

ИССЛЕДОВАНИЕ ГРУПП КЛИЕНТОВ ЛОГИСТИЧЕСКОЙ ОРГАНИЗАЦИИ С ПРИМЕНЕНИЕМ МЕТОДОВ КЛАСТЕРИЗАЦИИ



**ЕКАТЕРИНА
ТИТОВА,**
РЭУ им. Г.В. Плеханова, факультет
маркетинга,
L.L.M.
аспирантка



**ОЛЬГА
ДЯТЛОВА,**
РЭУ им. Г.В. Плеханова,
факультет маркетинга,
аспирантка

Деятельность любой компании, независимо от сферы, в которой она существует, всегда подразумевает взаимодействие со своими клиентами. Такие отношения носят как объективный, так и чисто субъективный характер. До некоторого момента как в научной литературе, так и на практике большее внимание уделялось вопросу разработки такого процесса взаимодействия с клиентами, когда все параметры выполняются эффективно, вовремя, клиент удовлетворен исполнением договорных обязательств, перечисление денежных средств по таковым обязательствам произведено своевременно, и, соответственно, компания получила свою долю прибыли.

Это очевидное правило действует и сейчас, однако теперь большее внимание уделяется имиджу компании, и на смену объектно-ориентированному приходит субъектно-ориентированный подход, который становится все более индивидуально-ориентированным. Естественно, возникает проблема, как учесть огромное число факторов, которые влияют на взаимодействие клиента с компанией и прогнозировать их дальнейшие отношения. Именно на данном этапе становится актуальным применение метода кластеризации — одной из задач Data Mining для отслеживания клиентских групп, о которой в дальнейшем будет идти речь.

В настоящее время на базе методики Data Mining, в которой получил распространение информационный подход к анализу, создается большинство прикладных аналитических решений в бизнесе и многих других областях.

Термин Data Mining имеет в англоязычной среде несколько определений.

Data Mining дословно переводится как «добыча данных» или «раскопка данных» и определяется как *обнаружение в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретаций знаний, необходимых для принятия решений в различных сферах человеческой деятельности.* [2, с. 42]

Зависимости и шаблоны, найденные в процессе применения методов Data Mining, должны быть нетривиальными и ранее неизвестными, знания должны описывать новые связи между свойствами, предсказывать значения одних признаков на основе других.

Data Mining — это не один метод, а совокупность большого числа различных методов обнаружения знаний. Существует несколько условных классификаций задач Data Mining, среди которых рассмотрим четыре базовых.

Классификация — это установление зависимости *дискретной выходной переменной* от входных переменных.

Регрессия — это установление зависимости *непрерывной выходной переменной* от входных переменных.

Кластеризация — это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов. Объекты внутри кластера должны быть похожими друг на друга и отличаться от тех, которые вошли в другие кластеры.

Ассоциация — выявление закономерностей между связанными событиями. Примером такой закономерности служит правило, указывающее, что из события X следует событие Y. Такие правила называются *ассоциативными*.

В дальнейшем исследовании мы будем использовать только некоторые из них, а именно, кластеризацию и классификацию.

Помимо указанных выше задач, также зачастую выделяют следующие: *анализ отклонений* (deviation detection), *связей* (link analysis), *отбор значимых признаков* (feature selection). [4]

В отличие от задачи регрессии, в классификации на выходе мы имеем дискретную переменную, которую называют меткой класса. В решении задачи классификации заранее известно все множество классов, к которым может относиться объект, и, по своей сути, решением задачи классификации и является поиск класса, к которому относится объект. Опреде-

АННОТАЦИЯ:

В статье дается определение и классификация задач Data Mining, дается определение и приводится классификация алгоритмов кластеризации. Определяются условия и общая суть кластерного анализа данных применительно к логистической компании.

КЛЮЧЕВЫЕ СЛОВА:

Data Mining, кластеризация, логистика, лояльность клиентов.юс

ANNOTATION:

In this article we define what Data Mining is and arrange its goals; we give a definition of clusterization and classify its algorithms. We also define conditions and the essence of cluster data analysis in terms of logistics company.

