Visual Analytics Methods for Categoric Spatio-Temporal Data

T. von Landesberger* TU Darmstadt Darmstadt, Germany
Sebastian Bremm† TU Darmstadt Darmstadt, Germany
Natalia Andrienko‡ Fraunhofer IAIS Bonn, Germany
Gennady Andrienko§ Fraunhofer IAIS Bonn, Germany
Mária Tekušová¶ SHMU Bratislava, Slovakia

Abstract

We focus on visual analysis of space- and time-referenced categorical data, which describe possible states of spatial (geographical) objects or locations and their changes over time. The analysis of these data is difficult as there are only limited possibilities to analyze the three aspects (location, time and category) simultaneously.

We present a new approach which interactively combines (a) visualization of categorical changes over time; (b) various spatial data displays; (c) computational techniques for task-oriented selection of time steps. They provide an expressive visualization with regard to either the overall evolution over time or unusual changes.

We apply our approach on two use cases demonstrating its usefulness for a wide variety of tasks. We analyze data from movement tracking and meteorologic areas. Using our approach, expected events could be detected and new insights were gained.

1 Introduction

Spatio-temporal data consist of three main components: geographical space, time, and thematic attributes describing various properties of places and spatial objects. These components are very different in kind, which makes the data complex and difficult to visualize and analyze. Visual displays suitable for representing one of the components give very limited possibilities for representing the other components. Particularly, maps, which are the primary means to represent space, are very weak in representing time and changes occurring over time. The effectiveness of map animation raises serious doubts [34]. Space-time cube, where two dimensions represent space and the third dimension time [18,21,26], also has its drawbacks (occlusions and distortions of both space and time due to projection) and limitations with respect to the number of spatio-temporal objects that can be effectively viewed and the length of the time interval. Therefore, researchers dealing with spatio-temporal data often combine cartographic representations of the spatial component of the data with other types of display showing the temporal and thematic components [5]. While changes of numeric characteristics are explored by combining maps with such displays as temporal line plots (time graphs), exploration of categorical changes has not been sufficiently addressed yet.

The term "categorical data" refers to thematic attributes assuming nominal or ordinal values, which denote possible states of objects [6]. For example, states of a person may be 'at home', 'at work', etc. The categoric data can either stem from direct data observations (e.g., health states of a person) or be derived from other data (e.g., by clustering). The states can change over time, i.e., an object can move from one category to another. This kind of change is in the following referred to as "categorical change" or "transition". We call data reflecting such changes over a time *time-varying categorical data* or *dynamic categorical data*.

We have designed Dynamic Categorical Data View (DCDV), a special type of display showing categorical changes, which is applicable to large numbers of objects and longer time periods. Being dynamically linked with a cartographic map display and space-time cube, DCDV enables visual exploration of categorical changes occurring in space and time. We support generic tasks [30]:

- *Overview*: see the overall frequency of the changes and their character (gradual, when only a few objects change each time, or abrupt, when many objects change simultaneously; coherent, when many objects change in the same way, or rather chaotic), periods of stability and periods of intensive changes.
- *Zoom and filter*: consider a subset of objects, selected time steps, or particular transitions between categories.
- *Details-on-demand*: using the dynamic links to maps and other displays through coordinated highlighting and filtering, find answers to queries related to the spatial (where?), temporal (when?), and thematic (what?) data aspects [28]):
  - Where in space is category c at time t? Where are the objects that changed from category c_1 to category c_2 between times t_1 and t_2?
  - When did objects from category c appear at location l? When did spatial object (subset of objects) o belong to category c?
  - What is the category of the spatial object (subset of objects) o at time t? What changes did the object(s) o undergo from time t_1 to time t_2?

Our **contributions** are as follows:

- An approach to analyzing *spatially referenced dynamic categorical data*. It interactively combines geographic and categoric changes visualization and provides automatic data analysis support.
- Extended and novel algorithms for identification of *globally and focally representative time steps*.
  - A selection of *globally* representative time steps enables an overview of the mainstream data developments.
  - A choice of *focally* representative time steps is suitable for revealing unusual developments and for focusing on specific aspects of the data (hence the name).
- Visual-interactive user *guidance for setting of algorithm parameters*, when choosing a set of time steps for deeper examination.

We would like to emphasize that our algorithms do not select representative time steps individually, e.g. based on the statistical distribution of the categorical values at each time step or on the differences from the previous time step, but select combinations of time steps based on the similarities and differences of each time step with respect to all others. Different combinations are supportive for different analysis tasks; therefore, several algorithms are needed for finding good combinations according to a task at hand. We would also like to note that finding good combinations is non-trivial: even

*e-mail: tatiana.von.landesberger@gris.tu-darmstadt.de
†e-mail: sebastian.bremm@gris.tu-darmstadt.de
‡e-mail: natalia.andrienko@iais.fraunhofer.de
§e-mail: gennady.andrienko@iais.fraunhofer.de
¶e-mail: maria.tekusova@shmu.sk

IEEE Symposium on Visual Analytics Science and Technology 2012
October 14 - 19, Seattle, WA, USA
978-1-4673-4753-2/12/$31.00 ©2012 IEEE

for data with a relatively small number of time steps the number of possible combinations is far beyond the capacity of the user to test all of them and find useful ones. It should be also borne in mind that selecting all time steps, even for short time series, is usually not a good idea since interesting global patterns may be hidden among the numerous changes between neighboring time steps.

We have applied our approach to two use cases thereby demonstrating its usefulness for diverse user tasks. The first use case shows analysis of daily movements of many mobile phone users in a city. The second case focuses on analysis of weather patterns at many stations in two European countries over a long time period.

2 Related Work

Our work relates to the areas of spatio-temporal visual data analysis, visualization of (time-varying) categorical data and analysis of time series. As dynamic categorical data often stem from clustering of other time-varying data, we give an overview of the work in this domain as well.

Visualization of categorical data: The Parallel Sets approach by Kosara et al. [20] displays multivariate categorical data. In particular, it shows group memberships across several categorizations building upon parallel coordinates. Similarly to Parallel Sets, Interactive Sankey diagrams [10] show categorical data over several dimensions while improving the visual design. Mosaic plots [15] and KVMap [25] use space filling approach for multi-dimensional categoric data. Mosaic plots divide recursively the rectangular space according to number of objects in each category. KVMaps use regular splitting, where color coding shows the number of objects in each category.

Categoric data in geographic context is often shown on maps (e.g., by color coding of geographic areas according to category value) or using special visualization techniques. For example, for hierarchically organized categoric data, spatially-ordered treemap technique has been introduced [40]. Recently, Wood et al. [39] presented BallotMaps – an abstract visualization of spatial ordered data (voting rankings for voting regions). It shows the relative preference of a candidate for regions and for parties (categories) according to name order. All these approaches however do not deal with categories changing over time.

