Kalman Filters to Generate Customer Behavior Alarms

Josep Lluis de la Rosa[1], Ricardo Mollet[2], Miquel Montaner[3], Daniel Ruiz[4], Víctor Muñoz[5]

Abstract Aggressive marketing campaigns to attract new customers only covers customer churn, resulting in neither growth nor profitability. Retaining current customers, increasing their lifetime value, and reducing customer churn rates, thereby allowing greater efforts and resources to be dedicated to capturing new customers are the goals of a commercial director. But how can that loss be detected in time and avoided—or at least reduced? There is the 3A program to keep customers loyal, based on analyzed information from our customers, to construct an expert alarm agent and one-to-one retention actions. In this paper we show how to apply the Kalman filter and study how to configure it to predict the normal behavior of customers by projecting their consumption patterns into the future. Abnormal behavior detected by the Kalman filter triggers alarms that lead to commercial actions to avoid customer churn.

Keywords: relational marketing, customer retention and loyalty, one-to-one actions, Kalman filter, intelligent agents

1. Motivation: Why Retain Customers?

1.1 The 3A Loyalty Program

It is a marketing service that consists of the purchasing behavior of customer (by means of an engine called 2A) to retain the most valuable customers. The fact is that if a company has a relatively small number of customers, commercial technicians or agents can closely monitor each one of them. However, when the company has a large number—thousands or even hundreds of thousands—of customers, the sales network will be unable to fully and dynamically control the behavior of many of them, focusing efforts on capturing new customers and monitoring current, generally heavy use customers. Even though that information exists and can be obtained through well-organized database enquiries or found in company reports, it is usually unfeasible for a sales network to analyze the data for each client. In addition to the effort made by the

[135] Josep Lluis de la Rosa, Miquel Montaner, Victor Muñoz : EASY Center of the Xarxa IT CIDEM, University of Girona. Campus Montilivi (Edif. PIV) 17071. Girona, Catalonia. Tel. + 34 972418854, E-mail: {peplluis, mmontane, vmunozs}@eia.udg.es
[24] Ricardo Mollet, Daniel Ruiz, PSM-Carlson Marketing Group, Barcelona-Madrid, Spain, {ricardo, daniel_ruiz}@psm.es

sales network, a large part of company resources are aimed at capturing new customers: brand advertising, product advertising, promotions, discounts, direct marketing, etc. One example is the use of bold campaigns by cell phone companies in Spain in recent years to capture new customers. Unfortunately, in many of these situations limited resources and other strategic focuses mean that customer loss is only detected when it is already too late and recapturing them would be too difficult or too expensive. Retaining customers is profitable: published studies have indicated this; despite the difficulties studying the effects of an increase in customer loyalty on a company (it must be done using comparable test areas, under *the same* conditions, with reliable information about the competition and general market trends, etc.).

1.2 How do Customer Acquisition and Loyalty Programs Work?

First of all, customers are reached by exposing them to “the product or service” (advertising, sales networks, recommendations from other customers) and obtain “product or service proof”. However, to attain loyalty from customers—and to retain them—we must be able to strengthen the *head share* to ensure their purchasing actions are repeated (the purchases repetition is the differential factor to know whether the customer values the product) [7]. Product and/or service quality, adequate prices and customer service are obviously essential, but not sufficient, conditions for their loyalty. Customers “become accustomed” to those aspects and are no longer surprised: they “forget” the company. Satisfying customer expectations is an unending task: establishing mechanisms of continued communication with customers, using special promotions, reinforcing brand image and the image of products and services, etc. are also necessary to maintain and increase the *head share* of the product or service in the minds of customers as well as the emotional links with them—special treatment, privileges and deals only for customers.

Loyalty programs are meant to achieve this objective: in each transaction customers accumulate points they can exchange for products, gifts, etc. or, after a certain volume of transactions, they can take advantage of discounts they would not have had otherwise. For example, airline company loyalty cards let us accumulate points for each flight and then use them to obtain free flights, and let the company gather knowledge about the customer behavior. Likewise, accumulating transactions (flights) can lead to special treatment, privileges and exclusive deals that create particular bonds among customers and companies. In addition, reminding customers of the points they have obtained establishes periodic contact with them in a natural way. And, it should not be forgotten that a record of customer transactions means that an otherwise anonymous service becomes a valuable source of information.

