Analysis of Cluster Migrations Using Self-Organizing Maps

Denny[1,2], Peter Christen[1], and Graham J. Williams[1,3]

[1] Research School of Computer Science, The Australian National University, Canberra, Australia
[2] Faculty of Computer Science, University of Indonesia, Indonesia
[3] Australian Taxation Office, Canberra, Australia
denny@cs.ui.ac.id, peter.christen@anu.edu.au, graham.williams@ato.gov.au

Abstract. Discovering cluster changes in real-life data is important in many contexts, such as fraud detection and customer attrition analysis. Organizations can use such knowledge of change to adapt business strategies in response to changing circumstances. This paper is aimed at the visual exploration of migrations of cluster entities over time using Self-Organizing Maps. The contribution is a method for analyzing and visualizing entity migration between clusters in two or more snapshot datasets. Existing research on temporal clustering primarily focuses on either time-series clustering, clustering of sequences, or data stream clustering. There is a lack of work on clustering snapshot datasets collected at different points in time. This paper explores cluster changes between such snapshot data. Besides analyzing structural cluster changes, analysts often desire deeper insight into changes at the entity level, such as identifying which attributes changed most significantly in the members of a disappearing cluster. This paper presents a method to visualize migration paths and a framework to rank attributes based on the extent of change among selected entities. The method is evaluated using synthetic and real-life datasets, including data from the World Bank.

Keywords: temporal cluster analysis, cluster migration analysis, visual data exploration, change analysis, Self-Organizing Map.

1 Introduction

Migration analysis tracks changing behaviors of a population and monitors the changes in clusters over time. This can, for example, be applied to group customers' purchase patterns to identify high-responsive customers for marketing promotions [10]. Moreover, it can be used to perform customer attrition analysis to prevent churning (customer turnover or loss of customers) from happening in the future. In industries such as telecommunications where getting new customers is costlier than retaining existing ones, knowledge about churning customers is important [10]. With this knowledge, strategies can be derived to prevent churn. For example, giving free incentives and getting feedback on customer satisfaction were recently used to target customers of a pay-TV company

L. Cao et al. (Eds.): PAKDD 2011 Workshops, LNAI 7104, pp. 171–182, 2012.

who had a high probability to churn [1]. Migration analysis can be used not only for analyzing churning customers, but also to analyze migrations from a cluster of highly profitable customers to a cluster of average customers.

While recent work in this area [3,5] focused on the analysis of changes in the structure of clustering, this paper focuses on analyzing cluster changes at the micro level (entity migration), i.e. changes in cluster membership. For example, analysts might be interested to know the attributes that have changed significantly for those entities that belong to a lost cluster, or they might want to know where entities from one cluster have migrated to in the next snapshot dataset.

This paper contributes a method for analyzing and visualizing entity migrations using Self-Organizing Maps (SOMs) between two snapshot datasets. The paper presents a visualization method to reveal the path of migration of entities, and a framework to rank attributes based on the extent of their change among selected entities. It is worth to note that this paper does not address time series clustering [11], where the aim is to cluster time series with similar patterns. The approach presented here clusters entities at points in time (snapshots), and analyzes cluster migrations between such snapshots.

SOMs are useful for visualizing clustering structure changes, and they have several advantages in temporal cluster analysis. First, they are able to relate clustering results by linking multiple visualizations, and they can detect various types of cluster changes, including emerging clusters and disappearing clusters [3,5]. Second, SOMs create a smaller but representative dataset that follows the distribution of the original dataset [7]. Third, SOMs topologically map a high-dimensional dataset into a two-dimensional map with similar entities being placed close to each other [7]. Importantly, SOMs offer various map visualizations [14] that allow non-technical users to explore high-dimensional data spaces through a non-linear projection onto a two-dimension plane.