A special case of categorical data are **clustering results**. Zhou et al. [41] and Lex et al. [24] propose approaches for *clustering result comparison*. They both consider several groupings of data objects according to different clustering results. The visualization is similar to Parallel Sets [20], however it focuses on individual objects. The view connects each object across all clusterings while minimizing edge crossings. These approaches do not regard time dependency of the data and put strong emphasis on individuals rather then groups.

Visualization of time-varying data: A recent book on visualization of time series [2], building upon the survey presented by Aigner et al. [1], provides a broad overview of techniques for visual analysis of time series. Most relevant to our work is the approach of Hao et al. [13], which proposes an importance-driven time series visualization: more important time intervals are provided with more screen space. Data which are more up-to-date or have higher variance are deemed more important. Moreover, Ziegler et al. [43] visualize time series based on the changes between pairs of time steps and in this way, important value changes can be identified.

Visualization of time-varying categoric data: ThemeRiver [14] approach shows time changes of topic categories over times in a stacked chart where bar sizes denote the number of objects in each group (i.e., number of articles per topic). The developments in the number of objects in each category can be examined. A similar approach is the History Flow [36], which displays the development of documents by authors. The focus is on identification of important themes/authorships in each time step. Both approaches, however, do not include in the view the changes of group membership between time steps.

Sequence data also can be seen as time-varying categorical data without an explicit time dimension. There are several approaches for visual analysis of such data. They disregard the exact time and concentrate only on states (sequence values) and their changes. Often they are represented as state-change or *state transition graphs*, where each node is a state and an edge represents a state change. Edge thickness and node size represent the number of objects [7, 35, 37]. The main disadvantage is that the exact time of transitions is not visible in the graph. It can only be seen in another linked view. Moreover, these approaches do not take spatial information into account.

Recently, two new approaches for analyzing **group changes over time** have been presented [8, 33]. The first work by Turkay et al. [33] deals with clustering results of time-varying data. The visual analysis shows to which cluster an individual object belongs over time (similar to [24]) and focuses on assessment of cluster quality development. Secondly, the work by Bremm et al. [8], extends Parallel Sets approaches [20] with time dimension and algorithmic data analysis for examination of group changes over time. Both approaches, however, disregard geographic location of the data, so spatio-temporal analysis is not possible. Moreover, the first work does not provide time selection and the second work offers only a simple algorithmic analysis of time steps with a black-box approach to time step choice.

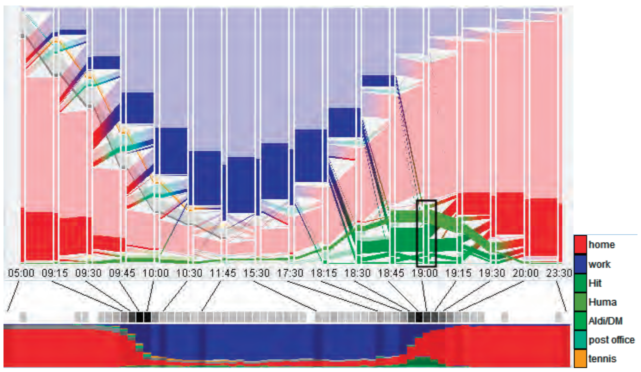
Spatio-temporal visual data analysis: The paper by Guo et al. [11] demonstrates two approaches to exploration of categorical changes that are applied in geovisualization. In this work, categories of spatial objects (states of the USA) are defined by clustering of combinations of values of multiple time-varying attributes. The authors suggested two complementary ways to visualize the cluster membership of the objects over time. The first is the classical “small multiples” approach [32] where a small map for each time step in which the cluster membership of the objects is represented by color. The second is a matrix with the rows corresponding to objects and columns to time steps; the cells are colored according to the cluster membership. The ”small multiples” are good for exploring the spatial distribution of the categories at each time step and the differences between the distributions at different time steps. The matrix is suitable for exploring the evolution of each individual object. However, both displays do not support well enough the exploration of object groupings and detection of coherent changes of multiple objects. Besides, it is cognitively difficult to examine a large number of maps showing data at different time steps. The displays also do not provide a convenient overview of the overall evolution.

Automatic identification of significant time steps in time series data: The above-mentioned approaches to visualization of categorical data cannot be straightforwardly applied to long time series of categorical values. There may be not enough screen space to represent all time steps and changes between them. Hence, there is a need for scalability in the approaches, e.g., representing all time steps in a summarized form and selected time steps in more detail. Good selection of time steps plays a crucial role.

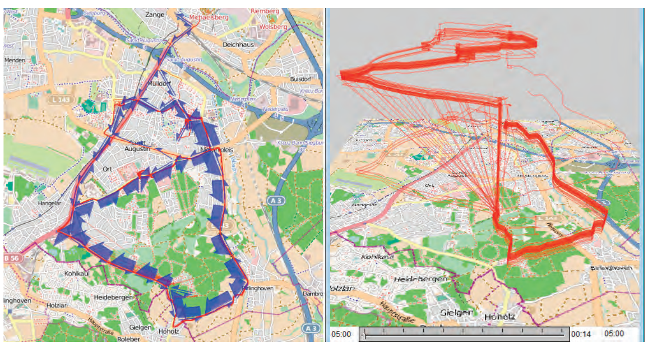
There are numerous visual and computational methods for identification of important time steps in time series [12, 23, 42, 43], videos [27, 31] or other types of data. They try to compress the underlying data set into a small set of key time steps that would best represent the data set. There are various methods, such as largest difference between time steps, regular spaced time steps with succeeding removing of less important time steps. They mostly work with quantitative data, so they need to be extended or adapted to categorical data.

3 APPROACH

In our approach, we interactively *combine a geographic view on the data locations with the dynamic categorical data view (DCDV)* allowing visual analysis of categories and their changes (see Figure 1). The geographic view applies geovisualization techniques appropriate to the data type [4], the categorical view builds upon the ideas presented in [8]. This view is briefly introduced in Section 3.1. We link the two views: objects and time steps identified in one view are propagated to the other view allowing for simultaneous analysis of both spatial data organization and of categorical changes (see Figure 1).



(a) Dynamic categorical data view with highlighting (DCDV)



(b) Linked geographic views

Figure 1: The presented visualization of spatio-temporal categoric data links the dynamic categorical data view (a) and various geographic views (b). This figure shows a summary of about 1 year of daily trajectories of a person. Different colors depict different locations and associated activities (see legend on right). The categoric view (a) allows for tracking of group memberships for highlighted objects (i.e., trajectories in this case) over time. In this case, all objects in the green group at 19:00 were selected (see black box on the top). The selected objects are shown in the linked geographic views (b).

3.1 Dynamic Categorical Data View (DCDV)

DCDV shows category sizes (i.e., number of objects having a category value) and counts of transitions (i.e., objects changing categories) over time (see Figure 2). It contains three parts.

