ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Visualizing pay-per-view television customers churn using cartograms and flow maps

David L. García[1] and Àngela Nebot[1] and Alfredo Vellido[1] *

1- Dept. de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya. Barcelona TECH - Spain

Abstract. Media companies aggressively compete for their share of the pay-per-view television market. Such share can only be kept or improved by avoiding customer defection, or churn. The analysis of customers' data should provide insight into customers' behavior over time and help preventing churn. Data visualization can be part of this analysis. Here, a database of pay-per-view television customers is visualized using a nonlinear manifold learning model. This visualization is enhanced through, first, the reintroduction of the local nonlinear distortion using a cartogram technique and, second, the visualization of customer migrations using flow maps. Both techniques are inspired by geographical representation.

1 Introduction

Faced by increasing competition, media companies must aggressively compete for their share of pay-per-view television markets. Such share can only be kept or improved by avoiding customer defection, or churn, a phenomenon faced by any service provider across industries [1]. The analysis of customers' data should provide insight into customers' behavior over time and help preventing churn. Exploratory data visualization can be part of this analysis. In this study, a database of pay-per-view television customers is explored through visualization using a nonlinear dimensionality reduction (NLDR) method.

Many NLDR methods have been proposed over the last few years [2]. Some of them belong to the manifold learning family. Methods of this family attempt to describe multivariate data through nonlinear low-dimensional manifolds embedded in the observed data space. The nonlinearity of these methods entails local distortion in the mapping of the data from the observed space into their visualization space, through processes of compression and stretching of the manifold in the form of local *magnification*. As a result of this local distortion, the data visualization is difficult to interpret, given that the coordinates of visual representation are non-trivial combinations of the observed data attributes.

Here, the customer database is analyzed using Generative Topographic Mapping (GTM, [3]), a NLDR latent variable model of the manifold learning family. Knowledge extraction through visualization requires the use of methods that guarantee the interpretability of results [4, 5]. For this, the GTM visualization is here enhanced through, first, the reintroduction of the local distortion using a cartogram technique and, second, the visualization of customer migrations using Flow Maps. Both techniques are inspired by geographical representation.

*This research was partially supported by Spanish project TIN2012-31377.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

2 Methods

2.1 Generative Topographic Mapping and its Magnification Factors

Generative Topographic Mapping [3] is a latent variable model for data visualization and vector quantization, in which a sample of K regularly spaced latent points $k = 1, \ldots, K$ are mapped into the observed data space. Each of them defines a prototype point $\mathbf{y}_k$, which is the image of the former according to $\mathbf{y}_k = \mathbf{W}\Phi(\mathbf{u}_k)$, where Φ is a set of M nonlinear basis functions ϕ_m, and $\mathbf{W}$ is a matrix of adaptive model weights.

The set of prototypes $\mathbf{y}_k$ belongs to a smooth manifold that wraps around the observed D-dimensional data $X = \{\mathbf{x}_n\}_{n=1}^N$. Assuming they lie close to the manifold, the conditional distribution of the observed data variables, given the latent variables, $p(\mathbf{x}|\mathbf{u})$ takes the form of a noise model with variance β^{-1}:

$$p(\mathbf{x}|\mathbf{u}, \mathbf{W}, \beta) = (\frac{\beta}{2\pi})^{D/2} \exp\{-\frac{\beta}{2}\sum_{d=1}^{D}(x^d - y^d(\mathbf{u}))^2\}, \quad (1)$$

From this, we can marginalize the latent variables and obtain an analytical expression for the likelihood of the model. The adaptive parameters of the model can thus be optimized within a maximum likelihood (ML) framework. Details of this procedure can be found, for instance, in [3].

As a result of the ML optimization, we obtain a closed expression for $p(\mathbf{u}_k|\mathbf{x}_n)$, also know as responsibility r_{kn} of each latent point k for the generation of each observed data point n. This probability can be used to obtain data visualization in the form of either a *posterior mode projection* of $\mathbf{x}_n$: $k_n^{mode} = \arg\max_{\{k_n\}} r_{kn}$ (assigning each observed data point to the latent point with the highest responsibility for its generation), or a *posterior mean projection* $\mathbf{u}_n^{mean} = \sum_{k=1}^{K} r_{kn}\mathbf{u}_k$ (so that a data point is placed on latent space continuum according to a responsibility-weighted combination of all latent point locations).