This paper analyzes the migrations of entities given two observations for each entity in two datasets, $D(\tau_1)$ and $D(\tau_2)$, and two trained SOM maps, $M(\tau_1)$ and $M(\tau_2)$, with τ_1 and τ_2 being the time snapshots of the two datasets. The datasets should contain the same type of information on the same entities at different points in time, therefore all entities in dataset $D(\tau_1)$ should exists in dataset $D(\tau_2)$, and vice versa. This type of data is known as longitudinal data in statistics [6] which enables analyzing migrations and measuring change.

2 Self-Organizing Maps

A SOM is an artificial neural network that performs unsupervised competitive learning [7]. Artificial neurons are arranged on a low-dimensional grid, commonly a 2-D plane. Each neuron j has an d-dimensional prototype vector, m_j, where d is the dimensionality of dataset D. Each neuron is connected to neighboring neurons, with distances to its neighbours being equidistant in the map space. Larger maps generally have higher accuracy and generalization capability [15], but they also have higher computation costs.

Before training a map, the prototype vectors need to be initialized [7]. The preferred way is to use linear initialization which uses the first two largest principal components. At each training step t, the best matching unit b_i (BMU) for training data vector x_i from a dataset D, i.e. the prototype vector m_j closest to the training data vector x_i, is selected from the map. Using the batch training algorithm [7], the values of new prototype vectors $m_j(t+1)$ at training step $(t+1)$ are the weighted averages of the training data vectors x_i at training step t, where the weight is the neighbourhood kernel value h_{b_ij} centered on the best matching unit b_i (commonly Gaussian).

A SOM can be visualized using various methods that allow non-technical users to explore a dataset. These methods include component plane visualizations and u-matrix visualization [14]. The prototype vectors of a trained SOM can be treated as 'proto-clusters' serving as an abstraction of the dataset. The prototype vectors are then clustered using a traditional clustering technique to form the final clusters [15].

3 Related Work

The similarity between two datasets can be shown by comparing two data hit histogram visualizations [14]. A data hit histogram shows the frequency with which data vectors are mapped to each node. This technique can indicate changes over time in the data mapping, but it is difficult to interpret these changes simply by comparing the data hit histograms, because this kind of visualization introduces inconsistencies in visualizing changes in datasets. If a vector is mapped into a dense area of the SOM, a small change in the data may cause it to be mapped to a different node. On the other hand, when a data vector is mapped into a sparse area, the same magnitude of change in the data vector might not cause this vector to be mapped to a different node. In addition, this visualization is not designed to show the migration paths of entities because it does not link the same entities on these two map visualizations.

Another way to visualize the movement of individual entities over time is by using the trajectory of an entity on a SOM. However, this technique can only show the movement of one entity. It would be very cluttered to show trajectories of many entities simultaneously. This visualization has been used to observe the evolution of a bank in terms of their financial data [13]. When an entity migrates to a new cluster in period τ_2 that is not represented on the original SOM (map $M(\tau_1)$), it leads to high quantization error in the visualization of the entity on map $M(\tau_2)$.

Changes in entities over time can be visualized using Trendalyzer developed by Gapminder[1]. This visualization shows multivariate time series data using an *animated bubble chart*. The bubble chart can plot three numerical attributes. The position of a bubble on the X and Y axes indicates values of two numerical attributes, and the size of the bubble represents the value of a third numerical attribute. The X and Y axes can use linear or log scale. The colour of the bubble can be used to represent a categorical attribute. The time dimension is controlled by a sliding bar. As the time advances, the position and the size of the bubbles

[1] `http://www.gapminder.org/`

change to reflect values at the time. The slider can be used to choose the time and animate the movement of the bubbles. While this visualization is suitable to illustrate changes over time, it is not effective for large datasets. A large number of bubbles would make the visualization too dense, thus making it difficult to discover patterns in large datasets. Furthermore, this visualization can only show three numerical attributes and one categorical attribute at one time.

