

# МОДЕЛЬ ФОРМИРОВАНИЯ ПАКЕТА УСЛУГ СОТОВОЙ СВЯЗИ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОЙ ЛОГИКИ

Н.Л. Додонова, А.С. Александров

Самарский государственный аэрокосмический университет им. ак. С.П. Королева  
(национально исследовательский университет)

Интенсивный рост объёма генерируемой информации, представленной в электронном виде, в том числе и данных деятельности абонентов сотовой связи, делает актуальным решение задачи разбиения множества абонентов на заранее неизвестное количество подмножеств с различными видами телекоммуникационной активности. Возникает задача формирования групп абонентов со схожими предпочтениями в области связи. Задачи такого типа относятся к задачам анализа больших массивов данных, которые приобрели большое значение с развитием вычислительных технологий.

Количество абонентов сети «МегаФон» и их параметров достаточно разнообразно и велико. Например, количество строк (фактов) в таблице одной только активности абонентов составляет 3 795 657 327 операций, без учёта различных справочников и только в поволжском филиале за один месяц.

На данный момент в Поволжье 17134754 действующих абонента. Очевидно, что для формирования пакета каждому абоненту потребуется много времени (расчёт стоимости каждого пакета, определение состава и др.). Другой, более серьёзной проблемой является само количество этих полученных пакетов, равное количеству абонентов. Поскольку поддержка актуальной информации о состоянии пакета (стоимость, состав) является достаточно трудоёмкой задачей. Для решения этой проблемы необходимо разбить исходное множество абонентов на группы «похожести» и уже каждой группе формировать пакет услуг.

В качестве алгоритма кластеризации данных был выбран алгоритм Fuzzy C-Means. На рисунке 1 представлена схема алгоритма кластеризации.

Нечёткое разбиение входного множества абонентов (векторов) осуществляется в ходе минимизации целевой функции:

$$J_{FCM}(X, k, m) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \rho^2(x_i, c_j).$$

В качестве функции расстояния в алгоритме используется евклидова мера расстояния:

$$\rho(x_i, x_j) = \sqrt{\sum_{t=1}^d (x_{it} - x_{jt})^2}.$$

Данный алгоритм применяется к таблице фактов абонентов («обучающее» множество).

В результате работы алгоритма мы получаем кластеры, которые являются нечёткими множествами и каждый абонент принадлежит различным кластерам с различной степенью принадлежности.

После получения нечётких кластеров необходимо выяснить наиболее важные характеристики абонентов в каждом кластере. Необходимо провести анализ значений характеристик в полученных кластерах, чтобы понять, почему абоненты попали в один кластер. Для выявления значимости характеристик по кластерам будем использовать алгоритм нечёткого вывод.

