

МЕТОД ГЛАВНЫХ КОМПОНЕНТ И КЛАСТЕРНЫЙ АНАЛИЗ В МАРКЕТИНГЕ

Решетняк Ю.Б.

Национальный фармацевтический университет, г. Харьков

В маркетинговых исследованиях довольно часто приходится сталкиваться с задачами, содержащими значительные объемы исходных данных. Большим может быть, как количество объектов исследования (наблюдений), так и число признаков, характеризующих эти объекты. Такими задачами, в частности, являются сегментирование рынка и многие другие. Для подобных задач важное прикладное значение имеют методы кластерного анализа — набора алгоритмов, позволяющих упорядочивать объекты в сравнительно однородные группы.

Целью настоящей работы является исследование возможности применения метода главных компонент (МГК) для решения задач кластеризации в маркетинговых исследованиях.

Кластерный анализ не накладывает ограничений на вид рассматриваемых объектов, и позволяет рассматривать множество исходных данных практически произвольной природы. Если объекты кластеризации представить как точки в n -мерном пространстве признаков (n — количество признаков, характеризующих объекты), то сходство между объектами определяется через понятие расстояния между точками. Способ вычисления расстояния между объектами или группами объектов отражает специфику решаемой прикладной задачи. Например, в случае непрерывных переменных может быть задано

евклидово расстояние: $r = \sqrt{\sum_{i=1}^n (X_{ij} - X_{ik})^2}$, где X_{ij} — значение i -го фактора в j -м наблюдении.

В опросных листах при проведении маркетинговых исследований, как правило, содержатся родственные вопросы, и полученные данные обычно мультиколлинеарны. МГК позволяет перейти от исходного множества коррелированных факторов X_1, X_2, \dots, X_n к новому набору объясняющих переменных Z_1, Z_2, \dots, Z_n с нулевыми коэффициентами ковариации между ними: $cov(Z_i, Z_j) = 0, (i \neq j)$. Он позволяет уменьшить количество объясняющих переменных до нескольких первых главных компонент и, таким образом, облегчить задачу кластеризации.

МГК легко реализуем в программной среде R. Для этой среды с открытым исходным кодом существуют многочисленные дополнительные библиотеки-модули, которые значительно расширяют ее функциональность. R обладает большим набором опций для визуализации данных.

Метод главных компонент хорошо приспособлен для маркетинговых исследований, позволяет формировать новое пространство объясняющих переменных с нулевой мультиколлинеарностью, снижать размерность этого пространства, облегчает визуализацию данных.