Migration patterns can also be analyzed by finding changes in cluster composition. An approach to address changes in cluster composition over time and changes in cluster membership of individual data points has been proposed [8]. First, the data of the first period is clustered using k-means. Then, the resulting clusters are labeled based on descriptive statistics of the clusters. This analysis is repeated for subsequent periods. A bar chart is used to visualize changes in cluster size over time, and a pie chart to show the proportion of each cluster at a certain period of time. This method cannot detect new clusters, lost clusters, and the movement of cluster centroids, because it assumes that the structure of the subsequent clustering results would be the same as the first one.

Existing methods to analyze cluster migration have limitations. For example, the SOM trajectory method [13] is not suitable for large datasets. Other methods assume subsequent clustering structures in a snapshot dataset to be the same as in the previous one. The migration analysis method presented in this paper is suitable for analyzing migrations of large snapshot datasets and analyzing attribute interestingness based on changes in a selected group of migrants.

4 Visualizing Migrations

Because SOMs follow the distribution of the dataset it trained on, there is a problem with visualizing cluster migration paths on a single map. When a new region emerges in dataset $D(\tau_2)$, this region is not represented on map $M(\tau_1)$. As a result, when visualizing entities that migrate to a new region in period τ_2, this region is not represented by a prototype vector on map $M(\tau_1)$, as is illustrated in Figure 1. Similarly, when visualizing cluster migration paths only on map $M(\tau_2)$, entities which migrate from a lost cluster cannot be visualized properly, because the lost region is no longer represented on map $M(\tau_2)$. Therefore, migration paths that are involved with new clusters or lost clusters will not be shown properly by using one map only (either map $M(\tau_1)$ or map $M(\tau_2)$).

Migration paths should be visualized using both maps, because they follow the distribution of the dataset they were trained on. However, showing migrations using two 2D maps can be complicated. For example, a three-dimensional visualization could be used to show migrations between two 2D maps, but this can result in a cluttered visualization if many migration patterns are shown. As an alternative, migrations can be shown using two 1D maps. The challenge is to transform 2D maps into 1D maps.

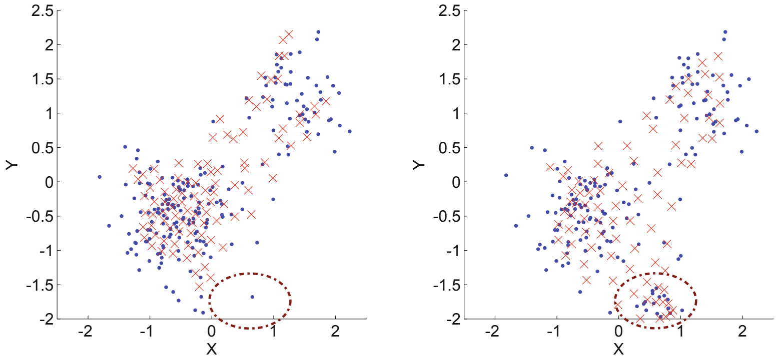


Fig. 1. Data (blue dots) and prototype vectors (red crosses) of the maps $M(\tau_1)$ (*left*) and $M(\tau_2)$ (*right*) trained using two synthetic datasets $D(\tau_1)$ and $D(\tau_2)$ respectively. A new cluster has emerged in dataset $D(\tau_2)$ (marked with the red circle in the right hand plot). It is not represented by any prototype vectors on map $M(\tau_1)$.

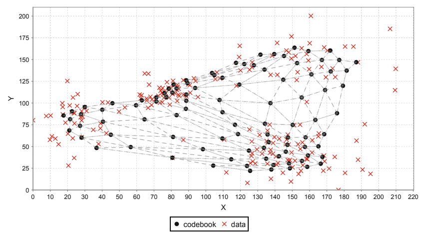
4.1 Transforming 2D Maps to 1D Maps

There are two approaches to transforming a SOM from 2D to 1D. In the first, a 1D map can be created by training the map using the prototype vectors of the 2D map as the training vectors. Experiments performed in this research show that this approach is not effective to visualize migrations, as discussed in the following section. In the second approach, a 1D map is created based on the clustering result of the 2D map.