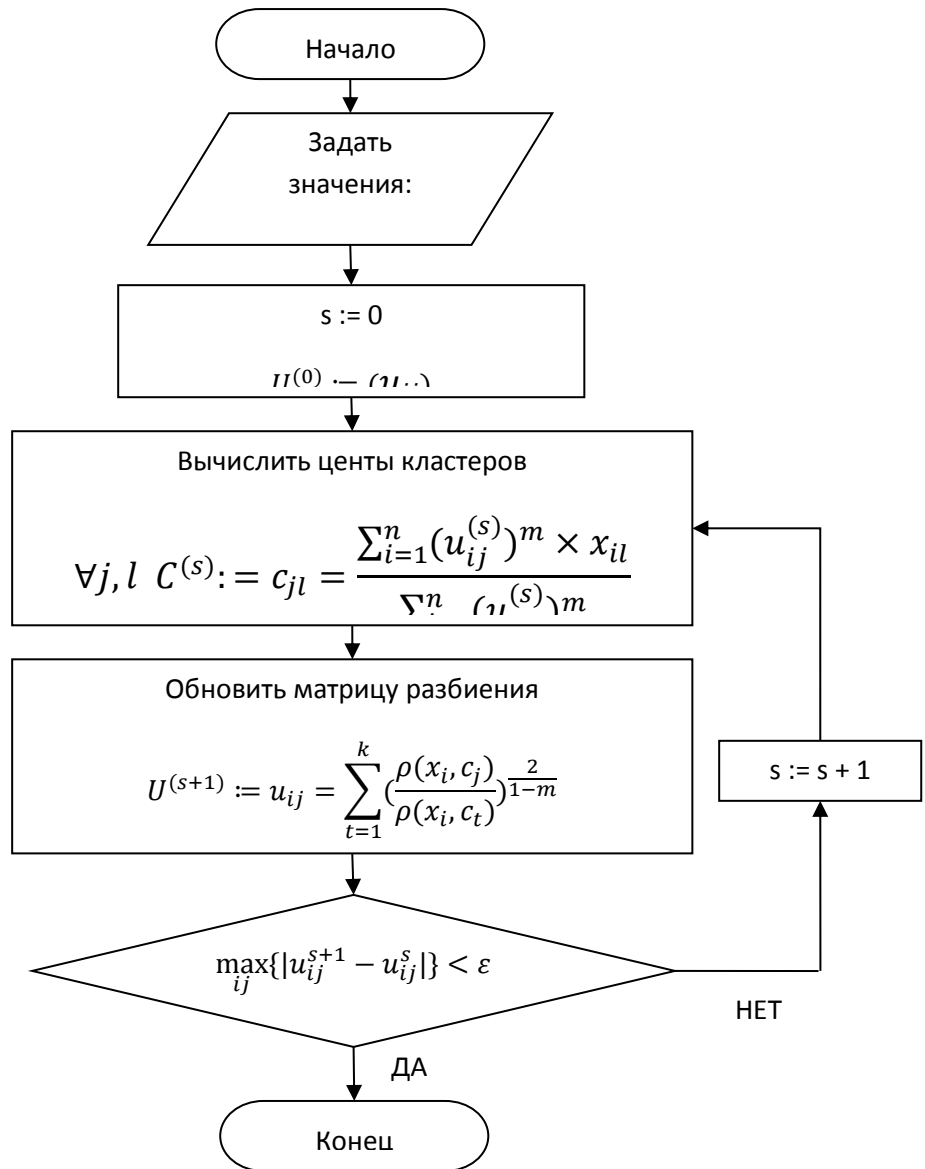


Рисунок 3.1 – Схема алгоритма кластеризации

В схеме используются следующие обозначения:

- $d \in \mathbb{N}$  – размерность пространства векторов данных;
- $n \in \mathbb{N}$  – мощность обучающей выборки;
- $X \subset \mathbb{R}^d$  – обучающая выборка векторов данных;
- $i \in \mathbb{N}: 1 \leq i \leq n$  – номер вектора обучающей выборки;
- $x_i \in X$  –  $i$ -й вектор выборки;
- $k \in \mathbb{N}$  – количество кластеров;
- $m \in \mathbb{R}: m > 1$  – коэффициент нечёткости;
- $\varepsilon \in (0,1) \subset \mathbb{R}$  – критерий остановки;
- $s \in \mathbb{N}$  – номер итерации;
- $U \subset \mathbb{R}^{n \times k}$  – матрица степеней принадлежности (разбиения), где  $u_{ij} \in \mathbb{R}$ ;
- $C \subset \mathbb{R}^{k \times d}$  – матрица, которая содержит центры кластеров;
- $c_j \in \mathbb{R}^d$  – центр кластера  $j$ , вектор размерности  $d$ ;
- $j \in \mathbb{N}: 1 \leq j \leq k$  – номер кластера;
- $l \in \mathbb{N}: 1 \leq l \leq d$  – номер координаты вектора;
- $\rho(x_i, c_j)$  – функция расстояния (мера сходства), определяющая степень принадлежности вектора;
- $x_{il}, c_{jl} \in \mathbb{R}$  –  $l$ -е координаты векторов  $x_i$  и  $c_j$  соответственно;

Выбираем один из кластеров и абонентов из этого кластера, степень принадлежности которых к выбранному кластеру больше чем к остальным. Таким образом, получаем чёткое множество, которое будем считать универсумом для рассматриваемого далее нечёткого множества. После этого необходимо вычислить наиболее значимые признаки в этом кластере. Алгоритм вычисления значимости признака строится на теории нечёткого вывода, поскольку абоненты, даже находящиеся в одном кластере, потребляют услуги не в одинаковом объёме и сделать вывод о составе формируемого пакета можно лишь приблизительно. С учётом количества абонентов универсума будем проводить анализ не для всех абонентов, а лишь для тех представителей универсума, которые относятся к ядру рассматриваемого далее нечёткого множества. После чего обобщим результат обработки на остальных абонентов универсума. В случае малого количества объектов в ядре можно расширить число анализируемых абонентов, заменив ядро нечёткого множества на  $\alpha$ -срез при подходящем значении  $\alpha$ . Подобную процедуру проводим для всех полученных кластеров.

