ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2009, Vol. 43, No. 6, pp. 351–354. © Allerton Press, Inc., 2009.
Original Russian Text © V.M. Grinyak, S.M. Semenov, E.I. Kogay, 2009, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, 2009, No. 12, pp. 6–9.

Model for Planning Seasonal Sales in Corporate Information Systems

V. M. Grinyak, S. M. Semenov, and E. I. Kogay

Received September 10, 2009

Abstract—This paper addresses automated enterprise control, namely, the use of modern information technologies to plan sales. A new model intended for seasonal products and linked with the ideas of cluster analysis is proposed. Its construction and efficiency are demonstrated using real data.

Key words: nomenclature of items, periodicity of sales, seasonality parameters, reliability of planning.

DOI: 10.3103/S0005105509060065

INTRODUCTION

The activities of present-day manufacturing and commercial enterprises are characterized by an extremely high level of various costs, which leads to permanent balancing on the edge of profitability. The price of management errors under such conditions becomes very high; this is a basic incentive for the introduction of information equipment, i.e., corporate information systems (CISs) at enterprises. The basic objective of such corporate systems is the creation of an information base for management decisions; the information content and form of its presentation should be made to secure the fewest erroneous decisions possible.

The nomenclature of the materials and products of a large company can have dozens or hundreds of thousands of items. The known technologies supporting the enterprise activity under these conditions (ERP and MRP) are oriented first of all to minimizing stocks in storage, which allows a reduction of costs and the attraction of working capital. On the other hand, in a competitive environment, it is not less important to keep customers, which implies the guaranteed on-time delivery of the demanded product and constant expansion of the list of stock.

Under these conditions, the planning of sales becomes a crucial function of the corporate information system of an enterprise. Today's corporate systems (for example, Russian Galaktika, IC, international SAP, Axapta, etc.) use various technologies of automation of sales planning, from the simplest types to fairly advanced ones. All of them are actually based on retrospective analysis of data and their extrapolation to forthcoming periods by use of the methods of regression and statistics.

Planning the sales of seasonal products is a very important and difficult problem, as it implies the identification of items whose the sales are seasonal, evaluation of the periodicity of sales, and scheduling the distribution of sales by periods. These parameters should provide for the maximum accuracy of planning, but all current CISs (in their standard configuration) leave it to the user to select these parameters in an exclusively intuitive, "manual," fashion; for a larger list of items the proper evaluation of seasonality parameters may turn out to be impossible.

In this work we examine a model for the problem of planning seasonal sales designed for use in today's CISs and propose a method for the automated evaluation of seasonality parameters based on the technology of cluster analysis.

REPRESENTING THE BASIC MODEL

In problems of economic analysis the minimum period for the evaluation of a certain parameter is as a rule, a day, as related to specific data. In problems of planning use is largely made of "condensed" periods, i.e., weeks, decades, months, quarters, etc.

Let X_k be the value of the selected indicator (it can be, for example, profit from sales, number of sales, proceeds from sales, etc.) in a period with number k. The model of the change in the value of indicator X over time can be expressed by formula

$$X_k = G(k) + \eta(k), \tag{1}$$

where $G(k)$ is a function expressing a determined law for the evolution in the value of X (trend) and $\eta(k)$ is a random value characterizing the digression of the actual value of the indicator from its trend (here and hereafter we will accept that $\eta(k)$ is an uncorrected random value with a zero mathematical expectation). In the solution of the planning problem, the known function $G(k)$ is used to extrapolate the values of indicator X, while properties $\eta(k)$ are based on evaluating the validity of planning (for example, expressed in the form of confidence intervals) [1].

Assume the trend $G(k)$ is a function presented by:

$$G(k) = Ag(k), \qquad (2)$$

where $g(k)$ is a periodic function with period J, so that $g(k) = g(k+J)$ and $\sum_{k=1}^{J} g(k) = 1$, A is a value characterizing the summary magnitude of the indicator X during the period. Function $g(k)$, therefore, assigns a schedule of distribution of indicator X during the period. (Comment: the authors hope that the use of the term "period" for both the number of time interval k in formula (1) and of the property of the periodic function $g(k)$ in formula (2) will present no difficulty for the reader).

From practice it is well known that the period appropriate for the sales of seasonal products is 1 year, which unambiguously identifies the value J corresponding to this time interval. In this context, the main problem of planning sales is, therefore, identification of those items for which $g(k)$ is in fact a periodic function and determination of the function $g(k)$ variety within the yearly period.

