткой логики в значительной степени ограничено сложностью генерации правил и определения терм-множеств входных и выходных аргументов системы. На практике создание базы нечетких правил и выбор функций принадлежности аргументов зависит от успешной агрегации знаний многих специалистов и многократных проверок модели. В работе предлагается алгоритм для автоматизации построения нечетких систем вывода Мамдани, который предполагает первоначальное исследование пространства входных и выходных переменных для выявления терм-множеств системы, формирования нечетких правил и подбора оптимальных функций принадлежности для каждого нечеткого множества.*

## ВВЕДЕНИЕ

Многие современные интеллектуальные системы, основанные на нечетких и нейронечетких вычислениях, демонстрируют свою эффективность и надёжность при решении многих слабоформализуемых задач[1]. Однако создание таких систем, основанных на знаниях, требует постулирования правил вывода человеком-экспертом, кроме того нахождение адекватных функций принадлежности (ФП) для всех терм-множеств входных и выходных переменных модели может быть произведено путем неоднократной настройки или обучения. Существующие методы, автоматизирующие построение систем вывода Мамдани, используют априорную информацию о числе термов для каждой переменной для разбиения диапазона возможных значений между нечеткими множествами, создания правил вывода со связями между переменными и правилами по принципу "каждого с каждым" и последующего обучения алгоритмами локальной или глобальной оптимизации[2]. Практически довольно сложно определить, какой тип ФП лучше использовать для описания конкретного терм-множества, тем более, что для получения результата используется набор правил вывода[3]. Другим подходом, рассматриваемым в предлагаемой работе, является автоматизация путем предварительного анализа обучающей выборки на предмет выделения кластеров в объединенном пространстве входных и выходных переменных. Каждый кластер будет представлять собой зависимость (или другими словами правило вывода), определяющую влияние входных аргументов на результат. Концентрация векторов, таким образом, будет означать наличие общих характеристик у векторов. Компоненты этих векторов можно в этом случае представить как объекты терм-множеств. Поиск типов ФП и их параметров

для таких множеств можно осуществить с учетом концентрации векторов внутри кластеров с помощью нечетких мер.

## I. НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ КАК ИНСТРУМЕНТ ВЫДЕЛЕНИЯ ТЕРМ-МНОЖЕСТВ И ПРАВИЛ ВЫВОДА

Согласно поставленной задаче построение системы нечеткого вывода заключается в поиске регионов концентрации обучающих примеров и выявлении взаимосвязи между терм-множествами входных и выходных переменных. Центр каждого кластера, обнаруженного в результате процедуры кластеризации, будет соответствовать центрам (L-R)-ФП терм-множеств переменных, описывающих эту группу точек. На рисунке 1 показан пример кластеризации нечеткими гиперэллипсами, заметим, что даже точки, не попавшие в регионы максимальной концентрации векторов, обладают вектором принадлежности к каждому из кластеров и окончательная форма каждого кластера учитывает распределение разрозненных векторов.

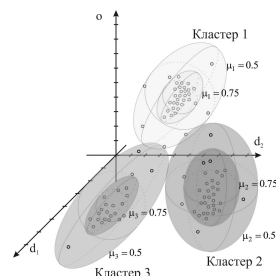


Рис. 1 – Пример кластеризации Густаффсона-Кесселя с указанием границ

На рисунке указаны области кластеров: «внутренние» более темные гиперэллипсы ограничивают области пространства с точками, принадлежащими к данному кластеру со степенью

принадлежности  $\mu_i > 0.75$ , а «внешние» - с  $\mu_i > 0.5$ . Использование нечетких гиперэллипсов по сравнению с гиперсферами выглядит более естественно, т.к. в большей степени отражает взаимозависимость между компонентами входа  $d_i$  и выходным значением  $o_i$ .  $\mu_i$

Проекции векторов кластера  $k$  на ось формируют набор значений  $\{y_i, f_i\}$ , где  $y_i$  - значение входного или выходного параметра, в зависимости от того, для какой переменной строится ФП, т.е.  $y_i = d_{ij}$  или  $o_i$ ,  $f_i$  - степень принадлежности к кластеру  $k$ . Набор  $\{y_i, f_i\}$  подходит к описанию дискретного нечеткого множества и теперь по этим значениям необходимо восстановить вид ФП. Система нечеткого вывода, постоянная по предлагаемому методу для рисунка 1 представлена на рис.2