KEYWORDS:

Data Mining, cluster analysis, logistics, customer loyalty.

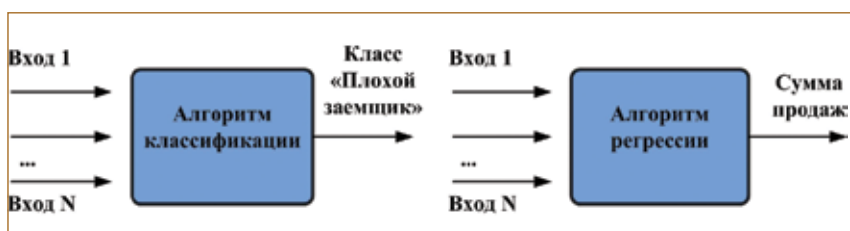


Рис. 1. Иллюстрация задачи классификации (слева) и задачи регрессии (справа)

ление класса осуществляется на основе анализа имеющихся свойств объекта. В свою очередь, в задаче регрессии получаемая переменная оказывается непрерывной (пространство действительных чисел). На рис.1 проиллюстрированы примеры задач классификации и регрессии. [2, с. 43]

Непосредственное применение кластерного анализа в общем виде представляет собой следующие этапы:

- отбор выборки объектов для проведения кластеризации;
- определение множества переменных, по которым происходит оценка объектов в выборке. В случае необходимости — нормализация значений переменных;
- процесс вычисления значений меры сходства между объектами;
- использование метода кластерного анализа для создания групп сходных объектов (кластеров);
- рассмотрение результатов анализа. [1]

Классификацию алгоритмов кластеризации можно свести к нижеследующим:

I. Иерархические и плоские.

Иерархические алгоритмы (алгоритмы-таксоны) строят целую систему вложенных разбиений выборки на непересекающиеся кластеры. На выходе мы получаем дерево кластеров, корнем которого будет вся выборка, а листьями — наиболее мелкие кластеры.

Плоские алгоритмы строят только одно разбиение объектов на кластеры.

Табл. 1

Сравнительная таблица алгоритмов

Алгоритм кластеризации	Форма кластеров	Входные данные	Результаты
иерархический	произвольная	число кластеров или порог расстояния для усечения иерархии	бинарное дерево кластеров
минимальное покрывающее дерево	произвольная	число кластеров или порог расстояния для удаления ребер	древовидная структура кластеров
последняя кластеризация	произвольная	последовательность порогов расстояния	древовидная структура кластеров с разными уровнями иерархии
выделение связанных компонент	произвольная	порог расстояния R	древовидная структура кластеров
c-средних	гиперсфера	число кластеров, степень нечеткости	центры кластеров, матрица принадлежности
k-средних	гиперсфера	число кластеров	центры кластеров

II. Четкие и нечеткие.

Четкие (непересекающиеся) алгоритмы ставят каждому объекту выборки в соответствие номер кластера, а именно, каждый объект принадлежит только одному кластеру.

Нечеткие (пересекающиеся) алгоритмы ставят в соответствие каждому объекту набор вещественных значений, определяющих степень отношения объекта к кластерам. В данном

случае каждый объект относится к каждому кластеру с некоторой вероятностью. [4]

Далее, приведем сравнительную таблицу алгоритмов кластеризации [1].

Далее перейдем к рассмотрению задачи кластеризации. Отличие ее от задачи классификации заключается в том, что на выходе мы имеем только кластеры (массивы похожих объектов, в которые нужно разместить все имеющееся множество данных), причем количество кластеров может быть неизвестным, а выходная переменная отсутствует. В итоге отмечается лишь схожесть объектов. Если необходимо объяснить полученный результат (образовавшиеся кластеры), потребуется его дополнительный анализ и интерпретация (рис. 2)

Эти задачи имеют значительный спектр применения в экономике и иных сферах. Попробуем указать наиболее известные способы.

