

## **О ПРИМЕНЕНИИ ТЕХНОЛОГИЙ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ОЦЕНКИ СПРОСА ЮВЕЛИРНЫХ УКРАШЕНИЙ**

**Чеблакова Е.А.**

**Научный руководитель д-р техн. наук Ченцов С.В.**

***Сибирский федеральный университет, Институт космических и информационных технологий***

*В работе предлагается использовать интеллектуальный анализ данных для повышения спроса на ювелирные украшения и повышения собственной прибыли ювелирного предприятия.*

### **Введение**

По мере того, как компании собирают миллионы информационных показателей о своих клиентах и пытаются найти способ эффективно использовать эти данные для повышения собственной прибыли, интеллектуальный анализ данных (data mining) становится все более популярным технологическим направлением.

Применение существующих программных продуктов на предприятии не позволяет выделить некоторые правила, взаимосвязи или тенденции в больших наборах данных. Все они имеют определенные ограничения и некоторые недостатки.

Все это приводит к необходимости поиска новых подходов. Одним из таких подходов могло бы стать использование интеллектуальных информационных технологий, интенсивно развиваемых последние два десятилетия.

Для оценки спроса ювелирных изделий необходимо создать методику, на основании которой станет возможным оценить:

1. какие типы изделий пользуются спросом,
2. какой стиль исполнения изделий предпочтителен,
3. спрос на популярные виды изделий в ближайшее время;
4. модные тенденции и актуальность изделий.

### **1. Математические методы кластеризации**

Одной из наиболее важных задач при применении кластерного анализа в маркетинговых исследованиях является анализ поведения потребителя, а именно: группировка потребителей в однородные классы для получения максимально полного представления о поведении клиента из каждой группы и о факторах, влияющих на его поведение. Эта проблема подробно описана в работах Клакстона, Фрая и Портиса (1974), Киля и Лэйтона (1981).

Важной задачей, которую может решить кластерный анализ, является позиционирование, т.е. определение ниши, в которой следует позиционировать новый продукт, предлагаемый на рынке. В результате применения кластерного анализа строится карта, по которой можно определить уровень конкуренции в различных сегментах рынка и соответствующие характеристики товара для возможности попадания в этот сегмент. С помощью анализа такой карты возможно определение новых, незанятых ниш на рынке, в которых можно предлагать существующие товары или разрабатывать новые.

#### **1.1. Формальная постановка задачи**

Дано — набор данных со следующими свойствами:

- каждый экземпляр данных выражается четким числовым значением;
- класс для каждого конкретного экземпляра данных неизвестен.

Найти:

- способ сравнения данных между собой (меру сходства);

- способ кластеризации;
- разбиение данных по кластерам.

Формально задача кластеризации описывается следующим образом.

Дано множество объектов данных  $I$ , каждый из которых представлен набором атрибутов. Требуется построить множество кластеров  $C$  и отображение  $F$  множества  $I$  на множество  $C$ , т. е.  $F : I \rightarrow C$ .

Отображение  $F$  задает модель данных, являющуюся решением задачи. Качество решения задачи определяется количеством верно классифицированных объектов данных.

Множество  $I$  определим следующим образом:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\} \quad (1)$$

где  $i_j$  — исследуемый объект.

Каждый из объектов характеризуется набором параметров:

$$i_j = \{x_1, x_2, \dots, x_h, x_m\} \quad (2)$$

В нашем случае, такими параметрами являются тип изделия, стиль исполнения, проба, сплав и др.

Каждая переменная  $x_h$  может принимать значения из некоторого множества:

$$x_h = \{v_h^1, v_h^2, \dots\} \quad (3)$$

В данном примере значениями являются действительные числа.

Задача кластеризации состоит в построении множества:

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\} \quad (4)$$

Здесь  $c_k$  — кластер, содержащий похожие друг на друга объекты из множества  $I$ :

$$c_k = \{i_j, i_p \mid i_j \in I, i_p \in I \text{ и } d(i_j, i_p) < \sigma\}, \quad (5)$$

где  $\sigma$  — величина, определяющая меру близости для включения объектов в один кластер;  $d(i_j, i_p)$  — мера близости между объектами, называемая расстоянием.