The main view (1) shows details for category sizes (bars) and group membership changes (edges) for selected time steps. Bar height encodes the number of objects in a category in a time moment. Bar color denotes the category. Edge width reflects the number of objects participating in a certain transition between two selected time moments. Edge color corresponds to the categories - color smoothly changes from start category to end category. Exact counts for visual elements can be accessed on demand. Note that empty areas between bars and edges having the shape of triangles do not convey any information.

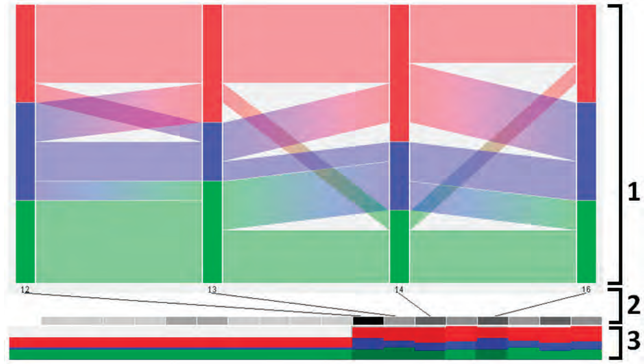


Figure 2: The dynamic categorical data view (DCDV). The upper part (1) shows the number of objects in a category (staced bars) and number of objects per categorical change (edges) in the selected time steps. Middle part (2) shows the distribution of selected time steps over the analyzed period as lines connecting corresponding time points in views 1 and 3. Grayscale heatmap indicates the intensity of the transition activity for all time moments. Lower part (3) shows object distribution among the categories in each time step over the whole time period.

The middle part (2) indicates the distribution of time steps selected for detailed inspection in the main view (1). It also shows the counts of transitions between categories using a gray color scale. In this way, the user can detect high and low activity periods in the data.

On the bottom (3), the sizes of color-coded categories in each time step over the whole time period are shown giving a general overview of the object distribution over categories and time.

The data visualization relies on selection of time steps for detailed inspection. This selection can be done both interactively and algorithmically. The users can click on steps of interest in the main view or transitions in an additional t-t plot [16] showing transition counts between each pair of time steps as a triangular heatmap matrix. The algorithmic analysis can select globally and focally representative steps depending on user's analysis goals (see Section 4).

DCDV provides *interactive features* for deeper analysis of the data. It allows the user to track a set of interesting objects over time (see Figure 1). These objects can be chosen in various ways: according to their location, movement pattern, group membership, or group changes. The selected objects are highlighted in both geo- and categoric views. DCDV is linked to further geographic data views and algorithmic data analysis functions (see Section 5).

4 TRANSITION-BASED TIME SELECTION

The analysis of categorical changes (i.e., transitions), as proposed in this paper, relies on selection of representative time steps. Regarding the limited screen space and cognitive human skills, we propose to present to the user only such combinations of time steps that uncover important data properties. We have therefore developed several algorithms for user-guided automatic time selection. These algorithms can be applied to the whole data set (all objects in all time steps) or to various subsets (e.g., selected objects, time intervals, or categories). The time selection is a guided process, where the user is provided with visual feedback on the sensitivity of the results to the algorithm parameters. The results can be further adjusted in an interactive way, where the user chooses the time steps to additionally include in or exclude from the suggestion.

In the following, we present the new algorithms and the visual-interactive interface for time selection.

4.1 Transition Weights (Distances)

For task-specific time selection, the user may employ transition weights w that influence the selection of time steps. Larger weights are assigned to more important transitions.

The weights often represent distances (dissimilarities) between categories. These distances can be determined by

- geographic location of regions $L(c)$ (e.g., activity locations),
- data values (e.g., cluster distances as Euclidean distance of cluster centers)
- context-dependent differences between categories (e.g., shopping and work are more distant than shopping and sport).

Setting some weights to zero serves as *filtering*, so that irrelevant transitions are not taken into consideration.

The choice of right weight calculation method is left to user expertise. Especially, if clustering is used, the choice of appropriate clustering method and its parameters are not tackled in this paper as we concentrate on time selection.

4.2 Definitions

We consider the type of data containing objects having certain categorical attribute values and certain locations in different time steps. Objects are denoted as $o_i \in O$, $i = 1, \dots, n$, where n is the number of objects in the data set. The time steps $t \in T$, $t = 1, \dots, \tau$, are assumed to be equally spaced, or their spacing is irrelevant to the analysis. We allow that an object o_i in time step t_j may not exist, or the data about this object may be missing. The number of available objects at time step t_j is denoted as n_{t_j}. The number of objects existing in at least one of the time steps t_i and t_j is denoted $n_{t_i \vee t_j}$. Note that, in the data analysis, the user may focus only on a subset of time steps T^S, $T^S \subset T$ or on a subset of objects O^S, $O^S \subset O$.

We define that object o_i in time step t_j has a categorical value c_k (i.e., is a member of group c_k, is classified as c_k or has class c_k) as $C(o_i, t_j) = c_k$, $c_k \in C$, where $C = \bigcup_{k=1}^{m} c_k$ is the set of all categories. m is the number of distinct categories (i.e., the number of categorical values/states or classes). Moreover, we define a categorical value c_0 (missing), where missing or non-existing objects in specific time steps belong $C(o_i, t_j) = c_0$. In the algorithms, we also include the category c_0 in the calculations. Therefore, we extend C as $C^0 = C \cup \{c_0\}$. For simplicity, we refer to C^0 as C, i.e., $C^0 \equiv C$.

We define that object o_i is member of a categorical change (i.e., transition) $Tr^{t_1,t_2}_{c_k,c_l}$, when $C(o_i, t_1) = c_k \wedge C(o_i, t_2) = c_l$. This object transition is denoted as $Tr^{t_1,t_2}_{c_k,c_l}(o_i)$. Please note that t_1 and t_2 are not necessary consecutive time steps, i.e., there may be any number of other time steps between them.

We define the number of objects participating in a certain transition $Tr^{t_1,t_2}_{c_k,c_l}$ as $N(Tr^{t_1,t_2}_{c_k,c_l})$. The number of all objects that change categories between time steps t_1 and t_2 are defined in Equation 1.

$$N(Tr^{t_1,t_2}) = \sum_{c_k \in C, c_l \in c, c_k \neq c_l} (N(Tr^{t_1,t_2}_{c_k,c_l})) \tag{1}$$

We denote the weight for the transition Tr_{c_k,c_l} as w_{c_k,c_l}. We assume that all weights are non-negative $w_{c_k,c_l} \geq 0$, and $\exists k,l : w_{c_k,c_l} > 0$. We define $mw = \max_{\forall c_k \neq c_l}(w_{c_k,c_l})$.