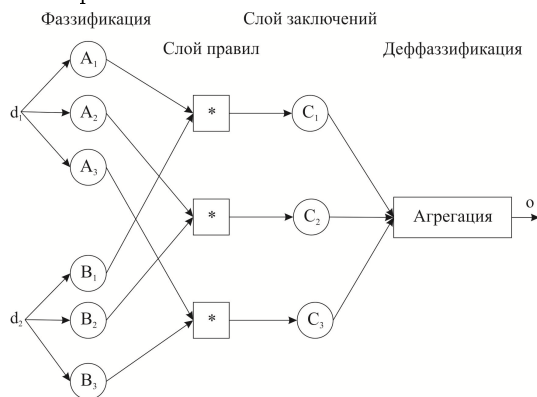


Рис. 2 – Нечеткая система вывода Мамдани для трех кластеров

На рисунке  $A_1, A_2, A_3$  - множества описывающие терм-множества для первой переменной,  $B_1, B_2, B_3$  - для второй,  $C_1, C_2$  и  $C_3$  - для выходной.

## II. НАСТРОЙКА ПАРАМЕТРОВ ФУНКЦИЙ ПРИНАДЛЕЖНОСТИ

В качестве типов ФП для системы используются стандартные (L-R)-функции (треугольная, трапецевидная, гауссова и колоколообразная, но предлагаемый подход не накладывает ограничений на вид ФП). Настройка заключается в поиске параметров ФП, выражающих степень группировки точек у центра и пологость функции. Задачи подобного типа, связанные с анализом групп разнородных объектов и определением их характеристик, достаточно эффективно решаются с использованием генетического алгоритма [4]. В качестве фитнес функции берется функция минимизации среднеквадратичной ошибки между значениями полученной ФП и нечетким значением принадлежности вектора, содержащего анализируемую компоненту к

нечеткому кластеру  $f_i$ . Хромосома представляет собой набор значений всех параметров ФП за исключением центра. Для нахождения оптимального решения генерируются несколько популяций для каждой из заданных пользователем типов ФП и запускается параллельно несколько генетических алгоритмов для каждого типа ФП. Среди всех полученных решений (для всех типов ФП) выбирается наилучшее решение, обладающее минимальным фитнесом. Однако полученные таким образом ФП, представляют собой только начальное приближение (хотя и довольно близкое к оптимальному), т.к. они были рассчитаны независимо друг от друга. Требуется дополнительная настройка параметров ФП на первоначальной выборке  $=\{d_i, o_i\}$ . Коллективная настройка параметров выполняется генетическим алгоритмом и для поиска решения было предложено взять хромосому, кодирующую все параметры всех ФП, использующихся в нечеткой системе за исключением центров ФП. Создается значительный объем популяции, в которой представлены различные ФП с разными параметрами, и запускается процедура естественного отбора. Значение фитнес-функции представляет собой среднеквадратичную ошибку реального выхода системы от значений  $o_i$ .

## III. ЭКСПЕРИМЕНТ

Оценка эффективности алгоритма произведена при решении задачи восстановления сложной поверхности, заданной уравнением  $z = \sqrt{x^2 + y^2} + 3 \cos(\sqrt{x^2 + y^2})$ . Входными значениями для алгоритма являлись значения 1000 точек, расположенных на поверхности  $f(x, y)$ , требовалось построить систему нечеткого вида типа Мамдани, позволяющую производить аппроксимацию функции по предложенному алгоритму и оценить ошибку аппроксимации модели. В результате алгоритм детектировал 34 нечетких кластера. Средняя квадратичная ошибка модели при тесте из 1000 точек, случайным образом взятых с идеальной поверхности, составила 7,4

## IV. СПИСОК ЛИТЕРАТУРЫ

1. Larose, D. T. Discovering Knowledge in Data / D. T. Larose // New York: John Wiley and Sons, 2005. — 217 p.
2. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федюлов. — М.: Горячая линия – Телеком, 2007. — 284 с.
3. Fuzzy Cluster Analysis / F. Hoepfner, F. Klawonn, R. Kruse, T. Runkler. — New York: John Wiley and Sons, 2000. — 288 p.
4. Data mining: Concepts and Techniques / J. Han, M. Kamber, J. Pei. — Waltham: Morgan Kaufmann, 2012. — 703 p.

Работа выполнена при финансовой поддержке гранта Министерства образования и науки в рамках госзадания «Наука» № 8.8113.2013 и гранта РФФИ № 13-07-00397а.