Let $g_i(k)$ be the function $g(k)$ for an item with number i. We insert value d_{ij} characterizing the metrics of distance between schedules of distribution of indicators X during the year for items with numbers i and j, so that

$$d_{ij} = \sum_{k=1}^{J} |g_i(k) - g_j(k)|. \qquad (3)$$

The matrix D corresponding to d_{ij} will be called the matrix of distances between yearly schedules of distribution.

Among the number of items there are, as a rule, those that are characterized by a similar "type of seasonality." Thus, the sales of some products are made mostly in the winter period, while the sales of others are in the summer, and of a third group, in the off-season. In products with the same type of seasonality functions $g(k)$ will be similar. In terms of model presentations (3), such products are characterized by points with coordinates $g(k)$ closely located in J-dimensional space. Matrix D is the input data to determine those fields (clusters) of the J-dimensional space in which the largest number of closely located points $g(k)$ are grouped. The items appropriate to such clusters can be identified as products with the same type of seasonality and, therefore, that are characterized by the same schedule of distribution $g(k)$ for solution of the problem of planning.

Thus, the representations described in (1), (2), and (3), the problem of locating the centers of clusters on the set of points $g_i(k)$ of the assigned metrics (3) for the identification of the points belonging to a particular cluster, as well as of the extrapolation of trends $G(k)$ according to (2), can described within the model.

THE METHOD OF PROBLEM SOLVING

There are several approaches to the solution of the problem of clusterization using the matrix D. The first group of methods is associated with clusterization for an assigned number of clusters (concise and nonconcise algorithms of C-means [2, 3], neural Kohonen networks [4], etc.). The second group does not require assignment of the number of clusters; the clusters are identified in the process of the algorithm's (mountain algorithm [5, 6]) operation.

In essence, clusterization for a given number of clusters in the considered problem means the assignment of the number of possible schedules of distribution $g(k)$. This approach is applied if the stock of products is characterized by clearly identified groups and each group has a type of seasonality of its own. In practice this situation is not frequently found. Thus, here we use an approach where the number of clusters is determined by the parameters of the algorithm itself.

The algorithm of mountain clusterization is that at first the points that can be centers of clusters are determined. Then, for each point the value of the potential for cluster formation in its vicinity is calculated. The more densely objects are located in the vicinity of the potential center of the cluster, the higher the magnitude of its potential. Searching for the centers of clusters among the points with maximum potential is done using selection by iteration.

In this work, the objects of clusterization themselves are proposed as centers of possible clusters, that is, the evaluation of the schedules of distribution is $g_i(k)$ for each item, so that

$$\hat{g}_i(k) = \frac{X_k^{(i)}}{\sum_{i=1}^{J} X_j^{(i)}}, \quad k = \overline{1, J},$$

where $X_k^{(i)}$ is the value of the selected indicator in the period with number k for the item with number i.

To calculate the potential of point $\hat{g}_i(k)$, the following function of the potential is selected:

$$P(\hat{g}_i(k)) = \sum_{j=1}^{M} \exp(-\alpha \hat{d}_{ij}),$$

where $\hat{d}_{ij} = \sum_{k=1}^{J} |\hat{g}_i(k) - \hat{g}_j(k)|$ is the evaluation of the corresponding coefficient of the matrix of distances, M is the number of items participating in the cluster analysis, α is a number characterizing the scale of distances d_{ij}; and $\exp(\cdot)$ is an exponential operator.

At the first step of clusterization, the potential of each point is calculated and the point $\hat{g}_i^{(1)}(k)$ with the highest potential $P(\hat{g}_i^{(1)}(k))$; is selected; exactly this point will be center of the first cluster. At the second step, the magnitudes of the potentials of all points are recalculated to exclude the effect of the potential of the cluster already found; to do this, the contribution of the center of the found cluster is subtracted from the current magnitudes of the potential using the formula

$$P_2(\hat{g}_i(k)) = P(\hat{g}_i(k)) - P(\hat{g}_i^{(1)}(k))\exp(-\beta\hat{d}_{iz}),$$

where β is a number characterizing the size of clusters and z is the index of the point that is the center of the first cluster. The center of the second cluster is point $\hat{g}_i^{(2)}(k)$ with the maximum value of potential $P_2(\hat{g}_i^{(2)}(k))$. The centers of all subsequent clusters are found in a similar fashion. The iteration procedure of the recalculation of potentials and identification of the centers of clusters continues until the maximal potential magnitude exceeds a certain assigned threshold.