Как правило, классификацию применяют в тех случаях, когда изначально известны классы. Это может быть анализ и отнесение к той или иной товарной группе нового продукта или разработки. Аналогична ситуация, если стоит задача о новом клиенте, которого необходимо отнести к той или иной потребительской группе, или же группе риска, если задача решается в контексте логистической компании. [2]

Если необходимо выявить связи и зависимости между различными факторами, применяется задача регрессии. Одним из ярких примеров являются «объемы продаж» - зависимая величина, на которую влияет огромное количество факторов: объемы продаж в прошлом, скачки курса валют, общая экономическая ситуация, деятельность конкурирующих организаций и т.п. Другой пример, при кредитовании встает вопрос о вероятности возврата. В этом случае необходимо учесть личные качества человека, наличие у него движимого и недвижимого имущества, сферы его жизнедеятельности и др.

Если у вас возросли объемы продаж и количество клиентов неуклонно увеличивается, становится очень сложно, а зачастую и вовсе невозможно, применить индивидуальный подход к каждому клиенту. В этом случае стоит обратиться к задаче кластеризации и разделить всех клиентов на различные группы. Каждая группа формируется на основе однотипных признаков, что позволит улучшить качество услуг. Такими признаками могут являться географическое расположение клиента, сфера его деятельности и т.п. После осуществления кластеризации достаточно легко выявить более и менее активные сегменты, выяснить какие из них убыточны, а какие — прибыльны.

Решение всех вышеназванных задач осуществляется с помощью алгоритмов и методов Data Mining. Данная методика сформировалась на основе множества дисциплин, в том числе математики, информатики, статистики, программирования.

На рисунке 3 проиллюстрировано, какого плана задачи можно решать с помощью Data Mining. [2, с. 46]

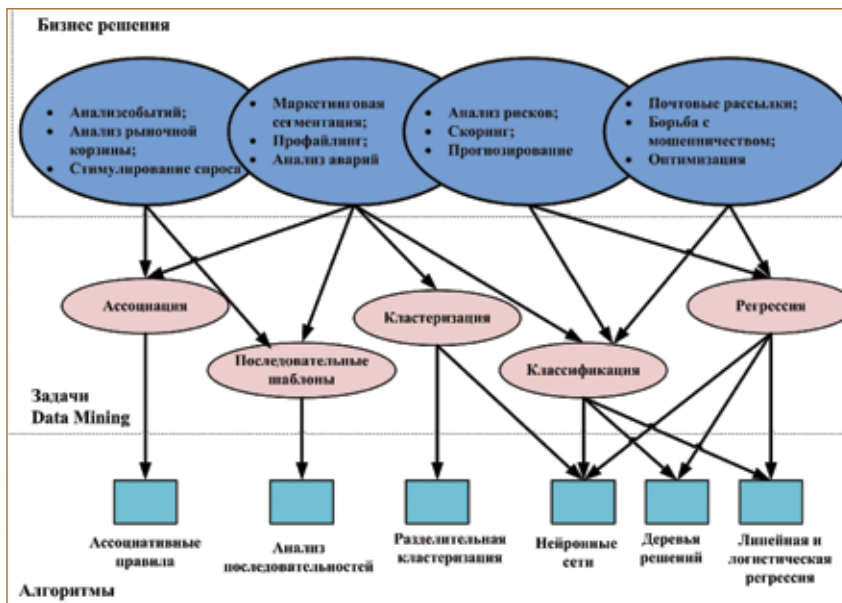


Рис. 3. Переход от реальной задачи к модели Data Mining

В процессе использования Data Mining неизбежно обнаруживаются новые свойства объектов и новые закономерности и процессы. Модель в методе Data Mining должна самостоятельно приобрести все необходимые свойства для максимально удобного отображения выявленных в имеющемся объеме данных всех присущих им закономерностей, причем в большинстве случаев они новые и ранее неизвестные. Информационный подход позволяет использовать комплекс методов (т.н. машинное обучение), которые обеспечивают возможность создания таких моделей — обучаемых моделей.

Одной из задач Data Mining является кластеризация. Соответственно, группа похожих объектов — кластер. Существует очень большое количество всевозможных определений кластеризации, далее приведем лишь несколько.