2. What does the 3A Customer Retention Consist Of?

3A is supported by an integrated management system consisting of various marketing, statistical, computer, communication and promotional techniques with the same objectives: generating trust, retaining customers and increasing their *life time value*.

2.1 General Description of the 3A Program

Briefly, the 3A program can be implemented in three separate phases:

1. **Datamining the database** through *statistical* techniques, to segment customers and detect their behavior [4] based on a small group of basic values (for example, in a transportation company, total shipments made, kilos sent, revenue per kilo sent). Subsequently, the behavior analysis of each customer can be made more sophisticated using multiple variables (for example, use of each of the products or services; time/urgency, documents, geographic areas...).

2. Establishing an **expert alarm agent** to personalize actions for individual customers and detect both positive and negative changes in their behavior. In other words, an agile and adaptable alarm system to ensure we act in time. Here, the agent needs the application of a filter, the Kalman filter, as we will show in the following sections.

3. Developing a different **action plan** according to customer type, behavior and profitability. The actions will have been pre-established according to anticipated changes in behavior, although the results of the alarm system will undoubtedly make us reconsider some of the actions or indicate new ones. The actions will be *one-to-one* intended only for those customers who have triggered a specific type of alarm.

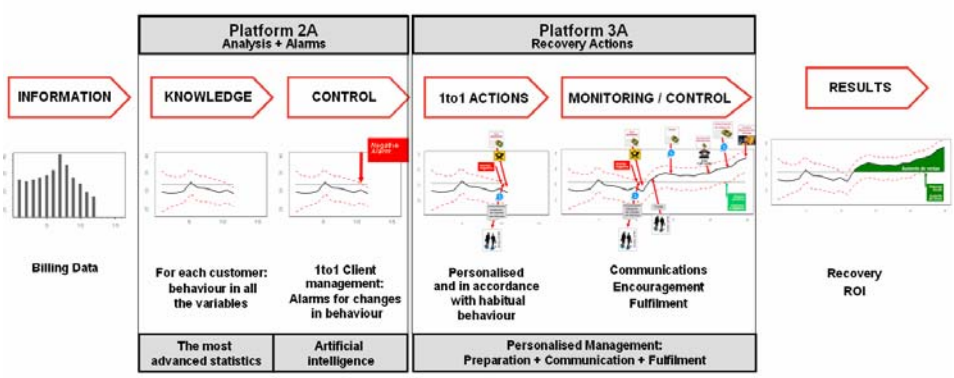

Fig. 1. The 3A program of loyalty actions of the PSM Company

2.2 Expert Alarms Agent and the Need of Kalman Filters

The objective, as we have explained in the description of the 3A program, is to develop an agile alarm system in an agent platform [3] which detects changes in customer activity. Based on what was explained in the previous section, it might seem natural for a good alarm system to detect changes in customer categories. However, this approach (creating alarms by clustering) has three serious disadvantages:

- Customers who are about to change categories can trigger an alarm even though the change in activity is very small. However, customers at the upper end of the interval can substantially reduce their activity before a category change occurs.
- Inclusion in a category of activity is usually based on interannual data (12^{th} order moving average, if we use monthly accumulations) so that the category change can occur with unacceptable delay relative the change in activity.

- We do not take into account the 'history' of customers in terms of their variability and seasonality of their behavior. We should not consider decreases in consumption—for example, over two consecutive months—of customers with stable trajectories in terms of the use of a product or service to be the same as decreases in consumption of customers with much more erratic behavior, or seasonal behavior (ice creams in summer, Christmas sweets in winter, etc).