The weighted number of transitions is calculated as in Eq. 2.

$$WN(Tr^{t_1,t_2}) = \sum_{c_k \in C, c_l \in C, c_k \neq c_l} (w_{c_k,c_l} \cdot N(Tr^{t_1,t_2}_{c_k,c_l})) \tag{2}$$

4.3 Algorithms for Time Selection

The algorithms select combinations of time steps that adequately represent the data developments for the task at hand. In this respect, we distinguish two types of the *analytical intent* of the user:

1. Get an overview of the mainstream data developments and overall trends. A selection of time steps supporting this goal is called *globally representative*. Globally representative selections can be made according to following criteria:
   (a) *Activity-based*: reveal the time periods of high and low transition activity.
   (b) *Similarity-based*: reveal similarities between time steps and repetitive patterns.
2. Get information on specific facets of the dataset and reveal uncommon development patterns. A selection of time steps supporting this goal is called *focally representative*.

1. Activity-based globally representative time selection: To provide a global overview of the data, our algorithm aims at giving low prominence to low levels of transition activity and high prominence to high level of activity. The algorithm selects a subset of time steps $T^S \subset T$ so that the relative number of transitions between any two consecutive time steps within this subset $t^S_i \in T^S$ and $t^S_j \in T^S$ is approximately equal. The number of transitions is normalized, which accounts for the changes in the object sets between time steps. The same number of transitions in a small data set (small number of objects N) may be more relevant than in a large data set.

We introduce *two variants of the algorithm* that are based on the number of objects changing groups: one taking into account the total number of transitions between two time steps and the other looking at category-specific transitions (see Algorithm 1a and b). Variant (a) relies on the total number of transitions between selected time steps. It it suitable for a general overview of the transition activity, however, it does not distinguish whether these transitions happened among few categories or were more widely distributed across categories. The category-specific number of transitions is captured in variant (b), which supposes that transitions between a pair of time steps $\{t_i, t_j\}$ are relevant if there is a significant difference between the number of transitions for all pairs of different categories (WDN). The calculation is presented in Equation 3.

$$WDN(Tr^{t_1,t_2,t_3,t_4}) = \frac{\sum_{c_k,c_l \in C} abs(WN(Tr^{t_1,t_2}_{c_k,c_l}) / - WN(Tr^{t_3,t_4}_{c_k,c_l}))}{(mw \cdot n_{t_1 \vee t_2 \vee t_3 \vee t_4})} \tag{3}$$

Note that the second variant can be used also for further selection of relevant time steps from a subset of consecutive time steps selected by the first variant (see Section 4.4).

Algorithm 1 ACTIVITY-BASED GLOBAL TIME SELECTION

$T^S = \{t_1\}$
$t_{ref} = t_1$
for all $t_i \in T, i > 1$ **do**
  `a) condition on total count`
  if $WN(Tr^{t_{ref},t_i}) / (mw \cdot n_{t_{ref} \vee t_i}) \geq d$ **then**
  `b) condition on category-based count`
  if $WDN(Tr^{t_{ref},t_{ref+1},t_i,t_{i+1}}) \geq d$ **and** $i < \tau$ **then**
    $T^S \leftarrow T^S \cup \{t_i\}$
    $t_{ref} \leftarrow t_i$
  end if{ for both conditions}
end for
$T^S \leftarrow T^S \cup \{t_\tau\}$
return T^S

Algorithm results depend on the setting of the parameter d. We provide the user with information on the time selection given a threshold d for an informed decision (see Section 4.4).

2. Similarity-based Global Selection of Time Steps This algorithm is meant to reveal repetitive patterns in object distribution across the categories, particularly, in periodic data. Data with periodic variation require specific ways of selecting representative time steps. When analyzing data for multiple periods (e.g. weather over many years), the analyst is not so much interested in changes between consecutive time steps and in differences between time steps within one period as in similarities and differences between the periods and long-time trends over many periods. For dealing with periodic data and for revealing repetitions in arbitrary data, we propose similarity-based time selection, which employs clustering of time steps by similarity of the corresponding data. The result is a set of time clusters, each containing a subset of time steps $T^K = \{T^{K_1}, ^{K_k}\}, T^{K_i} = \bigcup t_j^{k_i}$. Each time cluster gives a selection of time steps to be interactively explored. The user needs to explore all time clusters to construct a full picture of the data developments. For the time step clustering, the distance (amount of dissimilarity) between time steps can be measured as dissimilarity of the object classes taking into account their weights (see 4.1). The choice of clustering method and its parameters is user-defined. We provide several clustering algorithms (e.g., K-Means , SOM, DBScan) [38] for this purpose.

Algorithm 2 SIMILARITY-BASED GLOBAL TIME SELECTION

```
Determine distances of time steps
```
for all $t_i \in T$ **do**
  for all $t_j \in T, i \leq j$ **do**
    $dist(t_i, t_j) = \sum_{c_k \in C, c_l \in C} (w_{c_k,c_l} \cdot N(Tr_{c_k,c_l}^{t_i,t_j}))$
  end for
end for
$T^K \leftarrow$ `cluster(T)`
return T^K

3. Activity-based Focal Selection of Time Steps: This algorithm is used when the user wants to focus on particular data developments. There are two variants, which search for pairs of time steps with either very high or very low transition activity (the latter case means irregular transitions, i.e., outliers). This algorithm selects a set of time pairs formed by consecutive time steps $T^S = \bigcup \{t_i, t_{i+1}\}$, where the weighted number of transitions is higher/lower then a threshold d (see Algorithm 3a and b). The threshold d is user-defined using a visual-interactive interface (see Section 4.4). Please note that in case of searching for outliers in transitions, the normalization of transition count is not used (e.g., when only one object changes categories c_k, c_l). We also do not use weights for finding unusual transitions, apart from filtering out irrelevant transitions (i.e., with zero weight).

Algorithm 3 ACTIVITY-BASED FOCAL TIME SELECTION

$T^S = \emptyset$
for all $t_i \in T, i < \tau$ **do**
  `a) condition for high activity`
  if $WN(Tr^{t_i,t_{i+1}})/(mw \cdot n_{t_i \vee t_{i+1}}) \geq d$ **then**
  `b) condition for low activity`
  if $0 < N(Tr_{c_k,c_l}^{t_i,t_{i+1}}) \leq d$ **and** $w_{c_k,c_l} > 0$ **then**
    $T^S \leftarrow T^S \cup \{t_i, t_{i+1}\}$
  end if{ for both conditions}
end for
return T^S

4.4 Visual-Interactive Threshold Selection

The setting of proper parameters for algorithmic data analysis is usually essential for gaining high quality results. A good practice is to inform the user about the effect of the possible parameter choices on the result. In our tool, the selection of time steps in Algorithms 1 and 3 depends on the parameter d. If d is too small, all time steps can be selected. If d is too large, none or only the first and last time steps are selected. We provide the user with information on the distribution of the selected time steps depending on threshold d for an informed decision.

