

PROVISION OF REPRESENTATIVE SELECTION OF OBSERVATIONS IN CONDITIONS OF SHORT RANGE OF DYNAMICS

V.K. Kublikov, X.Yage

Odessa University

Abstract. Representative sample of observations in terms of a number of short ynamiky. The approach to ensure representativeness of the sample observation technical and economic performance of the company in a dynamic changing external and internal conditions of its operation. The proposed approach can improve the homogeneity of aggregate indicators studied, and it presentable in terms of a short series of observations.

Keywords: Short Range Of Dynamics, Representative Selection

ПРЕДОСТАВЛЕНИЕ ПРЕДСТАВИТЕЛЬСТВА ВЫБОРОВ НАБЛЮДЕНИЙ В УСЛОВИЯХ КОРОТКОГО ДИАПАЗОНА ДИНАМИКИ

В.К. Кубликов, канд. экон. наук, доц.,
Ха Халил Ягі, инженер

Абстрактные. Удостоверьтесь в представлении наблюдений образца в виде короткой серии динамик. Подход к обеспечению репрезентативности технических результатов и экономического наблюдения показывает, что компания динамично меняет внешние и внутренние условия своего бизнеса. Можно ли изучить предлагаемую методологию для повышения однородности агрегированных показателей и представить ее в виде краткой серии наблюдений.

Ключевые слова: Короткого Ряда Динамики, Репрезентативности

Введение :

Проведение анализа развития предприятия заключается в обработке имеющейся на данный момент информации о состоянии предприятия и наблюдавшихся ранее закономерностях его изменения, о конкретных условиях его функционирования, и превращении ее в информацию о будущем состоянии.

При исследовании такого сложного объекта как предприятие необходимо использовать стохастический подход, т.к. сложное взаимодействие технико-организационных факторов (ТОФ), влияющих на его развитие, невозможно отразить в пределах одной и той же детерминированной модели. Вместе с тем, использование стохастического моделирования определяется возможностями составления совокупности наблюдений, необходимой размерности и качественной структуры.

Основное требование при стохастическом моделировании — качественная однородность характеристик моделируемого объекта во всех наблюдениях. Но экономические явления столь сложны, что невозможно полное совпадение их количественных характеристик даже на однородных предприятиях [1, 2].

Другим условием использования стохастического моделирования является достаточная размерность выборки. Только при многочисленных выборках ($n > 100$) коэффициент корреляции может достаточно точно воспроизвести действительную ее величину [3]. Количество наблюдений должно в 3...4 раза превышать количество переменных в модели [4]. В экономическом анализе составление столь объемных совокупностей крайне затруднительно.

Предлагаемый подход позволяет повысить однородность изучаемой совокупности и повысить ее представительность. Рассмотрим две ситуации.

В первой ситуации по анализируемому предприятию имеется достаточно представительная выборка наблюдений, однако, вследствие изменения условий функционирования предприятия динамику изменения показателей довольно сложно описать какой-либо определенной функцией.

В силу дискретности измерения значений показателей будем рассматривать совокупность значений показателей в определенный момент времени как точки в k -мерном пространстве. В редких случаях эти точки рассеяны равномерно по всему пространству. Чаще они представляют собой некие обособленные группировки (облака).

При построении модели анализа пытаются аппроксимировать эти точки некоторой поверхностью, в частности, при линейном регрессионном анализе некоторой гиперплоскостью. Провести такую гиперплоскость через все облака затруднительно, т.к. отклонения точек от гиперплоскости будут составлять значительную ошибку модели. Но можно не находить наилучшую единую гиперплоскость, а аппроксимировать каждую отдельную область своей гиперплоскостью. Задача заключается в том, чтобы разделить между собой плотные группировки точек и строить модель для каждой из них. В предлагаемом подходе для решения задачи такого разделения используется аппарат автоматической классификации.