Here follows the need to filter the individual behavior of every customer instead of by clustering them. However, the *agent* must come across yet another problem: the lack of precision in alarms caused by data noise or by **the lack of exact activity bounds algorithms in the state of the art** [1]. In that respect, approaches used in many companies, such as percentage change (increases or decreases) based on annual, semiannual or quarterly accumulations, are not very reliable in terms of alarm sensitivity and specificity. And using Markov chains to predict the risk of collective customer loss is interesting as a *macro* analysis, but it has obvious disadvantages when we need a system to act in a personalized way. What therefore is the alternative? Our proposal is related to dynamic prediction models, especially the state space model, namely the Kalman filter. We have used Bayesian prediction techniques and dynamic models in monitoring and prediction processes in short-term time series in industrial, economic and commercial spheres of activity.

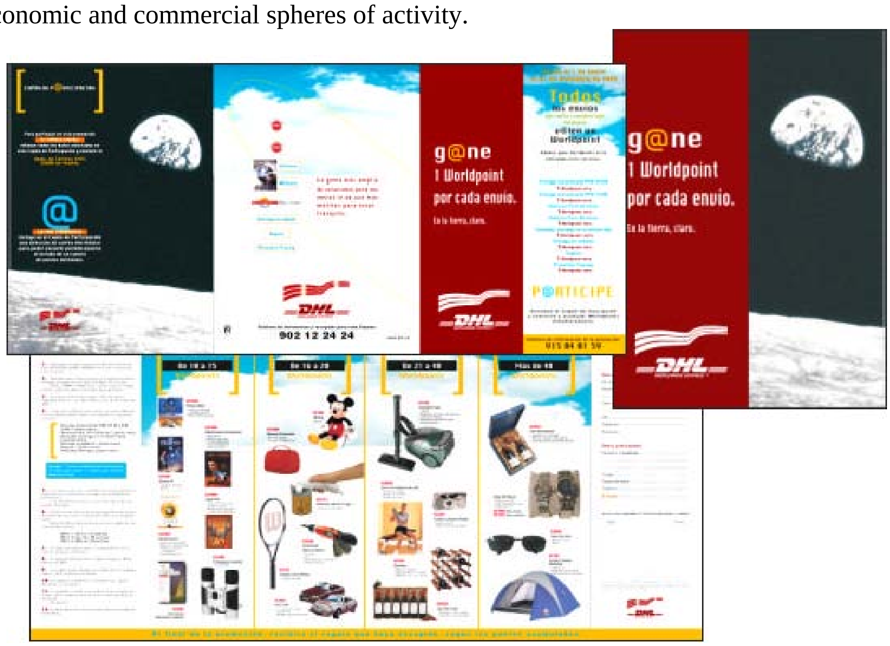

Fig. 2 Examples of generic actions based on models of customer activity with the 3A retention program

3. Applying the Kalman Filter to Predict Customer Purchasing Patterns

Kalman filter provides an estimation of model parameters, detailed predictions and the uncertainty measures associated with these predictions. The great flexibility of the Kalman filter model formulation makes it possible to deal with situations in which, during the process, structural changes may need to be monitored. Likewise, it is possible to introduce control components (interventions) and predict in advance or evaluate afterwards their impact on the system. The incorporation of sequential information allows adaptive estimation and automatic filtering based on Bayesian methods to update and extend the knowledge of the system. The linear dynamic systems formulated in state space constitute an extension of inference and prediction techniques associated with classic methods (exponential smoothing, EWMA [8] [9]). Tools developed for these models, like the Kalman filter equations [5] [6], make it possible to predict, filter, and smooth the trajectory corresponding to the state of the system, based on observations made of it. If the distribution of the activity indicators approaches normal distribution, the model used corresponds to a DLM (dynamic linear model). If customer behavior is stable, the values of the indicators will be constant (constant prediction function), while erratic behaviors will vary the indicators.

The activity observed in a customer during a certain period can be modeled as a 'real' level of underlying activity plus a perturbation originating in 'external factors'. The aim of the monitoring process is to estimate and predict that level of underlying activity to detect structural changes (tendencies, and significant, transitory or level changes).

The objective of the Kalman filter [5] is to estimate and predict customers' levels of underlying activity for each indicator detect structural changes in their behavior and generate an alarm. The Kalman filter is a set of mathematical equations that predicts an efficient recursive solution of the method of squared minimums. This solution lets us calculate a linear, unbiased and optimal estimator of the state of a process at any moment of time using the information available at moment *t-1*, and update those estimations with the additional information available at moment *t*.

