

## ЭВЕНТОЛОГИЧЕСКИЙ МЕТОД ГЛАВНЫХ КОМПОНЕНТ

Веретнова К.Ю.

научный руководитель доктор физико-математических наук, профессор

Воробьев О. Ю.

*Сибирский Федеральный Университет*

### Введение

Метод главных компонент — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Изобретен К. Пирсоном в 1901 г. Это один из способов выделить главный фактор, влияющий на группу объектов.

В настоящей статье приводится сравнение классического и событийного (эвентологического) метода главных компонент. Главное различие между этими методами — это то, что объектами применения классического метода являются конечные объекты, тогда как объектом применения эвентологического метода главных компонент являются события. Несмотря на разный род объектов между методами прослеживается определенная аналогия, которая будет рассмотрена далее.

### Классический метод главных компонент

Классический метод главных компонент предполагает работу с конечной совокупностью определенных объектов. В качестве таких объектов могут выступать объекты природы (люди, животные), экономические объекты (банки, предприятия). Совокупность объектов исследования — множество единиц одного рода, класса, которые обладают сходными свойствами, присущими всему классу.

Для каждой такой совокупности объектов необходимо выделить признаки. Признаки — это показатели, которыми можно охарактеризовать каждый объект из множества объектов изучения.

Например, если мы изучаем группу людей в условиях крайнего Севера[], то такими признаками могут выступать артериальное давление, пульс, количество приседаний, которое способен сделать человек в минуту.

В работе [] объектами изучения выступали банки в условиях экономического кризиса. Для них признаками являлись показатели ежемесячной балансовой отчетности.

После сбора информации данные формируются в матрицу, которая называется матрицей данных.

Матрицей данных  $X$  размерности  $(n \times m)$  будем называть матрицу, элементами которой являются  $x_{ij}$  — значение  $j$ -го показателя для  $i$ -го объекта, где  $n$  — количество исследуемых объектов,  $m$  — количество показателей, выбранных для характеристики совокупности исследуемых объектов,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ .

То есть каждый объект исследования представляется в виде точки в  $n$ -мерном пространстве показателей. Достоинство такого представления — это наглядность. Мы сразу можем представить облако точек-объектов в пространстве.

Задача метода главных компонент — это уменьшение размерности этого пространства  $n$ -мерного показателей, отыскание показателя, который более всего влияет на исследуемую совокупность.

Суть метода главных компонент — это отыскание такого признака, от которого совокупность объектов наиболее зависима.

Технически метод главных компонент заключается в отыскании матрицы векторов главных компонент.

То есть, изначально исследуемые данные представлены в виде матрицы:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}$$

В ходе реализации алгоритма метода находится матрица векторов главных компонент (матрица перехода к пространству главных компонент):

$$A = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ a_{21} & \dots & a_{2k} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{nk} \end{pmatrix},$$

где  $k$  — желаемая размерность пространства.

Каждый вектор  $a_i = \{a_{ij}\}_{j=1}^m$  матрицы  $A$  – вектор главных компонент, в котором каждая -ая координата соответствует  $j$ -му признаку. Тот признак, который имеет наибольшее значение по модулю в первом векторе главной компоненты – и есть главный фактор. В векторе второй главной компоненты признак, имеющий наибольшее по модулю значение – второй главный фактор и т.п.

Метод главных компонент позволяет отсеять существенные признаки для совокупности объектов от тех, которые не вносят различия между объектами. Используя метод главных компонент можно отбросить показатели, которые не влияют на исследуемое множество, уменьшить размерность данных.

#### **Событийный метод главных компонент**

Объектом применения эвентологического метода главных компонент является совокупность событий. Эти объекты отличаются по природе от объектов исследования классического метода главных компонент. Хотя объект исследования классического метода может стать объектом применения событийного метода главных компонент. Например, человека можно рассматривать как событие. Сам факт его существования можно рассматривать как событие.

Пусть есть эвентологическое пространство  $(\Omega, F, P)$ .

Проводя аналогию с классическим методом главных компонент, в эвентологическом методе так же выделяется совокупность исследуемых, уже не объектов, а событий. Пусть  $S \subseteq F$  – множество опасных событий.

Наступление, или не наступление, каждого такого события зависит от наступления, или не наступления других событий, факторных событий. Группа факторных событий – общая для исследуемой группы событий. Это аналог показателей для объектов в классическом методе главных компонент. Обозначим множество факторных событий  $V$ .