1) Кластеризация — это группировка объектов на основе близости их свойств; каждый кластер состоит из схожих объектов, а объекты разных кластеров существенно отличаются;

2) Кластеризация — это процедура, которая любому объекту X ставит в соответствие метку кластера Y . [4]

Кластеризацию, как правило, применяют в тех случаях, когда отсутствуют сведения о том, на какие классы/группы нам следует распределить имеющийся набор данных, либо же если объем информации настолько велик, что ручной анализ объектов достаточно затруднен. Необходимо отметить, что задача кластеризации неоднозначна по ряду причин:

- неизвестно количество кластеров
- определение близости объектов, уровень их схожести зачастую слишком субъективны. [5]

Задача кластеризации известна давно, и специалисты в различных областях оперируют рядом других терминов — таксономия, сегментация, группировка, самоорганизация. В Data Mining употребляется термин «кластеризация».

Остановимся на целях кластеризации и на ее применении в бизнес-аналитике. Цели кластеризации в Data Mining могут быть различными и зависят от конкретной решаемой задачи. Рассмотрим эти задачи.

Изучение данных. Разбиение множества объектов на группы помогает выявить внутренние закономерности, увеличить наглядность представления данных, выдвинуть новые гипотезы, понять, насколько информативны свойства объектов.

Облегчение анализа. При помощи кластеризации можно упростить дальнейшую обработку данных и построение моделей: каждый кластер обрабатывается индивидуально, и модель создается для каждого кластера в отдельности. В этом смысле кластеризация может рассматриваться как подготовительный этап перед решением других задач Data Mining: классификации, регрессии, ассоциации, последовательных шаблонов.

Сжатие данных. В случае, когда данные имеют большой объем, кластеризация позволяет сократить объем хранимых данных, оставив по одному наиболее типичному представителю от каждого кластера.

Прогнозирование. Кластеры используются не только для компактного представления объектов, но и для распознавания новых. Каждый новый объект относится к тому кластеру, присоединение к которому наилучшим образом удовлетворяет критерию качества кластеризации. Значит, можно прогнозировать поведение объекта, предположив, что оно будет схожим с поведением других объектов кластера.

Обнаружение аномалий. Кластеризация применяется для выделения нетипичных объектов. Эту задачу также называют обнаружением аномалий (outlier detection). Интерес здесь представляют кластеры (группы), в которые попадает крайне мало объектов, скажем, один-три. [1]

Первые две задачи наиболее популярны в бизнес-аналитике.

Применение кластеризации не так просто и однозначно, как может показаться на первый взгляд. Если вам нужно рассортировать картошку на большую и маленькую, человек часто будет стоять перед дилеммой, в какую корзину положить клубень среднего размера. Автоматизированный процесс сортировки, конечно, не столь мучителен, зато нетривиальная задача встает перед аналитиком, который выбирает и программирует алгоритм: итоговый вариант кластеризации будет существенно отличаться в зависимости от выбора варианта оценки близости между свойствами объекта. В Data Mining используется метрическая система сравнения, другими словами, это оценка расстояния - близости между объектами. Также, необходимо помнить, что кластеризация сама по себе не дает результатов. Она лишь распределяет массив данных по кластерам, а для получения необходимого эффекта требуется последующий анализ и интерпретация каждого кластера. Необходимо присвоить каждой ячейке (каждому кластеру) свое название, которое будет в полной мере отражать его суть для дальнейшего анализа. Затем необходимо детально и подробно проанализировать каждый кластер: выявить его мощность (количество объектов, которые по результатам кластеризации попали в него) и иные показательные характеристики.

Целью данного исследования является выявление групп клиентов, оптимально реагирующих на изменения логистической компании, анализ и прогнозирование поведения как отдельно взятых клиентов, так и групп в целом, обеспечение возможного контроля происходящих в компании бизнес-процессов в условиях выявленной динамики этих групп и отдельно взятых клиентов.

Задачи исследования состоят в проведении кластеризации по отношению к клиентам, выявлении динамики и прогнозирования групп в целом, отдельно взятых групп и поведения каждого отдельно взятого клиента и описании соответствующих выводов по тому, какие изменения необходимы для обеспечения более высокой лояльности и какая коррекция бизнес-процессов компании необходима, в зависимости от изменений, происходящих в группах.