In this representation the customer is described with a set of variables, called state variables, containing all the information relative to the customer at a certain moment of time. This information should allow customer behavior to be inferred and future behavior to be predicted. Based on the information obtained about past and predicted customer behavior, we will decide to set off an alarm when an important structural change is detected. First we will define the model used in the Kalman filter and then we will specify the criteria used to set off alarms according to the results of the filter.

3.1 Definition of a User Model by means of Kalman Filter

The state of the filter is characterized by two variables:

- $\hat{x}_{k|k}$, the estimator of the state in time *k*. (i.e. the number of transactions)
- $P_{k|k}$, the error covariance matrix (precision of the estimation of the state)

Prediction

- $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k$ (prediction of the state)
- $P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$ (prediction of the estimated covariance)

Update

- $\hat{y}_k = z_k - H_k \hat{x}_{k|k-1}$; $S_k = H_k P_{k|k-1} H_k^T + R_k$
- $K_k = P_{k|k-1} H_k^T S_k^{-1}$ (Kalman gain)
- $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \hat{y}_k$ (Updated estimation of the state)
- $P_{k|k} = (I - K_k H_k) P_{k|k-1}$ (Estimation of the updated covariance)

where z_k is the real value of the state at time *k*.

The prediction phase uses the estimation from the previous step to estimate the current state. In the update phase, the information about current step measurements is used to refine the prediction and obtain a new, more precise estimator.

3.2 Simplification of the Model

To create the customer activity model with the Kalman filter, we formulate the following hypotheses to simplify the model:

- The customer activity for period *t* **will be the same** as in period *t-1*. Therefore, the transition matrix *F* is the identity matrix.
- We do not know which external factors can modify customer behavior. Therefore, the vector u_k is void.
- The measurement of the variable is correct; therefore the matrix *H* is the identity matrix.

With these assumptions our simplified user model is as follows:

Prediction

- $\hat{x}_{k|k-1} = \hat{x}_{k-1|k-1}$ (prediction of the state)
- $P_{k|k-1} = P_{k-1|k-1} + Q_k$ (estimated prediction of the covariance)

Update

- $\hat{y}_k = z_k - \hat{x}_{k|k-1}$; $S_k = P_{k|k-1} + R_k$
- $K_k = P_{k|k-1} S_k^{-1}$ (Kalman gain)
- $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \hat{y}_k$ (updated estimation of the state)
- $P_{k|k} = (I - K_k) P_{k|k-1}$ (updated estimation of the covariance)

3.3 Filtering Process

According to the previous simplified user model, the filtering process will consist of:

<u>Initialization</u>

- Establish initial state values: *x* ($x_{0|0}$) and *P* ($P_{0|0}$)

<u>Iterations (beginning with *k=1*)</u>

- Prediction calculation:

  $\hat{x}_{k|k-1} = \hat{x}_{k-1|k-1}$; $P_{k|k-1} = P_{k-1|k-1} + Q_k$

- Update: the update is done with the known value of the analyzed variable at moment *k* ("Input Value" column of the table of significant data, which is the real value of the variable or the seasonally adjusted value when the series is seasonally adjusted previously) and the values calculated in the prediction.

  $\hat{y}_k = z_k - \hat{x}_{k|k-1}$; $S_k = P_{k|k-1} + R_k$; $K_k = P_{k|k-1} S_k^{-1}$

  $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \hat{y}_k$ $P_{k|k} = (I - K_k) P_{k|k-1}$

As we can see, these equations use the variables Q_k and R_k, which are the errors at moment *k*. We propose calculating them in the following way:

$$R_k = Q_k = \bar{x}_k$$

where $\bar{x}_k$ is the median in the last *k* periods.

Example:

In Table 1 we can see an example of the Kalman filter calculation for a DHL customer and a specific indicator. Initial values have been set at $x_{0|0} = 10$ and $P_{0|0} = 5$.