Алгебра событий  $A_V$ , порожденная множеством событий  $V$ , содержит всевозможные террасные события

$$ter(X) = \bigcap_{x \in X} x \cap \bigcap_{x \in X^c} x^c, X \subseteq V.$$

В настоящей работе будем называть их факторными террасными событиями.

Каждое факторное террасное событие влияет на наступление событий из  $S$ .

В событийном методе главных компонент теряется наглядность представления исследуемого объекта в пространстве показателей.

Событийный (эвентологический) метод главных компонент – способ выявления главного событийного фактора (факторного террасного события) среди алгебры событий  $A_V$ , порожденной множеством событий  $V$ , влияющих на наступление событий  $s \in S$ .

Задача эвентологического метода главных компонент - это уменьшение размерности множества факторных террасных событий, отыскание фактора, который более всего влияет на наступление, или не наступление, опасных событий из  $S$ .

Суть эвентологического метода главных компонент – это отыскание такого террасного события из  $A_V$ , от которого наиболее всего зависимо наступление, или не наступление опасных событий.

Данный метод позволит снизить размерность рассматриваемых факторных событий, выделить основные факторы влияющие на опасное событие.

Рассмотрим простейший случай, когда  $|S| = 1$ , т.е. рассматривается одно опасное событие  $s \in S$ .

Для того, чтобы выделить среди всех террасных событий наиболее влияющие на событие  $s$ , введем меру зависимости опасного события от факторных террасных событий и ввести алгоритм, по которому будут искаяться главные факторные события.

Пусть  $\delta(s, v)$  – мера зависимости между событиями  $s \in S$  и  $v \in V$  относительно вероятности  $P$ . Мера зависимости  $\delta(s, v)$  обладает следующим свойством: чем больше вероятностная зависимость между факторным и опасным событиями, тем больше величина  $\delta(s, v)$ , тем сильнее положительная вероятностная зависимость. Аналогично, чем меньше вероятностная зависимость, тем меньше величина  $\delta(s, v)$ , тем сильнее отрицательная вероятностная зависимость.

Тогда, если

$$\delta(s, v_1) = \begin{cases} \max_{v \in A_v} \delta(s, v) \\ \min_{v \in A_v} \delta(s, v) \end{cases}$$

то в случае максимума  $v_1 \in A_v$  – главная событийная компонента, вероятностно положительно влияющая на опасное событие  $s$  более других (относительно выбранной меры зависимости  $\delta$ ) факторных террасных событий из алгебры событий  $A_v$ , порожденной множеством факторных событий  $V$ . А в случае минимума  $x_1^c$  – главная событийная компонента, вероятностно положительно влияющая на опасное событие  $s$  более других дополнений событий из  $A_v$  (относительно той же меры зависимости  $\delta$ ). Если таких компонент несколько, каждый раз выбирается любая из них.

Затем по тому же критерию из оставшихся событийных компонент выбираются  $v_2, v_3$  и т.д.

В качестве меры зависимости между событиями можно рассматривать, к примеру,

$$\delta(s, v) = P(s|v) - P(s) \rightarrow \begin{cases} \max_{v \in A_v} \delta(s, v) \\ \min_{v \in A_v} \delta(s, v) \end{cases}$$

Применяя эвентологический метод главных компонент, можно выстроить террасные факторные события в порядке убывания меры вероятностной зависимости. Это позволит увидеть, какие события с наибольшей вероятностью порождают наступление опасного события  $s$ , а какие наоборот – не оказывают никакого влияния.

**Анализ аналогии между классическим и эвентологическим методом главных компонент**

Оба метода исследуют конечную совокупность объектов. В случае классического метода – это конечные элементы вещественного мира, в случае эвентологического метода объектами изучения являются события и их наступление.

Эвентологический метод главных компонент является продолжением классического метода, расширяет класс объектов, которые могут быть изучены с помощью этого метода.

Как и в классическом, в событийном методе главных компонент выделяются признаки, характеризующие исследуемую группу объектов – событийные факторы.

Однако в случае эвентологического метода главных компонент теряется наглядность представления отдельного объекта исследования как точки в пространстве признаков.

Несмотря на разный род объектов, для которых применяются два метода (классический и событийный), суть обоих методов – снижение размерности, выделение главных влияющих на исследуемые объекты факторов.