Во второй ситуации доступная выборка наблюдений по одному предприятию не отвечает требованиям к объему выборочной совокупности. В то же время возможно получение дополнительных данных по другим предприятиям. При использовании метода "звездолет" именно так и поступают. Однако, обобщенная совокупность наблюдений по ряду предприятий может не удовлетворять принципу однородности в

силу различных условий развития предприятий, специфичности хозяйственных ситуаций, сложившихся на них, особого взаимодействия ТОФ на каждом из предприятий. В графическом представлении такой ситуации в условном трехмерном пространстве (рис. 1), точки, соответствующие различным предприятиям, выделены различными обозначениями.

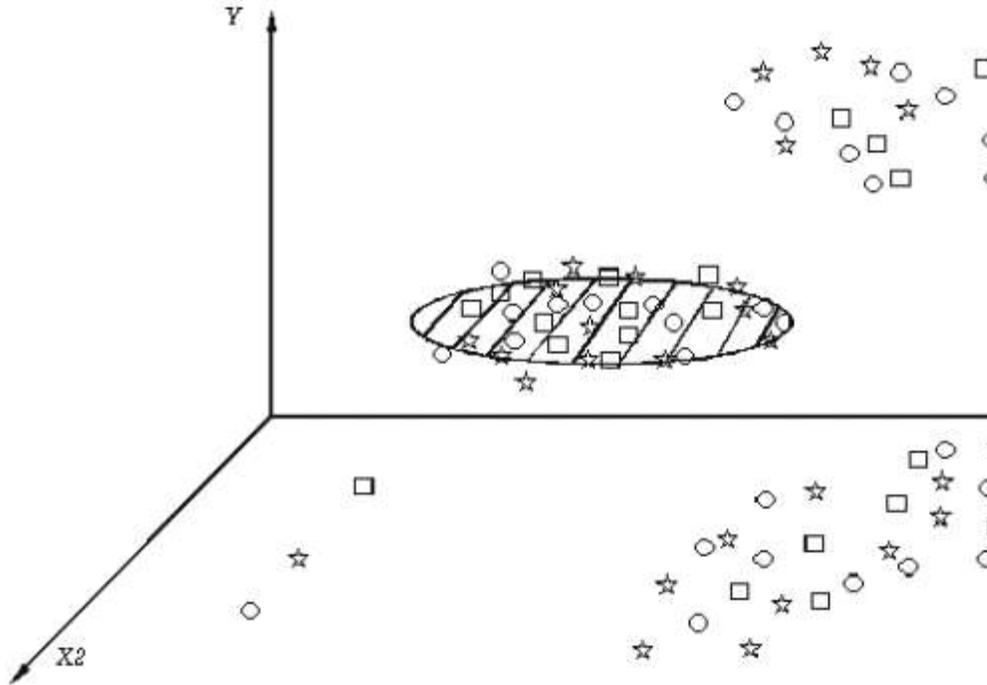


Рис. 1. Графическое представление в условном трехмерном пространстве значений показателей по нескольким предприятиям

При этом может накладываться и первая ситуация, связанная с изменением временных условий функционирования. Поэтому группы точек, соответствующие одному предприятию, одновременно разнесены в пространстве и близко соседствуют с точками, относящимися к другим предприятиям. Близость точек, соответствующих различным предприятиям, свидетельствует о сходстве производственно-хозяйственных ситуаций, сложившихся на предприятиях в периоды, которым соответствуют эти точки. Общая совокупность показателей по всем предприятиям непригодна для построения достаточно адекватной модели. Однако, можно выделить компактную группу, состоящую частично из значений показателей различных предприятий, достаточно однородную для проведения анализа. Решение этой задачи также основывается на методах автоматической классификации.

Существуют различные подходы к решению задачи автоматической классификации. В предлагаемой модели используется вариационный подход, основанный на экстремизации функционала, характеризующего качество полученного разбиения. Наибольший интерес в вариационном подходе представляют рекуррентные алгоритмы, позволяющие обрабатывать данные, поступающие последовательно во времени [6, 7]. Использование рекуррентного алгоритма в данной модели позволит в

дальнейшем отнести точку, соответствующую моменту прогноза в модели анализа факторов развития, к одному из классов без выполнения заново классификационной модели.