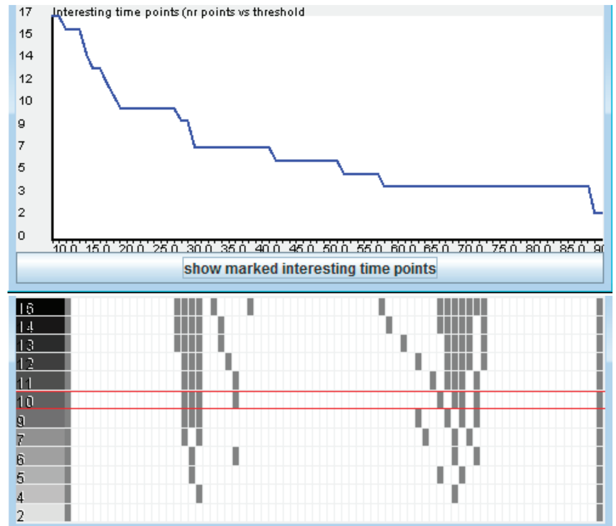


Figure 3: Choosing a threshold for time selection algorithms. Top: The impact of the threshold on the number of selected time steps. X-axis: threshold value, y-axis: number of selected time steps. Bottom: The impact of the threshold on the temporal distribution of the selected time steps. It shows a table, where rows correspond to threshold results for a selected number of time steps. Columns denote all time steps, where the selected time steps are shown in grey. The setting highlighted in red is chosen for visualization.

For the threshold selection view, we determine a set of threshold values $d \in D$, for which we calculate the selected time sets $T^S(d)$. We use thresholds $d = d_0 + i * d_s$, $i = 0, \ldots, i_{max}$, where d_0 is an initial threshold, d_s is the step between threshold values and i_{max} is the stopping criterion. The values of d_0 and d_s depend on the applied algorithm. For the algorithms $1a, 1b$ and $3a$, where $0 \leq d \leq 1$, we take small values of the initial threshold and the step: $d_0 = 0.05$ and $d_s = 0.05$. The number of time steps in algorithms $1a, 1b$ and $3a$ is generally non-increasing with growing d. Note that the Algorithm $1a$ selects time steps in a successive manner with a fixed addition of the last step. There can be special cases where a time step is added close to the end leading to an increase in the selected time steps by one when using larger d.

In the algorithm $3b$, the number of selected time steps decreases with increasing d, as we look for atypical events in the data $(... \leq d)$. So we start with a high d_0 and decrease it constantly. $d_0 = (int)0.05 * n$ and $d_s = -1$.The choice of d_0 is motivated by the general rule of thumb that $\leq 5\%$ threshold is commonly assumed as atypical [22]. The i_{max} determining the stopping of i increase is determined by the data – when there are no relevant times selected, because the threshold is too high (too low in case of Algo. $3b$).

We provide the user with a view on the *relationship between the threshold and the number of time steps* selected by the algorithm (see Figure 3 top). This view was inspired by the work on selection of dimension reduction threshold [17]. The parameter setting is difficult, as a small change of the threshold may result in a large decline of the number of time steps and vice versa. By looking at the line chart, the user may choose a relevant number of times by, e.g., "elbow criterion" [19], which is often used for determining the number of clusters as the point where adding more clusters does not add much information. This view, however, does not show a deeper insight into the distribution of time steps across the data set. This is tackled in a new view described below.

The new visualization shows *details on the impact of threshold setting on the time selection* (see Figure 3 bottom). It provides information both on the number of selected time steps for threshold d_i and on their positions. The number of selected time steps is shown in a gray color scale on the left together with the value of the threshold. The selected time steps are shown on the right as gray-shaded rectangles. Please note that we show only those thresholds that give different results in comparison to the previous threshold values. The motivation for this design choice can be seen in Figure 3 top. Many changes of threshold do not lead to changes in time selection. Therefore, we compact this view and show only different time selections. Moreover, this filtering also compensates for non-optimal selections of the initial threshold d_0 and the step d_s.

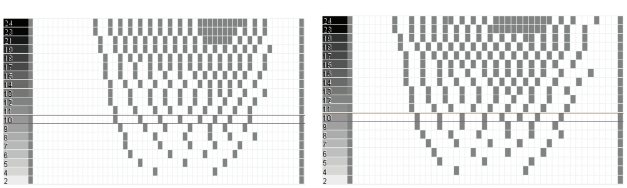
Calculation complexity in algorithms 1 and 3 is determined by the number of objects, number of time steps and thresholds used for final calculation ($O(N \cdot T \cdot D)$, where D is the number of thresholds used for selection proposals). Note that, assessment of the calculation complexity for calculating clusters in Algorithm 2 goes beyond scope of the paper. We therefore refer to [38] for complexity of cluster calculation, which is the main part of the algorithm.

5 Application Examples

We demonstrate the usefulness of our approach on two datasets from different application areas and of different structure. Section 5.1 shows visual analysis of movement data and Section 5.2 demonstrates our approach on meteorologic data.

(a) Milan regions (b) Trajectories on Wednesday (c) Trajectories on Sunday

Figure 4: Use case data. Milan has been divided into geographic regions forming categories (a). Mobile phone caller trajectories on Wednesday (b) and Sunday (c).

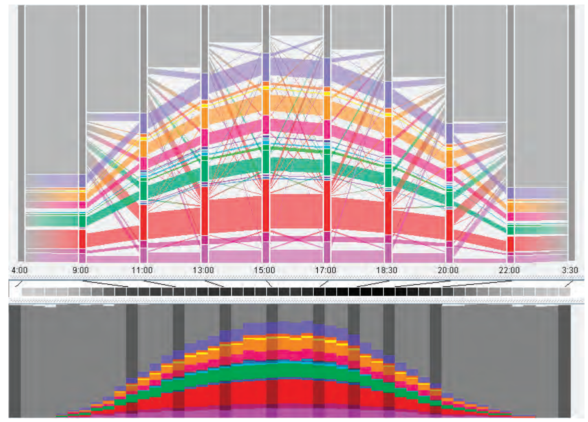


(a) Wednesday (b) Sunday

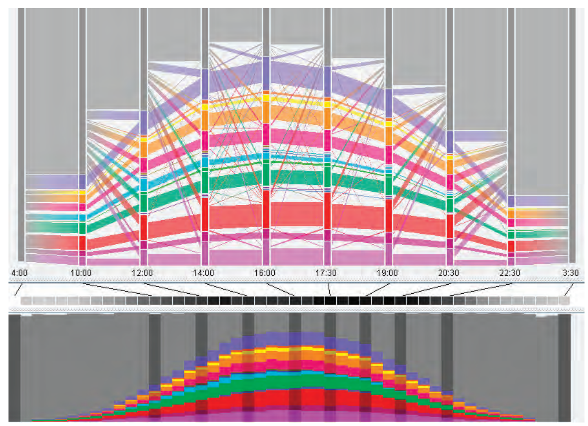
Figure 5: The overviews of the selected time moments w.r.t. the choice of threshold. It shows that the majority of phone call activity start later on Sunday (b) than on Wednesday (a).