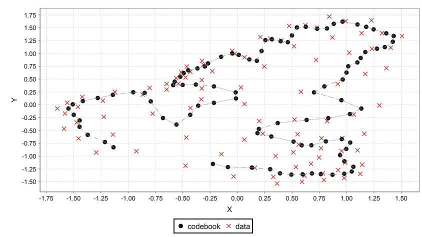
Training a 1D Map Using Prototype Vectors of the 2D Map. In transforming a 2D map to a 1D map, it will be easier to analyze the 1D map if the topological order of the 2D map is preserved. One way to achieve this is by using a SOM to order the prototype vectors of the 2D map onto a 1D map. The 1D map (e.g. Figure 2(b)) with the size $|M| \times 1$ is trained with the prototype vectors of the 2D map (e.g. Figure 2(a)), where $|M|$ is the number of units of the 2D map. The trained 1D map will have different prototype vectors compared to the initial 2D map, even though it is topologically ordered. After that, the prototype vectors of the 2D map are sorted topologically with the previously trained 1D map as the reference.

However, map folding might occur in the 1D map as shown in Figures 2(b) and 2(c). This could happen because the SOM attempts to follow the distribution of a higher dimensional dataset. For example, the ordered 1D map shown in Figure 2(b) is then clustered using k-means with four cluster (chosen based on the plot of the Davies-Bouldin Index [2]). In this figure, there are two nodes of the green cluster that are separated from the main green cluster. This scattered green cluster will make it harder for an analyst to see migration patterns, even though the map is topologically ordered.

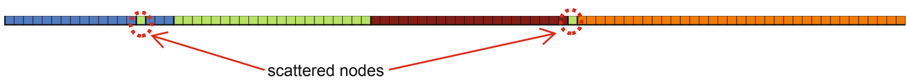
Showing migration from node to node between two 1D maps (such as the one shown in Figure 2(c)) can be difficult to analyze. The many migration paths



(a) Data vectors and prototype vectors of a 2D map trained using a synthetic dataset. The grey line shows connection between neighbouring units.



(b) Data vectors and prototype vectors of the 1D map trained using prototype vectors of the trained 2D map shown in Figure 2(a) as the training dataset.



(c) The clustering result of the 1D map shown in Figure 2(b) with four clusters. Notice that there are two folds (scattered nodes of the green cluster).

Fig. 2. Training a 1D map using prototype vectors of the 2D map

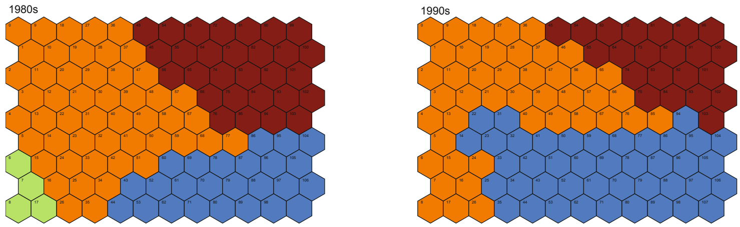
would result in a too crowded and too difficult to understand visualization. Arrows can come from most of the nodes on map $M(\tau_1)$ to the nodes on map $M(\tau_2)$. Alternatively, showing migration from cluster to cluster is easier to see and simpler, because the number of cluster is relatively smaller. In the next subsection, a migration visualization from cluster to cluster is developed.

Creating a 1D Map Based on the Clustering Result of a Trained 2D Map. In this approach, a 1D map is created based on a clustering result of a trained 2D map selected by a user. The user can use a plot of the Davies-Bouldin index [2] as a guide to choosing the optimal clustering result of the 2D map. The 1D map is visualized using stacked k bars where k is the number of clusters as shown in Figure 3(b). The height is in proportion to the size of the clusters of the 2D map. Colour is used to link the cluster on the 2D map and on the 1D map, as shown in Figure 3.