The distortion caused by the nonlinear GTM mapping in each point of the latent (visualization) space can be quantified and it is known as Magnification Factors (MF) [6]. The relationship between a differential area dA (for a 2-D visualization) in latent space and the corresponding area element in the GTM-generated manifold, dA', can be expressed in terms of the derivatives of the basis functions ϕ_m as $dA/dA' = \det^{\frac{1}{2}}(\Psi^T\mathbf{W}^T\mathbf{W}\Psi)$, where Ψ is a $M \times 2$ matrix with elements $\varphi_{mi} = \partial\phi_m/\partial u^i$ and u^i is the i^{th} coordinate ($i = 1, 2$) of a latent point.

2.2 Cartograms and their application to GTM

Cartograms, also known as density-equalizing maps, were originally devised as geographic maps in which the sizes of regions are proportional to underlying quantities of interest, such as their population density. This distortion takes the form of a continuous transformation from an original plane to a transformed one, so that a vector $\mathbf{x} = (x^1, x^2)$ in the former is mapped onto the latter according to $\mathbf{x} \to T(\mathbf{x})$, in such a way that the Jacobian of the transformation is proportional to an underlying *distorting variable* $\mathbf{d}$.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

A method for the creation of cartograms based on the physics principle of linear diffusion processes was proposed in [7]. In this method, the distorting variable $\mathbf{d}$ is let to *diffuse* over the map *over time* so that the final result is a map of uniform distortion in which the original locations have displaced according to the process, while preserving the integrity of the existing borders.

The standard diffusion equation takes the form $\nabla^2\mathbf{d} - \frac{\partial \mathbf{d}}{\partial t} = 0$, which has to be solved for distortion $\mathbf{d}(\mathbf{x}, t)$ to obtain the map location displacement in the cartogram, assuming that the initial condition corresponds to each map fragment being assigned its value of the distorting variable.

Creating cartograms for GTM entails replacing geographic maps by the GTM latent visualization map; geographic boundaries by the square regular grid formed by the lattice of latent points $\mathbf{u}_k$; and assuming that the level of distortion in the space beyond this square is uniform and equal to the mean distortion over the complete map, which is $1/K \sum_{k=1}^{K} J(\mathbf{u}_k)$, where $J = \det^{\frac{1}{2}}(\Psi^T \mathbf{W}^T \mathbf{W} \Psi)$, as described in the previous section. The cartogram method for GTM is described in detail in [8] and for a Batch-SOM model in [9].

2.3 Flow Maps for GTM visualization of customer migrations

Flow Maps are usually combinations of geographical maps and flow graphs that were originally devised to visualize evolution patterns such as population migrations. Again, we propose their use in NLDR-based visualization to display the evolution over time of individual points, here with GTM. With flow maps, we could track individual customers, anticipating the possibility of churn.

A method for the generation of Flow Maps using hierarchical clustering was recently proposed in [10]. In brief, its algorithm operates through six differentiated stages, which, for GTM, are: 1) *Layout adjustment*, enforcing a minimum separation distance among the nodes (in our case, each of the squares in the GTM lattice corresponding to individual latent points in the visualization space); 2) *Primary clustering*: merging of flow edges that share destinations, obtained by agglomerative hierarchical clustering. The resulting binary tree describes the branching structure of the Flow Map; 3) *Rooted clustering*, generated such that the root of the Flow Map is the root of the tree; 4) *Spatial layout*, which actually defines the flow hierarchical tree from the rooted hierarchical cluster solution; 5) *Edge routing*, in which edges are re-routed around the bounding boxes within the same hierarchical cluster to avoid unwanted crosses; 6) *Rendering*, in which each flow edge in the visualization map of GTM is rendered as a catmull-rom spline, generating an interpolation between the nodes of the spatial layout hierarchical tree. Their width is proportional to the magnitude of the flow.

3 Materials

For the experiments reported next, a proprietary database belonging to a Spanish pay-per-view television company was used. It includes monthly data from 33,992 customers, monitored for churn over 7 months, from March to September 2008. Their behaviour is described through 59 variables (not listed in full for lack

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

of space) corresponding to channel usage (36 variables) and customer-company interaction (23 variables, including post-sale, customer, and techical service; complaints and billing).