5.1 Unusual Movement of Multiple People

We analyze a dataset with 5,108,298 mobile phone call records from 367,730 customers of an Italian mobile phone company WIND collected over a period of 9 days in Milan area. From these data, we have constructed caller trajectories and divided them into daily tracks. The territory of Milan has been split into 307 Voronoi polygons built around the positions of the network antennas. Since these polygons are not meaningful in this case, we group them into larger regions reflecting the geography of the city (see Figure 4).



(a) Wednesday



(b) Sunday

Figure 6: Daily movements for Wednesday (a) and Sunday (b). The figure allows to compare the calling and movement behavior of people between a weekday and at weekend. It shows that the majority of phone call activity on Sunday start later then on Wednesday. Although the general movement patterns show high similarities, the proportion of people in regions varies, especially in the center of Milan (red) and in Milano 2 area (cyan).

To explore and compare the mobility of the phone users on a typical working day and at the weekend, we select from the database two subsets of trajectories: from Wednesday and from Sunday. We also remove the trajectories of stationary customers who made all their calls in the same cell or a few neighboring cells. This gives us 6,943 trajectories of mobile customers on Wednesday and 4,496 on Sunday. It signifies lower calling activity at the weekend. From these trajectories, we generate time series of visited regions with a time step of 30 minutes and visualize them on two DCDV displays.

With the help of tools for interactive threshold selection, we select 10 globally representative time steps for Wednesday and Sunday (see Figure 5). The selected time points indicate a later start of activity on Sunday than on Wednesday.

The behavior of callers between the selected time steps is shown in DCDV view (see Figure 6). The colors correspond to the city division. Grey color means that the locations of the callers are unknown; this occurs before the first call and after the last call. Main activity starts later on Sunday (by 9:30) then on Wednesday (by 8:30), based on globally relevant time point selection. Movement pattern between areas is very similar on both days, although time spans of these activities differ. The most visible flows are from "unknown" and to "unknown", while the flows between regions are much smaller. The summary view on the bottom shows: On Sunday compared to Wednesday, the presence of people by regions is proportionally lower in the Center (red) and Linate (green) and proportionally higher in e.g. San Siro (purplish red) and Milano 2 (cyan). The presence in Milano 2 is notably higher in the first half of the day.

A detailed inspection using the DCDV view (see Figure 6b) revealed, that there are much more activities on Sunday than on Wednesday in Milano 2 (cyan). Hereupon, we have used interactive selection of trajectories by visited areas and thus created a new instance of DCDV with the selected subset. On Wednesday, the presence of people in this region was rather constant, and there were no unusual change patterns (no focally representative time steps were detected). On Sunday, something particular was going on in Milano 2 in the first half of the day. Using the link to the map display, we select the trajectories of the people who were in Milano 2 at the selected representative time moments presented in Figure 7 top. Figure 7 left shows these trajectories. All or almost all trajectories meet in a single point. Hence, we can identify more precisely the place where unusual activities occurred on Sunday. Figure 7 right shows this place and its surrounding with more geographical detail. The place is near the metro station Cascina Gobba. By searching on the Internet for the possible reason of the unusual activities, we found several videos in YouTube showing the flea market at Cascina Gobba, which explains our finding.

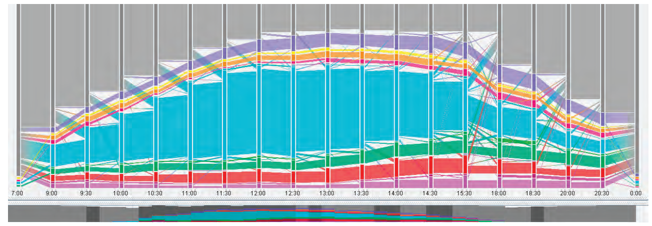
5.2 Analysis of Clustered Weather Data

This example demonstrates the use of DCDV for analyzing meteorological data. The goal of the analysis is to group similar weather characteristics and to find weather extrema. The assessment described in this use case was done together with an expert meteorologist – a coauthor of the paper.

We use a dataset with monthly weather-related data from 111 weather stations over Germany and Slovakia (see Figure 8) for the period of 241 months from January 1991 till January 2011. The complexities of this dataset are multiple attributes describing the weather (various attributes for temperature, wind force, cloudiness and precipitation), long time series, and periodic variation (seasonal variations over the year).

We analyze the data with help of clustering that groups similar value combinations of the weather attributes. This allows to decompose the whole analysis task into (a) analyzing the value distribution within each cluster and (b) making comparisons between the clusters. These tasks are well supported by visual techniques such as parallel coordinates and frequency histograms. However, these tools are insufficient for analyzing time-varying multidimensional data. Besides subtasks (a) and (b), the analyst needs to perform subtask (c): analyze how the cluster membership of the objects characterized by the multiple attributes changes over time. Such analysis is required in meteorology, however currently used tools focus on univariate time series (one element). This is the kind of task DCDV is designed to support.

Figure 9b shows the result of clustering in cluster-colored parallel coordinates plot (the choice of the clustering method and parameter settings is beyond the focus of this paper). The horizontal axes on the parallel coordinates plot (PCP) correspond to the four



(a) Representative time moments

(b) Trajectories for Milano 2 area

Figure 7: Visualization of trajectories of people visiting Milano 2 area (in cyan) on Sunday. a) DCDV view with representative time steps. b) Trajectories of the people who were in Milano 2 at these time steps. Right: Zoom into the area where all trajectories meet.

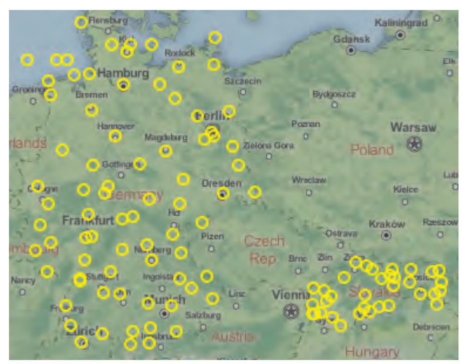
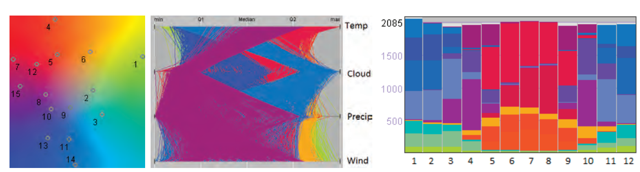


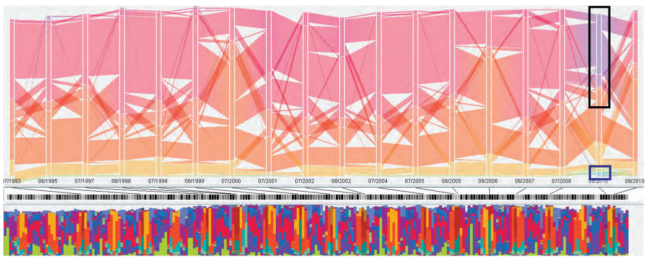
Figure 8: Locations of the weather stations used in the example. They comprise 111 weather stations over Germany and Slovakia.