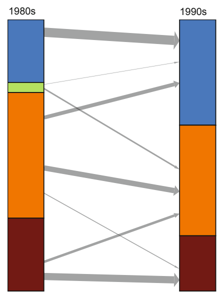
The migration paths are visualized (Figure 3(b)) using lines from the left hand map (period τ_1) to the right hand map (period τ_2). Using line arrows could clutter the visualization for large datasets with many clusters. Therefore, to avoid cluttered visualization of migration paths, a user can select interactively which migration paths from or to a cluster to show.

4.2 Visualizing Migrations

The number of entities of each migration path can be represented using colour or line thickness. Here, line thickness is used based on rank of number of migrants. To enable better visual data exploration, the visualization of migration paths is interactive. A user can acquire more details by right clicking on a migration



(a) Clusters of the maps



(b) Visualization of the migration paths

Fig. 3. The migration path visualization of the clusters on maps trained using World Development Indicator (WDI) datasets from the 1980s (*left*) and the 1990s (*right*). The green cluster located in the bottom left corner on the 1980s map is missing in the 1990s map.

path, a source cluster, or a destination cluster. The selected path or cluster is then analyzed as discussed in the next section.

5 Analyzing Attribute Interestingness of Migrants

As mentioned earlier, it can be of interest to know which attributes of a selected group of entities change more than others. A framework is developed to analyze attribute interestingness that can be applied to a variety of measures. Attribute interestingness related to migrations can be measured based on changes in attribute values for the selected entities between datasets $D(\tau_1)$ and $D(\tau_2)$. Then, the measurement results for each attribute are presented in a sorted table, as shown in Table 1.

Let $\mathcal{S}_{node}$ be the set of node indexes on map $M(\tau_1)$ that are selected by a user using brushing. The selection can be made based on ReDSOM visualizations [5], or clustering result visualizations (e.g. Figure 3(a)). Let $\mathcal{S}_{data}$ be the set of selected data indexes. $\mathcal{S}_{data}$ can be derived by mapping data vectors to nodes in $\mathcal{S}_{node}$ on map $M(\tau_1)$:

$$\mathcal{S}_{data} \leftarrow (\mathbf{x}_i \in D(\tau_1) \mid BMU\ (\mathbf{x}_i, M(\tau_1)) \in \mathcal{S}_{node}) \tag{1}$$

Alternatively, entities can be selected based on their involvement in a selected migration path shown in migration visualization, such as in Figure 3(b).

In measuring change, each entity can be represented by its best matching unit (BMU) on the trained SOM, instead of using its actual value. This method is more efficient for large datasets, since it avoids accessing the whole dataset. The BMU of entities $\mathbf{x}_i(\tau_1)$ on map $M(\tau_1)$ is the first measurement, where $i \in \mathcal{S}_{data}$. Similarly, attribute values of the BMUs of the same entities $\mathbf{x}_i(\tau_2)$ on map $M(\tau_2)$ are the second repeated measurement.

There are a number of ways to measure change in order to analyze the migration patterns of the set of entities $\mathcal{S}_{data}$ from time period τ_1 to τ_2. Magnitude of change for one entity over time can be measured using difference, percent change, and percent difference. These individual values of a set of entities need to be aggregated to obtain a magnitude of change of a set of entities $\mathcal{S}_{data}$. The aggregates of $\mathcal{S}_{data}$ can be calculated by averaging the difference, or averaging the percent change. Comparing average differences between attributes does not make sense when the attributes have different magnitudes or scales. The selection of ways to average percentage change is dependent on the context of the application.

