

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАСТЕРИЗАЦИИ ДЛЯ ЗАДАЧ «УМНОГО ГОРОДА»

Е.Ю. Юшков, 4 курс бакалавриата, Факультет информационных технологий и программирования, Университет ИТМО, г. Санкт-Петербург

Научный руководитель – Н.Н. Тесля, к.т.н., доцент, Факультет информационных технологий и программирования, Университет ИТМО, г. Санкт-Петербург

Особый интерес к методам анализа данных возник с развитием средств сбора и хранения данных, позволившим накапливать большие объемы информации. Среди методов анализа данных особое место занимают классификация и кластеризация. Классификация, при известной заранее группировке данных на подмножества (классы), устанавливает закономерность, по которой объекты группируются именно таким образом. Кластеризация же, основываясь на установленном отношении схожести элементов, устанавливает подмножества (кластеры), в которые группируются входные данные. Главное отличие кластеризации от классификации состоит в том, что перечень групп четко не задан и определяется в процессе работы алгоритма.

Кластеризация (кластерный анализ) - это метод обнаружения и разбиения множества связанных между собой объектов на группы, называемые кластерами. Внутри каждой группы должны оказаться похожие объекты, а объекты разных группы должны быть как можно более различны. Кластеризация нашла свое применение в разных областях. Интернет магазины, которые отслеживают историю покупок, могут использовать эту информацию для обнаружения групп клиентов со схожим потребительским поведением. Кластерный анализ применим даже в медицине, например, это систематизация болезней или их симптомов, лекарств и так далее. Кластеризация также применяется в социологических исследованиях. Допустим, на основе собранных данных о пользователях социальной сети посредством кластерного анализа можно выделить близкие по контексту группы пользователей. Или на основе проведенного опроса целевой аудитории по выбранным критериям можно выявить отдельные социальные группы.

Целью данной работы является изучение и сравнение методов и алгоритмов кластеризации данных для дальнейшего их применения в «умном городе».

Одна из основных задач концепции «умного города» предполагает разработку современных семантических сервисов и внедрение решений, ориентированных на интересы и потребности жителей. В настоящее время сервисы оказывают большое влияние на повседневную жизнь человека и позволяют эффективно использовать городское пространство. Для того чтобы определенный человек мог получить конкретную персональную рекомендацию по тому, к какому сервису обратиться, необходимо знать к какой группе (кластеру) жителей он относится. В «умном городе» кластерный анализ позволит обнаруживать и распределять жителей на отдельные подмножества со схожими признаками, для того чтобы в дальнейшем была возможность осуществлять рекомендации.

В представлении «умного города» во главе стоит термин «цифровая личность», который позволяет отразить личную и публичную информацию о человеке, появившуюся в результате широкого использования информации. «Цифровая личность» - это набор сведений в электронном виде, идентифицирующих объект, обладающих определенными атрибутами, предпочтениями и характерными особенностями, достаточный для проведения транзакций. Онтологическая модель «цифровой личности» представляет собой взаимосвязь трех классов атрибутов: неотъемлемые - являются врожденной частью сущности и не определены связями с внешними объектами, накапливаемые - накапливаются и изменяются с течением времени, и присвоенные - закреплены за сущностью, но не связаны с его внутренним характером. На

основе атрибутов онтологической модели сформирована спецификация параметров «цифровой личности», по которым можно осуществлять кластеризацию жителей умного города.

В результате работы было проведено аналитическое исследование методов кластеризации данных, выделены преимущества и недостатки использования данного подхода. На основе разработанной онтологической модели цифровой личности были выделены параметры и признаки жителя «умного города», на основе которых можно осуществлять кластеризацию.

Студент ФИТиП Юшков Е.Ю. _____

Научный руководитель: доцент ФИТиП Тесля Н.Н. _____