4 Experiments

The visual exploratory analysis of the experiments aims to identify potential customer churn routes through the combination of: a) The visualization of customer usage patterns using GTM. b) The enhancement of this visualization using a Cartogram representation. c) The visual representation of monthly customers' transitions using Flow Maps, aiming to discover potential churn and customer retention routes over the representation map of GTM.

The adaptive parameters of the GTM were initialized according to a standard procedure described in [3]. A 10×10 grid for the GTM lattice was used in all the experiments. The GTM input to the Flow Map algorithm includes: The GTM map layout (visualization lattice); the GTM model for the different months, in the form of the assignment of each customer to a given lattice node; and the flow from month-to-month GTM representations, in the form of cumulative customer information for each of the lattice nodes.

4.1 Results and discussion

The projection of the 33,992 customers in the GTM map is shown in Fig.1. The visualization includes the mean projection (top row, left) and the mode projection (top row, center) of the March 2008 data. They are accompanied by a visualization of the MF as a colour map, with white indicating highest MF and black, lowest MF (bottom row, left) and the corresponding cartogram of the MF with the mean projection overlaid onto it (bottom row, center). Beyond the distribution of individual customers, we are interested in their commercial TV package usage, which is identified in the segment partition of the GTM map also depicted in Fig.1 (top row, right), and in the spread over the map of the customer churn rate, which is again colour-coded with white indicating the highest churn rate and black, the lowest (bottom row, right).

Although similar results are available for the complete analyzed period, they are not shown due to space limitations. The same can be said for the corresponding flow maps: there is one for each of the 100 GTM map nodes for each month. For illustration, Fig.2 shows the detailed migration routes and the customer flow *from* a high-churn node (*number 84*, highlighted in Fig.1, bottom row, right).

Both data projections in the top row of Fig.1 suggest the existence of data structure, with both densely populated and completely empty spaces in the map. The display of the MF map in Fig.1 (bottom row, left) hints the existence of general segments, and this is visually emphasized by the cartogram display in Fig.1 (bottom row, center), which enhances the *mean projection* (top row, left) by explicitly reintroducing the MF-quantified distortion into the map. At least to some extent, this structure is likely to be dominated by separation due to customer diversity in their usage of distinct pay-per-view packages. This is

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

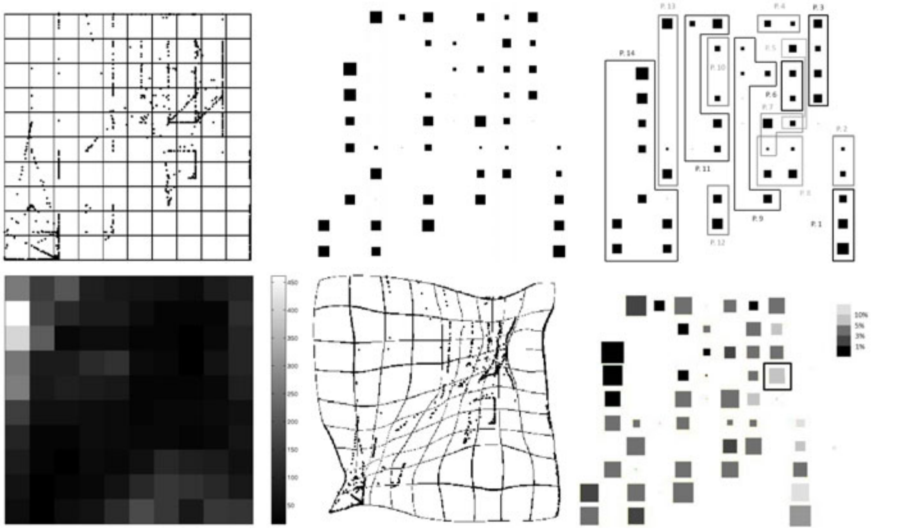

Fig. 1: Top row: *left*) *mean projection* of the customers over the GTM visualization lattice; *center*) corresponding *mode projection*, with relative square size in proportion to the ratio of customers assigned to individual nodes; *right*) Customer segments according to service packages, overlaid on top of the *mode projection*. Bottom row: *left*) MF colour map, with highest distortion in white; *center*) corresponding cartogram distorting the GTM lattice according to the MF; *right*) churn ratio as a colour map superimposed to the *mode projection*.