Каждого абонента можно охарактеризовать набором характеристик  $(V_1, V_2, \dots, V_n)$ . Указанные характеристики показывают различные показатели потребляемых услуг связи (длительность разговоров, количество трафика и др.).

Оценка влияния определённой характеристики (параметра) на пакет, с точки зрения значимости при формировании его состава, может быть описана в терминах естественного языка. Поэтому будем рассматривать их как лингвистические переменные, со значениями из множества {«незначительное», «среднее», «значительное»}. Множество принимаемых значений может быть расширено с помощью известных процедур (операции концентрации, дивергенции) [26].

Будем использовать процедуру нечеткого вывода для оценки значимости параметра. На вход подаем информацию о количестве абонентов, для которых значение параметра находится в отрезке  $[\alpha\lambda_{max}; \lambda_{max}]$ , где  $\lambda_{max}$  – максимальное значение параметра, а  $\alpha \in [0; 1]$ . На выходе получаем информацию о значимости параметра в виде значений функции принадлежности термам {«незначительное», «среднее», «значительное»}.

В ходе анализа предметной области из шестисот видов «логических» вызовов были выбраны семь. Рассмотрим набор характеристик абонента из таблицы фактов:

- $V_1$ : количество отправляемых SMS;
- $V_2$ : количество отправляемых MMS;
- $V_3$ : продолжительность местных вызовов;
- $V_4$ : продолжительность вызовов по России;
- $V_5$ : продолжительность международных вызовов;
- $V_6$ : продолжительность вызовов на стационарные (местные);
- $V_7$ : объём интернет трафика (Мбайт);
- $V_8$ : общая стоимость.

Зафиксируем характеристику  $V_i$ . Рассматривая количество абонентов универсума, для которых значение характеристики  $V_i - \lambda$  лежит на отрезке  $[\alpha\lambda_{max}; \lambda_{max}]$ , можем сказать, что их «малое количество», «среднее количество» или «большое количество». Далее под количеством абонентов будем понимать количество абонентов универсума со значением характеристики из указанного выше промежутка. Таким образом, получаем лингвистическую переменную  $X$  – количество абонентов. Лингвистическая переменная  $X$  может принимать значения, описываемые терм-множеством. Терм-множество имеет вид  $T = \{\tau_1, \tau_2, \tau_3\}$ , где  $\tau_1$  – малое количество абонентов,  $\tau_2$  – среднее количество абонентов,  $\tau_3$  – большое количество абонентов.

После анализа количества абонентов о каждой характеристике можно сказать, что она значительно влияет на состав будущего пакета, оказывает среднее влияние или незначительное. Таким образом, получаем лингвистическую переменную  $Y$  – степень влияния зафиксированной характеристики на состав формируемого пакета. Терм-множество лингвистической переменной  $Y$  имеет вид  $T = \{\tau_{1*}, \tau_{2*}, \tau_{3*}\}$  где  $\tau_{1*}$  – значительное влияние,  $\tau_{2*}$  – среднее влияние,  $\tau_{3*}$  – незначительное влияние характеристики в формировании пакета.

Очевидно, что характеристики  $(V_1, V_2, \dots, V_n)$ , имеющие значительное влияние станут основными в формирующемся пакете услуг, имеющие же незначительное влияние – могут стать бонусными.

Предложенная модель положена в основу автоматизированной системы формирования пакетов услуг связи для абонентов сети «Мегафон».

#### **Литература**

- 1 Телекоммуникационные технологии. Введение в технологии GSM / С. Б. Макаров, Н. В. Певцов, Е. А. Попов и др. – М.: Академия, 2008. –256 с.
- 2 Лапушкин К. Годовой отчёт МТС [Электронный ресурс] //ОАО «Мобильные ТелеСистемы» [сайт]. [2014]. URL: <http://www.report2013.mts.ru/ru/position.html> (дата обращения: 12.05.2014).
- 3 Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – М.: Горячая линия-Телеком, 2007. -284 с.
- 4 Додонова Н.Л., Кукольникова Е.А. Процедура нечеткого вывода состава региональных управляемых промышленных кластеров //Вестник поволжского государственного университета сервиса. Вып. 1(33). –Тольяти: ПВГУС, 2014. –С.52-57.