Statistical significance tests can also be used to measure whether changes in an attribute are significant for the chosen subpopulation. The use of these significance tests in this paper is merely for exploratory data analysis, because no treatments were provided to the entities that are mapped to selected nodes in $\mathcal{S}_{node}$, and most datasets are obtained from observational study rather than from experimental one. Observational data are not designed to make inferences of a larger population or to understand cause and effect [9]. Here, a P-value using the paired t-test for each attribute is used to evaluate the significance of changes in the attribute.

To calculate the t statistic for paired samples, one first calculates the difference between each matched pair to produce a single sample, then the one-sample t procedure is used [9]. It is important to keep in mind when analysing the P-value (e.g. Table 1) that the paired t-test calculates the t statistic without reference to the values of the first observations, such as in percentage change, so it can give the same t statistic value even though the differences may have different orders of magnitude for different attributes.

6 Application to the WDI datasets

The migration analysis developed in this paper has been applied to the World Development Indicator (WDI) datasets [16]. The datasets reflect different aspects of welfare of countries in the 1980s and 1990s.

Migration Path Visualization. Migration path visualizations in the WDI datasets between the 1980s and the 1990s is shown in Figure 3(b). With ReDSOM visualization [5], a number of interesting clustering structure changes were identified in these datasets, such as a missing cluster (the light green cluster), a shrinking cluster (the dark red cluster), and a growing cluster (the blue cluster), as shown in Figure 3(a).

The migration patterns involving these clusters are shown in Figure 3(b). It is straightforward from this visualization to see that countries that belong to the light green (lost) cluster, which consists of four South American countries: Brazil, Argentina, Nicaragua, and Peru, migrated to the blue cluster and the orange cluster in the 1990s. Knowing that member countries of the light green cluster experienced an economic crisis in the 1980s, migration of Argentina to the blue cluster that contains OECD and developed countries is an interesting path. This change is likely to be a result of rapid reforms in the late 1980s and early 1990s performed by many Latin American countries [17].

Also from the visualization, a small portion of members of the shrinking cluster (the dark red cluster) migrated to the orange cluster. This shrinking cluster consists mostly of African countries that were characterized by low school enrolment, high mortality rate, and high illiteracy in the 1980s. The orange cluster has moderate values of secondary school enrolment, birthrate, and mortality under 5 years old compared to the dark red cluster in the 1980s. None of the countries from the shrinking cluster in the 1980s migrated to the blue cluster in the 1990s that contains OECD and developed countries. It would be a big step for countries from the dark red cluster to migrate to the blue cluster. The blue cluster received its new members from the orange cluster and the light green cluster.

Analyzing Attribute Interestingness. It can be of interest to know which attributes have a greater degree of change for the cluster changes identified previously. Results of analysis are presented in sorted tables (e.g., Tables 1 and 2). The attributes are sorted based on the P-value (the second column) which is the probability that the old values and the new values in the attribute have no significant difference. The evidence of change is stronger with lower P-value. The commonly used 5% confidence level ($\alpha \leq 0.05$) is used here. The 1980s and 1990s mean values including their normalized values (z-score) are shown in the last two columns.

Based on hot spot analysis [4] of the lost cluster, the distinguishing characteristics were food prices inflation and consumer prices inflation. This analysis has shown that there are some other attributes that have significantly changed for these countries, e.g., an increase in measles immunization, a decrease in children labour, a decrease in mortality under five year old, and a decrease in infant mortality, as shown in Table 1. It is also interesting to note that these improvements in health were achieved without a significant change in the number of physicians (see at the bottom of the Table 1). These improvements can be explained by a UNICEF program known as GOBI (Growth monitoring, Oral rehydration, Breast-feeding, and Immunization) which was targeting poor children [12].

The countries in the shrinking dark red cluster on the 1980s map (Figure 3(a)) demonstrated significant changes in many aspects of welfare (Table 2). Advances in education is indicated by the illiteracy decreasing significantly, while school enrolment in primary, secondary, and tertiary increased significantly. Even though these changes show development, the welfare of these countries is still far behind those in the blue cluster. For example, the mean tertiary school enrolment in the dark red cluster in the 1990s is 4.9% (Table 2) whereas tertiary school enrolment in the blue cluster in the 1990s is 45.1%.