Неотрицательное значение  $d(i_j, i_p)$  называется расстоянием между элементами  $i_j$  и  $i_p$ , если выполняются следующие условия:

1.  $d(i_j, i_p) \geq 0$ , для всех  $i_j$  и  $i_p$ .
2.  $d(i_j, i_p) = 0$ , тогда и только тогда, когда  $i_j = i_p$ .
3.  $d(i_j, i_p) = d(i_p, i_j)$ .
4.  $d(i_j, i_p) \leq d(i_j, i_r) + d(i_r, i_p)$

Если расстояние  $d(i_j, i_p)$  меньше некоторого значения  $\sigma$ , то говорят, что элементы близки и помещаются в один кластер. В противном случае говорят, что элементы отличны друг от друга и их помещают в разные кластеры.

## 1.2. Набор данных для WEKA

Для построения модели кластеризации мы воспользуемся данными ювелирного предприятия. Ювелирное предприятие имеет 4 ювелирных салона в различных частях г. Красноярска (Центральный, Октябрьский, Свердловский и Советский районы). Были собраны данные обо всех посетителях ювелирных салонов, ювелирных украшениях за 2013 год, которые их заинтересовали, и о том, насколько часто посетители ювелирных салонов в итоге покупали приглянувшееся им украшение. Необходимо проанализировать эти данные для того, чтобы выделить различные группы посетителей и понять, нельзя ли определить какие-либо тенденции в их поведении. В нашем примере используется 1965 записей, и каждый столбец описывает определенный этап, который, как правило, проходит покупатель в процессе выбора и приобретения ювелирного украшения. Соответственно, значение 1 в столбце говорит о том, что посетитель прошел конкретный этап, а 0 — что посетитель этот этап не прошел.

Для анализа данных будем использовать свободное программное обеспечение Weka (Waikato Environment for Knowledge Analysis), написанное на Java в университете Уайкато (Новая Зеландия), распространяющееся по лицензии GNU GPL.

## 2. Результаты исследования

Полное исследование всех типов постановок задач анализа данных ювелирного предприятия в рамках данной работы представляется невозможным, поэтому исследования проводились при следующих ограничениях:

- За основу реализуемой модели взят кластерный анализ;
- Исследуемые данные поделены на 7 кластеров, как оптимальное число кластеров;
- Количество стилей ювелирных украшений - 10;
- Количество типов ювелирных украшений – 20.

После изучения существующих методов кластеризации было установлено, что наиболее подходящим методом для начальных исследований является метод **SimpleKMeans**, позволяющий нам самостоятельно настраивать число кластеров, в то время как другие методы (такие как, DBSCAN, FilteredClusterer, OPTICS и др.) не позволяют этого делать.

Как уже отмечалось выше, для разбиения 1965 строк на 7 кластеров с помощью электронных таблиц потребовалось бы несколько часов, однако WEKA выдает нам результат менее чем за секунду. Полученные результаты требуют дальнейшей интерпретации, исследования и изучения свойств и характеристик объектов для возможности точного описания, сформированных кластеров.

Каждый кластер характеризует определенные покупательские предпочтения клиентов, таким образом, на основании нашего разбиения мы можем сделать **некоторые полезные выводы**:

### Кластер 0

Группа покупателей, в которой при покупке ювелирного изделия предпочитаемой моделью является кольцо. Стиль, на который покупатели, прежде всего, обращают свое внимание - ар деко - стиль ясных геометрических форм, с четким построением симметричных композиций, главенствующую роль в которых играют прекрасно ограненные драгоценные камни. Этот стиль сочетает в себе простоту и роскошь.

Предпочитаемой технологией изготовления в данном кластере является монтаж. При данной технологии, украшения изготавливаются вручную, а также с использованием штампованных, чеканных, сканных или литых элементов.