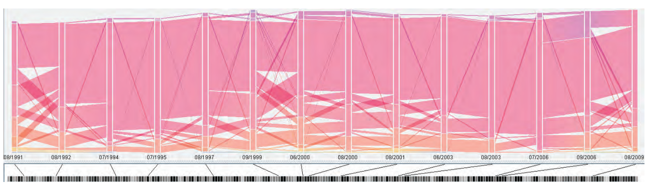
(a) Cluster colors (b) Cluster visualization in parallel coordinates (c) Number of objects pro cluster over year

Figure 9: Clustering result for weather data. a) Similarity-based assignment of colors to weather clusters. b) Parallel coordinates plot showing the cluster values. c) Number of objects in the clusters by months over all years.

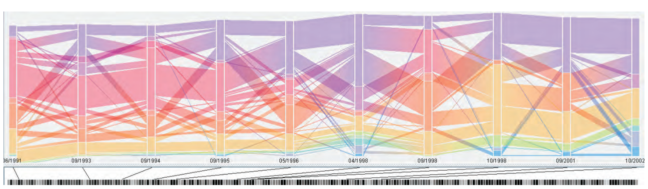
weather attributes (temperature, wind force, cloudiness and precipitation). The axes are quantile-scaled [3]; hence, the central position of an axis corresponds to the attribute's median. The colors for the clusters are chosen by projecting the cluster centers onto a 2D color space (see Figure 9a) so that similarity of cluster colors signifies similarity of the clusters. By selecting the clusters one by one (to reduce the overplotting), we see that red clusters correspond to high temperatures. Shades of orange represent quite warm weather but not so warm as shown in red. Green color corresponds to strong winds with lower temperatures, and shades of blue to low temperatures and either low or high cloudiness with low to medium precipitation. Red clusters with warm weather are more often found in summer months and blue and green clusters are more prominent in winter. The cluster distribution over a year shown in Figure 9c also confirms this finding. From now on, we treat these clusters as different weather types.



(a) Cluster 1: Summer months with predominantly warm weather



(b) Cluster 2: Summer months with stable extremely warm weather



(c) Cluster 3: Spring and autumn with changing weather conditions

Figure 10: Results of algorithm 2 identifying clusters of similar time moments in terms of weather types. Three example clusters are shown. a) warm weather cluster with one exceptional month (highlighted in black box), when weather was influenced by Iceland's volcano outbreak. b) Summer months with extremely warm weather. c) Spring and autumn variable weather conditions.

Since the weather changes periodically (seasonally over the year), looking at the changes from month to month is not very meaningful as the differences are usually high (i.e., all locations change their weather type) and the character of the changes is mostly known. It is more relevant to compare the weather in different years and to group time steps by similarity for revealing repeated patterns of weather types. Therefore, we employ Algorithm 2 to identify clusters of similar time moments, which we use for time selection in DCDV (see Figure 10).

Figure 10 shows time dependent data for three different time clusters. The upper two images portray time clusters consisting mostly of summer months. As seen, they are dominated by the two shades of red corresponding to very warm weather; one of them is extremely warm and clear (darker red) and the other is slightly cooler and more cloudy. The extremely warm and clear weather dominates in the second time cluster (b).It includes, in particular, August of 2003, which is still in the memory of Europeans for its extremely high temperatures with low cloudiness and low precipitation. More recent cases of similar weather were in July and September 2006 and in August 2009.

The time cluster 3 (see Figure 10c) includes early autumn and late spring months, which are characterized by strong weather variability over the territory. Interesting is that the latest time step is October 2002, i.e., such distributions of weather types as in this time cluster did not occur since then (i.e., until end-2010).

In perfect time clustering, we would expect homogeneous clusters, however the time clusters show some anomalies. As an exception, the time cluster 1 (see Figure 10a) includes May 2010 and the time cluster 2 (see Figure 10b) includes Septembers of 1999 and 2006. A closer look at the blue group in May 2010 (highlighted with black box in Figure 10a) within the geographic view (see Figure 11 right) reveals that all stations with this weather type were in Germany. So most part of Germany was cooler and more cloudy than the summer months in this time cluster but in Slovakia it was as warm as in the summer months. The meteorologist explained that this phenomenon can occur when two different pressures prevail in the regions. In this case, however, it could have been caused by a large outbreak of the Iceland Vulcano "Eyjafjallajökull" [29], which influenced the weather in northern and eastern Europe.

The time cluster 1 shows also few exceptional stations in May 2010 which are in green color (see Figure 10a, blue highlight). This corresponds to rather windy and cold weather. We find that these are mountain stations (Zugspitze, Feldberg and Strbske Pleso). The same stations appear in blue (rather cold, rainy weather) in the time cluster 3 (see Figure 10c). Analyzing the weather development in selected stations over the selected months in DCDV reveals that these stations are mostly characterized by worse weather than the others. However, there are exceptions. For example, in July 2002 both Zugspitze and Feldberg, and in July 2008 only Zugspitze moved to one of the red clusters (having exceptionally warm cloudless weather) (see Figure 12).

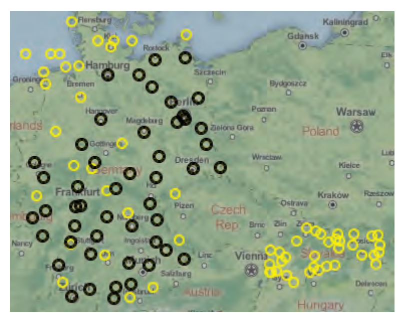


Figure 11: Locations of the weather stations selected in the DCDV view of Cluster 1. It shows locations with exceptionally cool weather in May 2010 (see Figure 10a). All selected locations are in Germany.

6 Discussion

The presented approach allows for visual analysis of categorical spatio-temporal data from various domains. It however has limitations. The main issues of our approach are scalability with regard to (w.r.t.) the number of categorical states, objects and time steps as well as the ordering of the categories in the view.