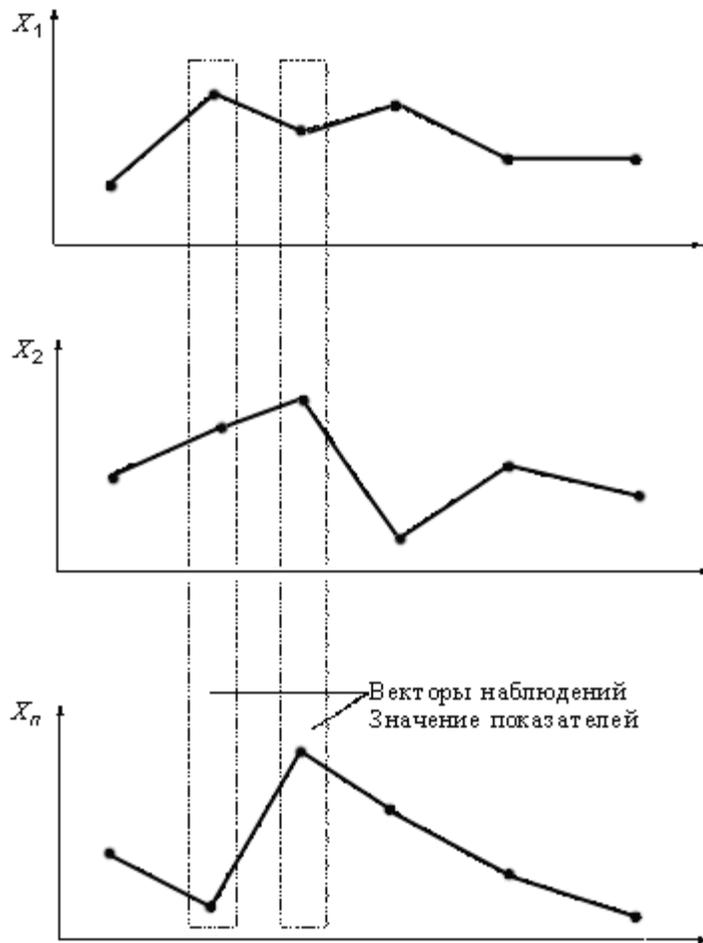


Рис. 2. Схема образования векторов наблюдений значений показателей

Рассмотрим постановку задачи автоматической классификации применительно к анализу совокупности показателей работы предприятия. Пусть имеется k рядов наблюдений, каждый из которых содержит n значений одного из показателей. Из одномоментных элементов этих рядов наблюдений образуем n векторов значений показателей размерностью k (рис. 2). Каждый вектор представляет собой как бы "моментный срез", совокупную характеристику состояния системы в определенный момент времени. При включении в рассмотрение данных по группе предприятий такие векторы составляются по значениям одноименных показателей каждого предприятия группы. Полученные векторы становятся объектами модели классификации.

Введем в рассмотрение k -мерное пространство X , в котором каждому вектору (объекту) соответствует точка $x \in X$.

Пусть заданная выборка объектов поступает последовательно во времени x_1, \dots, x_k, \dots в соответствии с заранее неизвестным законом распределения $P(x)$. Необходимо по этой выборке разделить пространство X на заданное число областей (классов) так, чтобы на этом разбиении выбранный критерий качества принимал максимальное значение.

Наиболее исследованным критерием качества является функционал [8, 9, 10]

$$\Phi = \sum_{i=1}^r M_i r / P_i,$$

где $M_i = \int x^i p(x) dx$ — первый ненормированный момент i -го класса;

$P_i = \int h_i(x) p(x) dx$ — вероятность i -го класса;
 $h_i(x)$ — характеристическая функция i -го класса.