Table 1. Results of experiment 1

k	z_t	$\hat{x}_{k\|k-1}$	$P_{k\|k-1}$	S_k	K_k	$\hat{x}_{k\|k}$	$P_{k\|k}$
0						10	5
1	11	10	5.1	5.85	0.8718	108.718	0.6538
2	11	108.718	0.7538	15.038	0.5013	109.361	0.376
3	9	109.361	0.476	1.226	0.3882	101.844	0.2912
4	12	101.844	0.3912	11.412	0.3428	108.068	0.2571
5	7.5	108.068	0.3571	11.071	0.3225	97.402	0.2419
6	8	97.402	0.3419	10.919	0.3131	91.953	0.2348
7	10	91.953	0.3348	10.848	0.3087	94.437	0.2315
8	6.5	94.437	0.3315	10.815	0.3065	85.414	0.2299
9	9.5	85.414	0.3299	10.799	0.3055	88.342	0.2291
10	6.5	88.342	0.3291	10.791	0.305	81.223	0.2287
11	7	81.223	0.3287	10.787	0.3047	77.803	0.2286
12	5	77.803	0.3286	10.786	0.3046	69.333	0.2285
13	5	69.333	0.3285	10.785	0.3046	63.445	0.2284
14	5	63.445	0.3284	10.784	0.3045	5.935	0.2284
15	6	5.935	0.3284	10.784	0.3045	59.548	0.2284
16	4	59.548	0.3284	10.784	0.3045	53.595	0.2284
17	6	53.595	0.3284	10.784	0.3045	55.546	0.2284
18	7.5	55.546	0.3284	10.784	0.3045	6.147	0.2284
19	5	6.147	0.3284	10.784	0.3045	57.977	0.2284
20	5	57.977	0.3284	10.784	0.3045	55.548	0.2284
21	4	55.548	0.3284	10.784	0.3045	50.813	0.2284
22	6	50.813	0.3284	10.784	0.3045	53.611	0.2284
23	8	53.611	0.3284	10.784	0.3045	61.647	0.2284
24	10	61.647	0.3284	10.784	0.3045	73.326	0.2284
25	11	73.326	0.3284	10.784	0.3045	84.494	0.2284
26	10	84.494	0.3284	10.784	0.3045	89.216	0.2284
27		89.216	0.3284				

Table 2. Results of experiment 2

k	z_t	*Baseline*	$\hat{x}_{k\|k-1}$	$P_{k\|k-1}$	$\hat{x}b^{+}_{k\|k-1}$	$\hat{x}b^{-}_{k\|k-1}$
1	11	11	10	5.1	144.263	55.737
2	11	11	108.718	0.7538	125.736	9.17
3	9	11	109.361	0.476	122.883	95.839
4	12	1.033.333	101.844	0.3912	114.103	89.586
5	7.5	10.75	108.068	0.3571	11.978	96.355
6	8	10.1	97.402	0.3419	108.863	85.941
7	10	9.75	91.953	0.3348	103.295	80.611
8	6.5	9.785.714	94.437	0.3315	105.721	83.152
9	9.5	9.375	85.414	0.3299	96.671	74.156
10	6.5	9.388.889	88.342	0.3291	99.586	77.098
11	7	9.1	81.223	0.3287	92.461	69.985
12	5	8.909.091	77.803	0.3286	89.038	66.568
13	5	8.583.333	69.333	0.3285	80.567	5.81
14	5	8.083.333	63.445	0.3284	74.678	52.213
15	6	7.583.333	5.935	0.3284	70.583	48.118
16	4	7.333.333	59.548	0.3284	7.078	48.316
17	6	6.666.667	53.595	0.3284	64.827	42.363
18	7.5	6.541.667	55.546	0.3284	66.778	44.314
19	5	6.5	6.147	0.3284	72.702	50.238
20	5	6.083.333	57.977	0.3284	69.209	46.745
21	4	5.958.333	55.548	0.3284	6.678	44.316
22	6	5.5	50.813	0.3284	62.045	39.581
23	8	5.458.333	53.611	0.3284	64.843	42.379
24	10	5.541.667	61.647	0.3284	72.879	50.415
25	11	5.958.333	73.326	0.3284	84.558	62.094
26	10	6.458.333	84.494	0.3284	95.726	73.262
27		6.875	89.216	0.3284	100.448	77.984

4. Alarm Activation

Alarms are activated according to a baseline. The baseline tells us the normal level of activity of the customer. When real activity is "considerably" far from the baseline, an alarm goes off.