Whether point $\hat{g}_i(k)$ is a part of a particular cluster is determined by the distance to the center of this cluster. A point is considered to belong to a cluster if the distance to it (compared to the distance to the centers of other clusters) is minimal and does not exceed some assigned magnitude β^*. But if the distance from the point to the centers of all found clusters is more than β^*, they are considered not to belong to any cluster.

SOLUTION RESULTS

Experimental planning of sales using the proposed technique was made using the actual data on the sales of spare motor parts by a large company; their number includes over 13000 items and regular sales of 2500 items occur. Considering the specificity of the products and the accounting policies of the enterprise, the following values of the problem parameters were used: the indicator X_k equals the quantity of products sold; the time interval between X_k and X_{k+1} is 1 month; period J is 12 months, $k = 1$ appropriate to January and $k = 12$ to December. Parameter a was selected as being equal to two (the maximum possible value d_{ij}), parameter β was assumed to be eight; and the maximum distance to the center of cluster β^* is 0.2.

In Fig. 1a the values of the potentials $P(\hat{g}_i k))$ at the first step of the algorithm (solid line), second and third steps (dotted line) and of three further steps (dots) are shown. It is seen that beginning with the fourth cluster the effect on the potentials of points $\hat{g}(k)$ becomes very small (i.e., a very small difference between the diagrams of the points in the figure). This means that at the periphery of the centers of the fourth and subsequent clusters the points are, in fact, not grouped. This is also validated by the analysis whose results are presented in Fig. 1a. It shows the number of items belonging to the first cluster (152 items), the second (31 items), third (45 items), and the fourth and fifth cluster (4 and 3 items, respectively). From the results it follows that from the considered set of points $\hat{g}(k)$ it is possible, in fact, to identify only three clusters with the described numbers of points.

Fig. 1b shows the values of $\hat{g}(k)$ points belonging to the first cluster (the beam of graphs of points in the left part of the figure) and values $\hat{g}(k)$ of the point that is at the center of the first cluster (heavy solid line, left part of the figure). In essence, this is the yearly dynamics of the sales of products belonging to the first type of seasonality, which is related to the standard dynamics of the first type of seasonality (thus, the yearly dynamics of the point that is the center of the cluster are used). The right part of Fig. 1b shows the actual values $\hat{g}(k)$ relating to the standard dynamics belonging to the beginning of the following year. It is seen that the left and right parts of the figure are quite consistent with each other, i.e., the products identified by the results of the analysis for the year as being seasonal products of the first type continue the same type of seasonality in the following year.

A similar picture also occurs for the second, and third clusters, as seen in Figs. 1d and 1e. From the comparative analysis of Figs. 1b, 1d, and 1e it is seen that in the considered array of data on the number of sales three types of seasonality can be deduced, i.e., low level of sales in January–March with a jump in April and a stable level until the end of the year (the first type of seasonality); a stable level of sales from January to July with growth in the autumn and a downturn by the end of the year (the second type of seasonality); and a the low level in the winter period and a high one from May to October (the third type of seasonality). Such types of seasonality are indeed typical of some groups of motor products.

Fig. 1e shows the quantitative distribution by clusters of those products whose sales are regular. The products with the seasonality of the first type account for 22% of all sales, while products with seasonality of the second and third types account for 4% and 11%, respectively. The remainder of the products, viz., 63%, are not seasonal. This relationship (over a third of the products are characterized by pronounced seasonality) confirms the relevance of the application of the respective procedures of planning.

The proposed model planning seasonal sales is adapted to the data of the corporate information system "1C: Commerce Control 8" and "1C: Factory Management 8" and implemented by the processing of platform "1C: 8," whose objective is to supply a manager with a sample of seasonal items and guides on planning their purchases and sales. Testing the program on an operating enterprise demonstrated the construction and self-consistency of the proposed technique.

