

НОВЫЕ ПОДХОДЫ К СОЗДАНИЮ ДИНАМИЧЕСКИХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Комарцова Л.Г.

МГТУ им. Н.Э. Баумана (Калужский филиал)
Россия, 248003, г. Калуга, ул. Баженова, д.2, e-mail: komlg@bmstu-kaluga.ru

Во многих реальных системах, связанных с интеллектуальной обработкой информации, необходимо обеспечивать эту обработку в режиме on-line. Основные проблемы, которые необходимо решать в этом случае, связаны с выбором соответствующего метода кластеризации. Наиболее известным методом кластеризации является k-means, который находит k различных групп данных (кластеров) и их кластерных центров как среднее векторов данных, находящихся внутри кластеров. Эта процедура минимизирует расстояние каждого вектора до его ближайшего кластерного центра. Обычно это достигается путем значительного числа итераций.

On-line версия k-means алгоритма основывается на априорном знании о распределении входных данных. Для каждого нового вектора обновляется только ближайший кластерный центр. Это правило известно как «локальный» k-means (или победитель забирает все), который может работать с данными, поступающими в интеллектуальную систему непрерывно.

Другие известные методы кластеризации DSA (Dynamic Clustering Algorithm, FCMC (Fuzzy C-means Clustering хотя и являются эффективными методами, работают в режиме off-line и требуют, как и алгоритм k-means, заранее определенного числа кластеров.

Менее известным является эволюционный кластерный метод (Evolving Clustering Method -ECM), который обеспечивает кластеризацию на протяжении всего жизненного цикла системы при непрерывном поступлении входных данных. Он предназначен для on-line кластеризации. Разработан комбинированный алгоритм, который включает специальную методику для уточнения кластерных центров при обучении в режиме off-line, что позволяет проводить смешанное обучение при функционировании ИС.

Основное достоинство алгоритма заключается в том, что, в отличие от других методов кластеризации, число кластеров заранее не определяется и может меняться в режиме on-line по мере поступления входных данных. Ограничение заключается в том, что для каждого кластера ищется максимальное расстояние между примером и ближайшим кластерным центром, и это расстояние не может быть больше заданной величины Dist. Однако в процессе эволюции (в режиме on-line) эта величина может меняться в зависимости от текущей ошибки кластеризации.

Проведенные эксперименты показали высокую эффективность нейросетевого классификатора на основе комбинированного алгоритма: этот классификатор быстрее обучается по сравнению другими алгоритмами и дает меньший процент ошибок классификации.