Особенности логистической компании заключаются в том, что существует несколько отделов, отвечающих за те или иные функции: *клиентский, юридический, бухгалтерия и операционный*. Причем особенностью клиентского отдела является то, что клиент обслуживается одним менеджером — так называемым, *ключевым менеджером*, который принимает решение о работе с клиентом на основе данных о клиенте по другим отделам, на основе субъективных и объективных характеристик по клиенту. В этом смысле будем рассматривать два типа лояльности: *объективную и субъективную*. Под **лояльностью** в целом будем понимать поведение клиента, характеризующееся неким набором показателей его отношений с компанией. **Объективной лояльностью** будем называть те характеристики по клиенту, которые объективно отражают его взаимодействие с компанией:

1. Средние выплаты по договорам на одного клиента;
2. Количество заключенных договоров на дополнительные услуги по одному клиенту;
3. Количество обычных договоров на одного клиента;
4. Реакция клиента на различные изменения;
5. Готовность клиента рекомендовать компанию другим людям.

Под **субъективной лояльностью** понимаем отношение клиента к определенным аспектам, таким как:

1. Скорость проведения платежей (бухгалтерский отдел);
2. Скорость проведения документации (юридический отдел);
3. Работа менеджера (клиентский отдел);
4. Оперативность работы (оперативный отдел);
5. Адекватность стоимости услуги;
6. Информация об услуге.

Все вышеперечисленные характеристики можно расширить и дополнить, в зависимости от того, что необходимо исследовать, какие результаты получить, и насколько подробная информация по клиентам необходима. Также стоит

отметить, что оценка по всем представленным пунктам должна проходить по некой общей шкале. Существуют так называемые качественные методы анализа данных, по которым оцениваются различные субъективные моменты, которые возникают в любых бизнес-процессах, а особенно когда речь идет о лояльности.

Но не будем останавливаться на этом, отметим лишь необходимость адекватного сопоставления объективных и субъективных факторов и выявления по ним некой общей шкалы.

В результате кластеризации необходимо:

1. Чтобы каждый кластер описывался неким набором параметров, отражающих лояльность клиентов в данной группе.
2. Получить вероятность принадлежности каждого -го клиента к каждому полученному кластеру. Это требуется для выявления динамики по клиентам и по группам, а также для возможности отслеживания VIP-клиентов.

Для этого логичнее всего использовать алгоритм нечеткой кластеризации. (За дополнительными расчетами вы можете обратиться к авторам этой статьи.)

Стоит отметить популярность использования различных методов кластеризации для анализа данных. Это объясняется относительной легкостью их использования за счет наличия специальных программ, которые с помощью нескольких итераций и малыми затратами по времени позволяют с максимальной точностью предоставить решение постоянно возникающих в любой компании наиболее популярных бизнес-задач.

Библиографический список

1. Гитис Л. Х. Кластерный анализ в задачах классификаций, оптимизации и прогнозирования. — МГГУ, 2001
2. Мандель И. Д. Кластерный анализ. — М: Статистика и Финансы, 1988
3. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям (+ CD): учеб. пособие. 2-е изд., перераб. и доп. — СПб.: Питер. — 2010. — 704 с.: ил.
4. Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl. Cluster Analysis, 5th Edition. — Wiley; 2011
5. Sigmund Halpern. The Assurance Sciences: An Introduction to Quality Control and Reliability. — Prentice Hall; Facsimile edition, 1978.



ООО «Фрегат»

www.fregatt.ru

Склады ответственного хранения

- погрузочно-разгрузочные работы
- обработка грузов в круглосуточном режиме
- хранение товаров
- инвентаризация
- обработка товара по специальным требованиям заказчика
- сортировка и переупаковка товара
- маркировка и стикеровка товара

Сдача в аренду складов класса А, В, С

124460, Россия, Москва, г. Зеленоград, 2-ой Западный пр-д, д. 3, стр. 1
 Телефоны: (495) 981-85-14, (495) 755-27-10, электронный адрес: otvetsklad@fregatt.ru