To determine this "considerable" distance, we define activity bounds, which establish a confidence interval around the activity of the customer. The alarm will sound when the activity bounds do not include the baseline of the customer.

Three activity bounds will be defined: the most restrictive (closest to the activity of the customer) will generate low-level alarms; the middle one will generate mid-level alarms; and the least restrictive (most removed from the activity) will generate the alarms at the highest level.

4.1 Activity Bounds

The activity bounds are calculated with the following formula:

$$xb_i = x_i \pm r \cdot \sqrt{P}$$

where x_i is the value of the activity, P is the variance of this value (so $\sqrt{P}$ is the typical deviation), and r is the tolerance index which, in a Gauss distribution, must be 1.96 to include 95% of the values, or 2.58 to include 99% of them.

4.2 Customer baseline

The customer baseline is calculated using the moving average method, which incorporates seasonal variation. This method contains five parts. The first four are optional and make up part of the process of seasonal adjustment of the series.

1. Calculation of the seasonally adjusted baseline.
2. Median of "Time Window" periods.

Calculation of the seasonally adjusted baseline

$$D_t = \frac{Y_t}{IVE_t}$$

where

$$IBVE_t = \frac{Y_t}{MM_t}, \quad \text{and} \quad IVE_h = IVE*_h \frac{L}{\sum_{j=1}^{L} IVE*_j} \quad \text{and} \quad IVE*_h = \frac{\sum_{t \in h} IBVE_t}{T-1}$$

and the calculation of the moving averages is as follows: Given a time window L, the moving averages for each period t are calculated, for odd L

$$MM(L)_t = \frac{1}{L} \sum_{i=t-\frac{L-1}{2}}^{t+\frac{L-1}{2}} y_i$$

where y_i is the value of the variable in the period i. Meanwhile, for even L (in this case the moving average with L = 2 is done again):

$$MM(L)_{t+0,5} = \frac{1}{L} \sum_{i=t-\left(\frac{L}{2}-1\right)}^{t+\frac{L}{2}} y_i$$

Calculation of the final baseline

A median is used for the baseline using the "Time window" parameter (N):

$$Baseline_t = \frac{\sum_{i=t}^{t+N} D_i}{N}$$

Example:

In Table 2 we can observe the calculation of the baseline and of the activity bounds of a customer; in this case only activity bounds of a single level have been created. In the example the values r=1.96 and n=12 have been chosen.

4.3 Example

In this plot we can see the activity of the customer from Table 1:

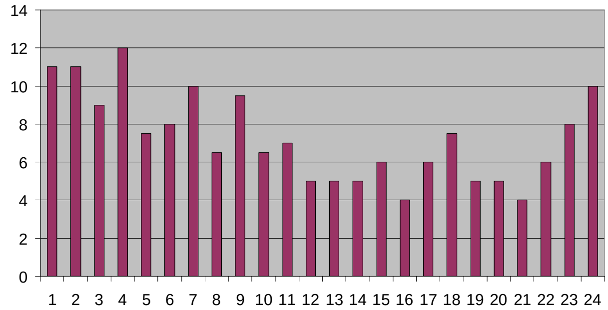

Fig. 2. Consumption of a DHL customer during 24 months

The result of the Kalman filter on said customer, adding the calculation of the activity bands and baseline is the following:

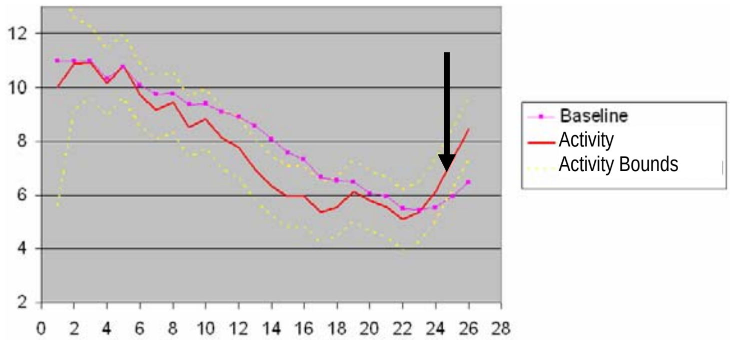

Fig. 3. Activity baseline and higher and lower bounds

We see how at the indicated point an alarm would sound, given that the trust interval of the prediction does not include the baseline of the customer. Each warning (alarm) will have an associated impact. The impact will be calculated with the formula given in the table of significant variables.