Здесь мы можем наблюдать достаточно высокий спрос на изделия, но в те же время присутствует интерес к изделиям меньшего веса и меньшей стоимости, чем изделия в данный момент производятся.

### Кластер 1

Покупатели этой группы всегда ищут модели ювелирных изделий 925 пробы и никогда не интересуются более дорогими изделиями. Они сразу же проходят в демонстрационный зал серебряных изделий, не тратя время на осмотр изделий из золота. Кроме того, их не особо интересуют эксклюзивные стилевые направления ювелирных украшений и отдают свое предпочтение стилю модерн - традиционному виду украшений с новыми оригинальными чертами. Наиболее популярная у покупателей модель в данном кластере – кольцо. Свой взгляд останавливают все также, на украшения, изготовленные вручную.

### **Кластер 2**

Эту группу следовало бы назвать «любители недорогих изделий массового производства», потому что посетители, попавшие в это кластер, более чем в 80-90% случаев покупают ювелирное изделие и определенно знают, какие изделия их интересуют.

Обратим внимание, что данные этого кластера демонстрируют интересные вкусовые предпочтения этих покупателей: самой выбираемой моделью являются серьги, а предпочитаемым драгоценным материалом является сплав красного золота 585 пробы. Как правило, они покупают модели в стиле русского узорочья - украшения, которые отличаются пышностью орнаментального декора и необычайным богатством красок.

### **Кластер 3**

Группа покупателей, в которой при покупке ювелирного изделия предпочитаемой моделью является кольцо.

Анализируя данные этого кластера, можно сделать следующий вывод: предпочитаемой технологией изготовления в данном кластере является монтировка.

Посетитель, попавший в этот кластер, гарантированно выбрал определенную модель ювелирного изделия из сплава белого золота 585.

Покупатели отдают свое предпочтение стилю модерн - традиционному виду украшений с новыми оригинальными чертами.

### **Кластер 4**

Эту группу посетителей можно было бы назвать «мечтатели». Они бродят по ювелирному салону, рассматривая ювелирные изделия, выставленные на внешних витринах, но никогда не заходят внутрь, и, хуже того, практически никогда ничего не покупают.

### **Кластер 5**

Предпочитаемой моделью в 5 кластере является кольцо. Посетители ювелирных отделов данного кластера предпочитают изделия, выполненные из серебра 925. Стилевое направление, на котором покупатели останавливают свой выбор - романтизм с его сентиментальным характером украшений, «идеальная ценность» которых в глазах владельцев намного превышает материальную. Все также покупателей привлекают изделия единичного исполнения, изготовленные вручную.

Посетители, попавшие в это кластер, приобрели 82-93% изделий от общего количества произведенных изделий, что является довольно хорошим показателем продаж. Для того, чтобы повысить интерес покупателя рекомендуется снизить вес ювелирных изделий (поскольку покупатели предпочитают изделия немного меньшего веса, чем предложено) и увеличить количество новых форм и вариаций ювелирных украшений.

### **Кластер 6**

Эту группу смело можно назвать «любители роскоши», поскольку в область их вкусовых предпочтений попадают изделия из сплава желтого золота 585 в стиле романтизм. Стоимость предлагаемых изделий начинается от 20 000 руб., что говорит уже о том, что здесь присутствуют потребители продукции высокого ценового сегмента.

Предпочитаемой технологией изготовления в данном кластере является монтировка – одна из самых трудозатратных технологий, но в то же время поражающая своим результатом, качеством и неповторимой красотой изделия.

## **3. Заключение**

В данной работе предложен подход к решению задачи оценки спроса на ювелирные украшения, выполнено построение модели на основе алгоритма кластерного анализа, проведена интерпретация полученных результатов, которая позволила сделать самостоятельные заключения о тенденциях внутри каждой группы, и обоснована возможность практического применения данной методики в ювелирном производстве.

С этой точки зрения, использование кластерной модели позволило извлечь из результатов нашего кластерного анализа достаточно полезную и любопытную информацию, которую мы не смогли бы получить с помощью других методов.