Данный функционал эквивалентен функционалу средневзвешенной дисперсии, поэтому характеризует разбросанность объектов внутри классов. Этот вид функционала используется в предлагаемой модели, т.к. наилучшим образом зарекомендовал себя в практических приложениях и имеет наглядную геометрическую интерпретацию [11].

При разработке модели автоматической классификации важным вопросом является выбор класса разделяющих функций. Практика свидетельствует, что разбиения, задаваемые линейными разделяющими функциями, являются в определенном смысле наилучшими. А именно, если стандартное значение функционала достигается на некотором разбиении, то это же значение достигается и на разбиении, которое можно задать с помощью линейных разделяющих функций [12]. Это означает, что алгоритм максимизации функционала можно конструировать на базе линейных разделяющих функций. Каждая линейная классификация задается с помощью $r(k+1)$ -мерного разделяющего вектора $p = (d_1, c_1, \dots, d_r, c_r)$, где d_i — константы, c_i — k -мерные векторы. Вектор задает r линейных разделяющих функций $f_i(x) = (c_i, x) + d_i$, $i = 1, \dots, r$.

Линейную классификацию, определяемую этими разделяющими функциями, обозначим через H_p .

Соответствующие функции принадлежности обозначим через $h_i p(x)$ и определим следующим образом:

$$h_i p(x) = \begin{cases} 1, & \text{если } f_i(x) = \max_{j=1, \dots, r} f_j(x) \text{ и } = \min_{j=1, \dots, r} \{j : f_j(x) = f_i(x)\} \\ 0, & \text{в противном случае} \end{cases},$$

где $f_i(x) = (c_i, x) + d_i$, $p = (d_1, c_1, \dots, d_r, c_r)$.

Разделяющая функция $f_i(x)$ выбирается так, что если для данного объекта ее значение оказывается наибольшим, то этот объект относится к i -му классу. Таким образом, решить задачу автоматической классификации — построить по имеющейся матрице данных систему функций $f_i(x)$, своим знаком указывающих принадлежность любого, а не только имеющегося в матрице данных, объекта к тому или иному классу. Определяемые такой функцией поверхности называют распределяющими поверхностями. При этом "хорошая" разделяющая поверхность должна разделять "горбы" функции плотности вероятности, проходя по "ложбинам" такой функции

[13, 14].

Учитывая изложенное, в качестве алгоритма классификации для модели отбора совокупности однородных наблюдений решено использовать рекуррентный алгоритм автоматической классификации, основанный на максимизации, в данном случае функционала, характеризующего плотность объектов внутри класса и использующего линейные разделяющие функции для задания разделяющих поверхностей. Данный алгоритм можно задать следующим образом [8].

Пусть имеется последовательность $S = \{x_1, \dots, x_n, \dots\}$. На каждом шаге алгоритма строится разбиение пространства X на r классов по выборке $f_i(x)$.

Алгоритм определяется индуктивно. Пусть построено разбиение выборки S_{n-1} на r классов. По этому разбиению строятся оценки моментов $Y_{n-1} = (v_{1n}, s_{1n}, \dots, v_{rn}, s_{rn})$,

где v_{in}, s_{in} — оценки вероятностей P_i и первых ненормированных моментов M_i соответственно. Затем вычисляется градиент r_{n-1} функционала Φ в точке Y_{n-1} . Используя r_{n-1} как разделяющий вектор, строим классификацию $H_{r_{n-1}}$. Далее для простоты будем обозначать $H_{r_{n-1}} = H_{n-1}$. На n -м шаге вновь поступившая точка i по классификации H_{n-1} относится к одному из r классов, другими словами, по разбиению выборки S_{n-1} и X_n строится разбиение выборки S_n .