Table 1. Attributes of the countries in the lost green cluster on the 1980s map in Figure 3 sorted by significance of change between 1980s and 1990s in each attribute measured using the paired t-test. The third column indicates direction of change. The line separates those attributes with significantly changed values between 1980s and 1990s with confidence level of 0.05.

Indicator Name	P-value	$\updownarrow$	1980s value (norm)	1990s value (norm)
Immunization Measles	0.0014	↑	66.91 (-0.21)	91.11 (0.98)
Labor Force Children	0.0087	↓	10.89 (-0.28)	6.7 (-0.53)
Inflation Consumer Prices	0.0095	↓	2,995.94 (4.72)	5.81 (-0.19)
Inflation Food Prices	0.0096	↓	2,556.06 (4.54)	4.5 (-0.21)
Mortality Rate Under 5	0.0478	↓	69.73 (-0.27)	49.03 (-0.53)
Mortality Rate Infant	0.0819	↓	48.83 (-0.18)	35.95 (-0.45)
⋮	⋮	⋮		
Daily Newspapers	0.8787	↓	68.47 (-0.27)	63.54 (-0.31)
Physicians	0.9926	↓	1.43 (0.23)	1.43 (0.23)

Table 2. Attributes of the countries in the shrinking dark red cluster on the 1980s map in Figure 3 sorted by significance of change between 1980s and 1990s in each attribute measured using the paired t-test

Indicator Name	P-value	$\updownarrow$	1980s value (norm)	1990s value (norm)
Labor Force Children	0.0000	↓	35.44 (1.19)	27.03 (0.69)
Birthrate	0.0000	↓	44.12 (1.05)	39.65 (0.71)
Television Sets	0.0000	↑	26.23 (-0.82)	55.69 (-0.65)
Illiteracy Rate Adult Female	0.0000	↓	67.36 (1.07)	54.22 (0.61)
Illiteracy Rate Adult Total	0.0000	↓	56.41 (1.08)	44.88 (0.60)
Mortality Rate Under 5	0.0000	↓	188.46 (1.21)	162.46 (0.89)
School Enrollment Secondary Female	0.0000	↑	14.18 (-1.03)	25.59 (-0.70)
School Enrollment Secondary	0.0000	↑	19.24 (-1.03)	29.36 (-0.72)
Mortality Rate Infant	0.0000	↓	114.86 (1.21)	101.81 (0.94)
⋮	⋮	⋮		

7 Conclusion

This paper has presented a method to visualize the migration paths of clusters from two snapshot datasets $D(\tau_1)$ and $D(\tau_2)$. Migration paths between clusters cannot be visualized properly on a single map, either on map $M(\tau_1)$ or on map $M(\tau_2)$, because there might be some regions that are not represented on the other map. Showing migrations using two 2D maps can be complicated as it requires 3D visualization with rotation. The most effective way to visualize the migration path is by representing clusters of 2D SOMs using 1D maps. Furthermore, experiments showed that transforming a 2D SOM into a 1D SOM can create folding in the 1D SOM cluttering visualization.

The paper has also presented a framework to understand which attributes have changed significantly in a set of selected entities from two snapshot datasets. This selection can be made though interactive brushing on the SOM visualizations, such as through ReDSOM visualization [5] or visualization of a clustering result of the map. Alternatively, entities can be selected based on their involvement in a user-selected migration path based on the migration visualization.

A number of measures can be used to assess whether or not these changes are likely to be significant, such as using differences, percent change, or paired t-test. The paired t-test evaluates the significance of changes, however it does not measure the magnitude of changes, the test only evaluates differences in an attribute relative to the variability of individual differences. This framework to analyze attribute interestingness can be applied using other measures as well.