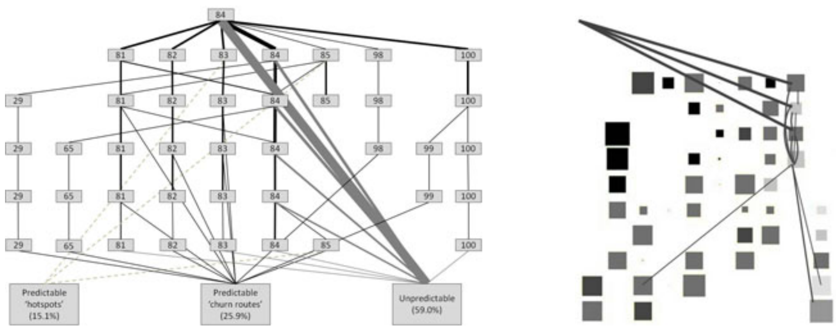

Fig. 2: Left) Customer migrations from GTM *node 84*. Nodes are numbered according to their column-wise location in the 10×10 GTM map (leftmost column, top-to-bottom: nodes 1-10; rightmost column: 91-100). Nodes are grouped into three categories: predictable *hotspots*, predictable churn routes and unpredictable ones. The thickness of lines is proportional to the volume of migration. Right) Corresponding Flow Map, showing the *geography* of the customer migrations outflowing from *node 84* over the GTM visualization map. Lines projected out of the map indicate churn.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

confirmed by the delimitation of customer segments shown in Fig.1 (top row, right), where packages are described as p.$\sharp n$. The right-hand side ones are mostly basic and economic commercial packages, whereas the ones on the center are mid-range thematic TV channels, and the ones on the left are premium packages. This structure reflects the company's commercial strategy.

The visualization of the churn rate by colour coding the GTM *mode projection* (bottom row, right) is most revealing, indicating the high propensity to churn in customers mapped into the bottom, right-hand side corner of the GTM map, which are fairly isolated from the rest of customers. This area corresponds to basic channel packages. The use in isolation of these non-specialized channel bundles thus reveals itself as the main gate to customer churn for the company.

The migration routes (Fig.2, left) for *node 84* allow us to identify *predictable* and *unpredictable* churn routes as well as *hot spots*, which are areas in the GTM map that would require preferential attention. These routes are overlaid in the GTM as Flow Maps in Fig.2, right. A total of 25.9% of churners with origin in this node show behaviours (in the form of movements across the visual map) that could be anticipated up to 3 months in advance.

In summary, the combined use of GTM maps, their cartogram representation and Flow Maps allows the data analyst to anticipate and prevent churn. For the analyzed company, it helped to anticipate 42.5% of the 3.795 customer requests for service cancellations over the 7 month period. A total of 75.2% of them were deemed to be suitable for preventive commercial action.

References

[1] J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, 34(10):2902–2917, 2007.

[2] J. A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer Verlag, 2007.

[3] C. M. Bishop, M. Svensén, and C. K. I. Williams. GTM: The Generative Topographic Mapping. *Neural Computation*, 10(1):215–234, 1998.

[4] A. Vellido, J. D. Martín, F. Rossi, and P. J. G. Lisboa. Seeing is believing: The importance of visualization in real-world machine learning applications. In *ESANN'11*, pages 219–226. d-side pub., 2011.

[5] A. Vellido, J. D. Martín, and P. J. G. Lisboa. Making machine learning models interpretable. In *ESANN'12*, pages 163–172. d-side pub., 2012.

[6] C. M. Bishop, M. Svensén, and C. K. I. Williams. Magnification factors for the GTM algorithm. In *IEE Fifth International Conference on Artificial Neural Networks*, pages 64–69. IEE, 1997.

[7] M. T. Gastner and M. E. J. Newman. Diffusion-based method for producing density-equalizing maps. *PNAS*, 101(20):7499–7504, 2004.

[8] A. Vellido, D. García, and À. Nebot. Cartogram visualization for nonlinear manifold learning models. *Data Mining and Knowledge Discovery*, doi:10.1007/s10618-012-0294-6.

[9] A. Tosi and A. Vellido. Cartogram representation of the batch-SOM magnification factor. In *ESANN'12*, pages 203–208. d-side pub., 2012.

[10] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow Map Layout. In *InfoVis'05, IEEE Symposium on Information Visualization*, pages 219–224. IEEE, 2005.