Adjusting the Parameters

As we have seen, there is a series of parameters that must be calibrated in the system:

- $x_{0|0}$: Value of the initial state.
- $P_{0|0}$: Value of the covariance of the initial state.
- r : Tolerance index used in the delimitation of the activity bounds.
- n : Number of periods used in the time window where the customer baseline

The values of r and n can be global for all the customers and indicators, but the parameters $x_{0|0}$ and $P_{0|0}$ must be adjusted individually.

- r is the tolerance index, used to calculate the activity bounds. In a Gauss distribution it is 1.96 to include 95% of the values, or 2.58 to include 99% of them. We propose **1.96** for the 1st bound, **2.58** for the 2nd and **3.5** for the 3rd.
- n, is used to calculate the baseline of the customer and represents the number of previous periods that are taken into account. Depending on the seasonal variation of the activity of customers, values of 12, 18 or 24 (months) can be chosen. In the DHL example we propose **12** months.

5. Future Work

The Kalman filter is applicable to predict the behavior of customers from a dynamic point of view. However, the individual values belong to customers and are calculated according to their activity. In conclusion, we propose the individual values as:

$$x_{0|0} = x_{1|1} \quad \text{and} \quad P_{0|0} = \left(\frac{x_{0|0}}{2r}\right)^2 - Q_{1|1}$$

In the future, we will compare this Kalman method of prediction with the interval uncertainty model [1] to create autotuning methods for this filter to avoid overbounding, and regarding the particular type of customers of every company. We expect the number of false behavior warnings will be less than it is with other methods.

Acknowledgements

The PSM company, the Spanish Ministry of Science and Technology (project TIN2006-15111/Estudio sobre Arquitecturas de Cooperación) and the EU (project No. 34744 ONE: Open Negotiation Environment, FP6-2005-IST-5, ICT-for Networked Businesses).

References

[1] Armengol J., Travé-Massuyès L., Vehí J., and de la Rosa J. Ll., A survey on interval model simulators and their properties related to fault detection ISSN: 1367-5788, Annual Reviews in Control**,** Vol. 24, No. 1. pp. 31-39, Springer-Verlag, Oxford, UK, 2002.

[2] Berry M., Linoff G. (1997). Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley. New York.

[3] Acebo, E., Muntaner, E., *Una plataforma para la Simulación de Sistemas Multiagente Programados en Agent-0*. Workshop de Docencia en Inteligencia Artificial, Congreso de la Asociación Española para la Inteligencia Artificial CAEPIA-TTIA, Gijón (2001). Utrecht University (2003).

[4] Bezdek J. C. (1981). Pattern recognition with fuzzy objective functions. Plenum Press. New York.

Gan, F. (1995). Joint monitoring of process mean and variance using exponentially weighted moving average control charts. Technometrics, 37: 446-453.

[5] Kalman, Rudolph E., A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME-Journal of Basic Engineering, Vol. 82, Series Dm pp. 35-45, 1960.

[6] Oliva, F., De Caceres, M., Font, X., and Cuadras, C. M. (2001). Contribuciones desde una perspectiva basada en distancias al fuzzy C-means clustering. XXV Congreso Nacional de Estadística e Investigación Operativa. Úbeda, 2001.

[7] J.Ll. de la Rosa, La Lucha por la fidelidad del Consumidor, Internet Global Conference IGC 2002, Internet Global Conference 4ª Edición, Barcelona, April 27-30, 2002

[8] Shumway, R. and Stoffer, D. (2000). Time series analysis and its applications. Springer-Verlag.

[9] West, M. and Harrison, J. (1999). Bayesian Forecasting and Dynamic Models. Springer-Verlag.