Real-life datasets from the WDI were analyzed using the framework devised in the course of this research revealing interesting changes. As an example, the countries in the lost cluster who experienced high inflations in the 1980s achieved significant improvements in health in the 1990s without a significant change in number of physicians which can be explained by a UNICEF program known as GOBI. This hidden context was found after analysis using the method developed in this paper.

Evaluations of the visualizations and methodologies developed in this paper using real-life datasets demonstrated that they can help analysts discover hidden context by providing detailed insights into reasons for cluster changes between snapshot datasets.

7.1 Future Work

Since SOMs can handle only numerical data, the methodology and visualization methods developed in this paper cannot be applied to categorical data without scale conversion. Further research should consider other clustering approaches that can work with categorical data and also mixed types of data.

Because this research provides a framework to analyze migration, new measures can be developed or evaluated to analyze attribute interestingness in the framework. The paired t-test used in this paper does not take into account changes in the rest of the population. If a selected group of entities has similar changes to the rest of the population for a particular attribute, the attribute does not add value for an analyst in understanding the changes. In other words, changes that occur locally need to be normalized by changes that occur globally. Further work needs to be done to evaluate other methods, such as the two-sample t-test, to evaluate whether the magnitude of change in the selected group is similar to the magnitude of change in the rest of the population.

References

1. Burez, J., den Poel, D.V.: CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. Expert Systems with Applications 32(2), 277–288 (2007)
2. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 1(2), 224–227 (1979)
3. Denny, Squire, D.M.: Visualization of Cluster Changes by Comparing Self-Organizing Maps. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS (LNAI), vol. 3518, pp. 410–419. Springer, Heidelberg (2005)

4. Denny, Williams, G.J., Christen, P.: Exploratory Hot Spot Profile Analysis using Interactive Visual Drill-down Self-Organizing Maps. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 536–543. Springer, Heidelberg (2008)
5. Denny, Williams, G.J., Christen, P.: Visualizing Temporal Cluster Changes using Relative Density Self-Organizing Maps. Knowledge and Information Systems 25, 281–302 (2010)
6. Diggle, P.J., Liang, K.Y., Zeger, S.L.: Analysis of Longitudinal Data. Oxford University Press, New York (1994)
7. Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer Series in Information Sciences, vol. 30. Springer, Heidelberg (2001)
8. Lingras, P., Hogo, M., Snorek, M.: Temporal cluster migration matrices for web usage mining. In: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 441–444. IEEE Computer Society (2004)
9. Moore, D.S.: The basic practice of Statistics. W H. Freemand and Company, New York (2000)
10. Reinartz, W.J., Kumar, V.: The impact of customer relationship characteristics on profitable lifetime duration. The Journal of Marketing 67(1), 77–99 (2003)
11. Roddick, J.F., Spiliopoulou, M.: A survey of temporal knowledge discovery paradigms and methods. IEEE Transactions on Knowledge and Data Engineering 14(4), 750–767 (2002)
12. Seear, M.: An Introduction to International Health. Canadian Scholars' Press Inc., Toronto (2007)
13. Serrano-Cinca, C.: Let financial data speak for themselves. In: Deboeck, G., Kohonen, T. (eds.) Visual Explorations in Finance with Self-Organizing Maps, pp. 3–23. Springer, London (1998)
14. Vesanto, J.: SOM-based data visualization methods. Intelligent Data Analysis 3(2), 111–126 (1999)
15. Vesanto, J., Alhoniemi, E.: Clustering of the Self-Organizing Map. IEEE Transactions on Neural Networks 11(3), 586–600 (2000)
16. World Bank: World Development Indicators 2003. The World Bank, Washington DC (2003)
17. Zagha, R., Nankani, G.T. (eds.): Economic Growth in the 1990s: Learning from a Decade of Reform. World Bank Publications, Washington, DC (2005)