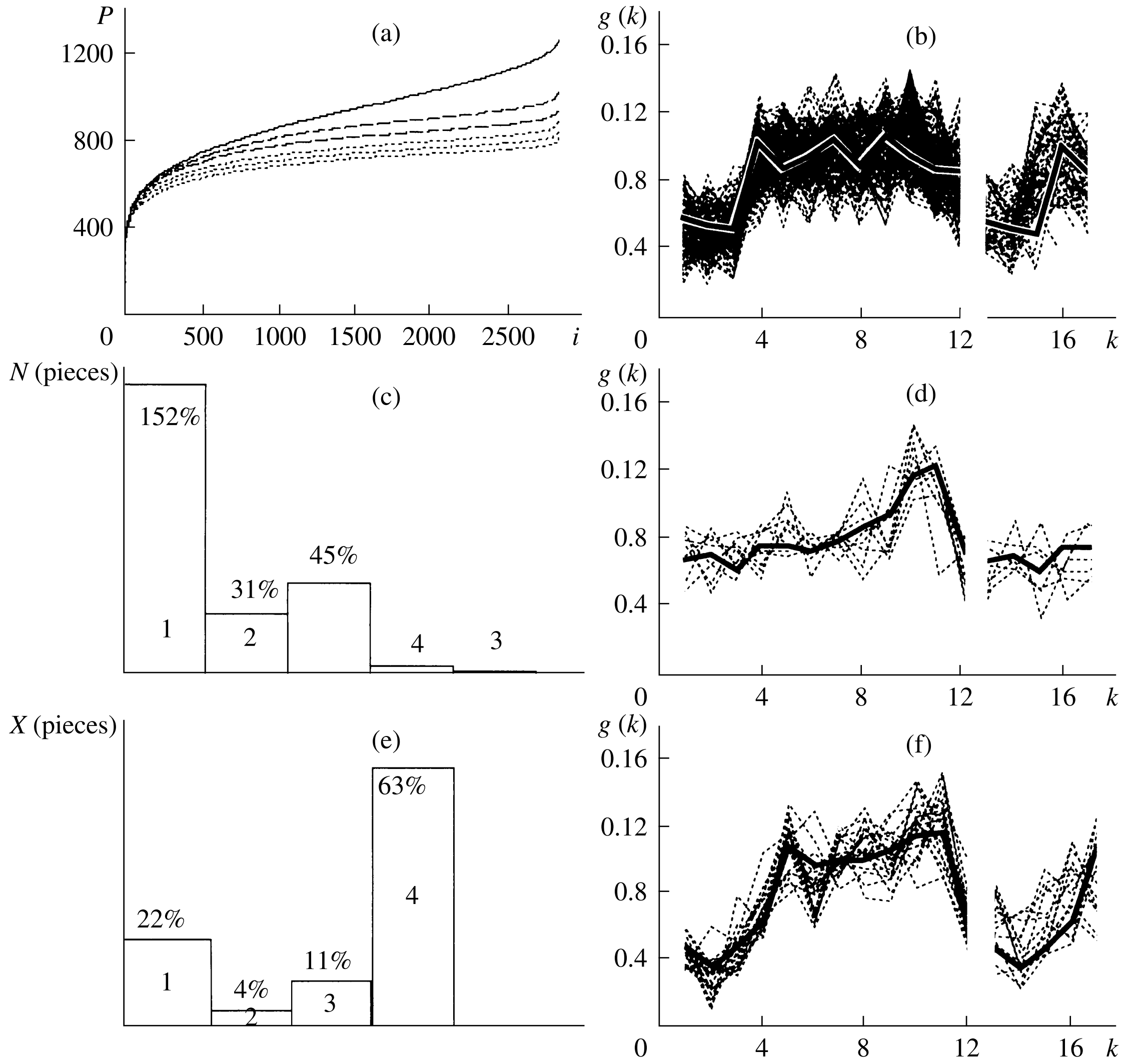

Fig. 1. Solutions of the problem of planning seasonal sales.

This work is aimed at expanding the functionality of corporate information systems.

REFERENCES

1. Grinyak, V.M. and Semenov, S.M., Model of Planning Sales in Corporate Information Systems, *Yestestvennye i tekhnichaskiye nauki*, 2009, no. 1.
2. Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function, New York: Plenum Press, 1981.
3. Babuska, R., Fuzzy Modeling for Control, Boston: Kluwer Academic Publishers, 1998.
4. Khaykin, S., *Neyronnye seti. Polnyi kurs* (Neuron Networks. Full Course), Moscow: Williams, 2006.
5. Yager, R., Filev, D., Essentials of fuzzy Modeling and Control, USA: John Wiley & Sons, 1984.
6. Shtovba, S.D., Introduction to the Theory of Fuzzy Sets and Fuzzy Logic: http://matlab.exponenta.ru/fuzzylogic/book1/index/php