Формально алгоритм максимизации функционала Φ можно записать в виде

$$v_{in} = \frac{1}{1 - \frac{1}{n} r v_{in} + \frac{1}{n} r h_{in-1}(x_n)},$$

$$s_{in} = \frac{1}{1 - \frac{1}{n} r s_{in} + \frac{1}{n} r h_{in-1}(x_n)},$$

$$Y_n = (v_{in}, s_{in}, i = 1, \dots, r),$$

$$r_n = \frac{\Phi}{(Y_n)} = (d_{in}, c_{in}, i = 1, \dots, r),$$

$$H_n = h_{in}(x), (i = 1, \dots, r).$$

Начальные условия алгоритма определяются соотношениями

$$v_{ir} = \frac{1}{r},$$

$$s_{ir} = x_i / r,$$

$$x_i \neq x_j, j < i, i = 1, \dots, r.$$

Данный алгоритм доведен до программной реализации. Моделирующая программа осуществляет автоматический перебор возможного количества классов, на которые производится разбиение общей плоскости, из диапазона значений определяемых пользователем. Работа программы начинается с занесения первых r объектов по одному в каждый класс. По этой начальной классификации определяются параметры модели, т.е. значения функционала разделяющих функций и функций принадлежности. Далее из совокупности выбирается следующий объект. Он относится по параметрам предыдущего шага к одному из классов, и затем параметры перерасчитываются с учетом информации вновь поступившего объекта. Итерации продолжаются до последнего момента, когда в рассмотрение вводится последний объект.

Результатом программы является набор классов с отнесенными к ним соответствующими моментными значениями показателей по одному предприятию (первая ситуация) либо по группе предприятий (вторая ситуация). Полученные данные могут использоваться в дальнейшем в моделях выделения приоритетных направлений развития и прогноза показателей.

В моделях прогнозирования определение класса, на базе которого будет строиться прогнозная модель, осуществляется по значениям показателей в момент, предшествующий точке прогноза. Такое отнесение объекта к классу возможно в силу рекуррентности алгоритма, использующего сведения только об очередном объекте, а не о величине "связи" этого объекта со всеми другими объектами.

Таким образом, использование предложенной модели позволяет повысить точность применяемых аналитических и прогностических моделей за счет повышения однородности исследуемой выборочной совокупности значений показателей.

Литература

- [1] Economic and mathematical methods in the analysis of economic activity of enterprises and associations / Butnik-Siversky AB, Sheremet AD, Reelyan Ya.R. And others - Moscow: Economics, 1983.
- [2] Eddous M., Stansfield R. Methods of decision making. - Moscow: Audit Unity, 1997.
- [3] Ezekiel M., Fox K. Methods of regression correlation analysis. - Moscow: Statistics, 1966.
- [4] EM Braverman, IB Muchnik. Structural methods for processing empirical data. - Moscow: Nauka, 1983.
- [5] Aizerman MA, Braverman EM, Rosonoer LI The method of potential functions in the theory of machine learning. - Moscow: Nauka, 1970.
- [6] Applied Statistics: Classification and Dimension Reduction / Aivazyan SA, Buchstaber VM, Eukov IS, Meshalkin LD - Moscow: Finance and Statistics, 1989.
- [7] EV Bauman and Dorofeyuk AA Recurrent algorithms for automatic classification // Automation and telemechanics. - 1981. - No. 3. - P. 95 - 105.
- [8] Tsyarkin Ya.E. Adaptation and training in automatic systems. - Moscow: Nauka, 1968.
- [9] Training systems for information processing and decision making / Lapko AV, Chentsov SV, Krahov SI, Feldman LA - Novosibirsk: Science, 1996.
- [10] Vapnik VN, Chervonenkis AF Theory of pattern recognition. - Moscow: Nauka, 1984.
- [11] Levin M.Yu., Fomin N.V. Proof of the convergence of recurrent learning procedures without a teacher, Vestn. Leningr. University, 1975. - No. 7. - P. 35 - 42.
- [12] Rozin B.B. Theory of pattern recognition in economic research. - Moscow: Statistics, 1973.
- [13] Zamkov OO, Tolstopyatenko AV, Cheremnykh Yu.N. Mathematical Methods in Economics. - Moscow: DIS, 1997.