Scalability w.r.t. the number of categorical states The presented approach is suitable mainly for a small number of categoric values. About ten categories can be handled in the visualization so that the views are well readable. However, often larger number of

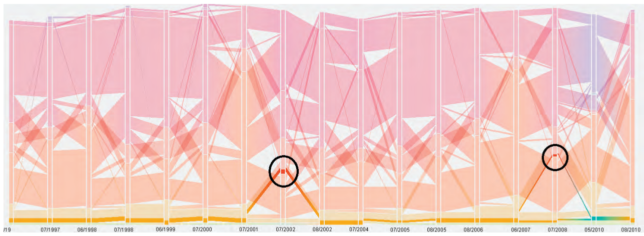


Figure 12: Weather conditions at two selected German Stations (Zugspitze and Feldberg), who had exceptionally cold weather in May 2010 in Cluster 1. These stations have normally colder and windier weather than other stations, however in July 2002 and July 2008 they also moved to red cluster – warm and cloudless weather.

categoric values exist in the data set. This can lead to overplotting problems. In order to support the analysis of larger number of categoric values, we provide a possibility to interactively merge several categories into one category. For example, the user can merge several similar activities into one broadly-defined activity such as tennis, walking and jogging can be integrated into sport. Alternatively, the user can merge categories that are not in her focus into an "other" category. It allows for targeted analysis of selected categories. As an illustration, Figure 13 shows a view on transitions for original categories (left) and for merged categories (right). The merged variant provides a more abstract view of the data with less overploting.

Scalability w.r.t. number of objects Data sets with large number of objects do not pose problems to the system as our approach concentrates mainly on groups of objects (defined as objects having the same categoric value). In this way, the data set is condensed into a much smaller set of object groups to be analyzed. If the number of groups is still large, the above-mentioned merging can be applied.

Scalability w.r.t. number of time steps In order to deal with a large number of time steps, we provide a guided selection of representative steps. The number of time steps for examination is determined by the user in an interactive informed way. This choice is determined by data properties and user preferences. For diverse datasets and user tasks, we proposed several time-selection algorithms (see Section 4.3).

Figure 13: Example of group merging.

Ordering of categories The ordering of categorical states is important both for good interpretation and readability of the visualization. Often, this ordering is not pre-defined. We currently provide interactive user-defined ordering. In this case, user experience and hypotheses about possible insights into the data determine the choice of ordering. Alternatively, the ordering can be calculated on the basis of transition counts. For example, approaches on the minimization of edge crossings for cluster comparison visualization [24, 41] or on placement of graph nodes on a line with minimization of edge crossing/edge length [9] can be used. However, these approaches have severe limitations. Not only the finding of optimal solution is a NP hard problem, but, for time-dependent data, it is not clear whether to apply these algorithms on all or only selected time steps. A global minimization may not be optimal for the selected time points and local calculation for selected time points would imply changing the ordering each time when time selection changes. Therefore, we currently rely on user-defined ordering.

7 Conclusions and Future Work

We have presented a new approach for visual analysis of spatio-temporal categorical data supported by algorithms for selection of globally and focally representative time steps based on categorical changes. The approach involves dynamic bi-directional linkages between geographic and categorical views, which are essential for comprehensive analysis of this class of data with respect to its main components: space, time, and categorical attributes. The user can select objects or places in space and see how they develop over time, or select object groups according to their temporal behavior and locate them in space, or select time moments in the categorical view and see corresponding spatial situations. The linkages not only support the basic "when, where, what" queries but also overall understanding of processes developing in space and time.

To test our approach, we have applied it to two different datasets focusing on mobility of a large number of people and weather changes over long time. We have demonstrated how the selection of representative time moments and the links between the categorical and geographical views allowed us to discover interesting spatio-temporal behaviors.

We have asked experts in the two use case domains discussed in Section 5 for feedback on usability of our system.
1) A mobile phone company's researcher found the proposed approach to be very interesting, as it allows to detect general mobility patterns without breaking individual's privacy. Moreover, it enables the company to add value to the data that they have collected for different reasons such as billing, etc.
2) The second use case was developed in cooperation with an expert meteorologist. The system allowed for analysis of complex weather phenomena over longer time periods. Such analysis is required in meteorology, however not sufficiently supported by current tools.. Although the DCDV view is rather unusual, after explanation of its features it was well adopted. Especially, as it allowed for both comparison of time moments (bars) and for analyzing changes in weather situations between time moments (edges). Linking the geographic and categoric views provided new ways of comparing weather developments across regions and seasons.

In the future, we would like to conduct a user study testing the effectiveness of our approach. Moreover, we would like to explore the usefulness of the approach in other application areas including economy and finance, medicine, and biology.

Acknowledgements

This work was partially supported by DFG SPP 1335 projects "Visual Feature Space Analysis", "Visual Analytics Methods for Modeling in Medical Imaging" and "Visual Spatio-temporal Pattern Analysis of Movement and Event Data". We thank Kerem Gülensoy for help with the project.

References

[1] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47 –60, jan.-feb. 2008.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer-Verlag New York Inc, 2011.

[3] G. Andrienko and N. Andrienko. Constructing parallel coordinates plot for problem solving. In *1st International Symposium on Smart Graphics,*, pages 9–14, 2001.

[4] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007.

[5] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data*. Springer, Berlin, Germany, 2006.

[6] Y. Bishop and S. Fienberg. *Discrete Multivariate Analysis theory and practice*. MIT Press, 2007.

[7] J. Blaas, C. Botha, E. Grundy, M. Jones, R. Laramee, and F. Post. Smooth graphs for visual exploration of higher-order state transitions. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):969 –976, 2009.

[8] S. Bremm, T. von Landesberger, G. Andrienko, and N. . Andrienko. Interactive analysis of object group changes over time. In *EuroVA, Second Int. Workshop on Visual Analytics*, pages 41–44, 2011.

[9] J. Díaz, J. Petit, and M. Serna. A survey of graph layout problems. *ACM Comput. Surv.*, 34:313–356, Sep. 2002.

[10] P. Froehlich. Interactive Sankey Diagrams. In *IEEE Symp. on Information Visualization*, page 233, 2005.

[11] D. Guo, J. Chen, A. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461–1474, 2006.

[12] M. Hao, H. Janetzko, S. Mittelstädt, W. Hill, U. Dayal, D. Keim, M. Marwah, and R. Sharma. A visual analytics approach for peak-preserving prediction of large seasonal time series. In *Computer Graphics Forum*, volume 30, pages 691–700. Wiley Online Library, 2011.

[13] M. C. Hao, U. Dayal, D. A. Keim, and T. Schreck. Multi-resolution techniques for visual exploration of large time-series data. In *EuroVis, Eurographics/IEEE VGTC Symposium on Visualization*, pages 27–34, 2007.

[14] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver: Visualizing theme changes over time. In *IEEE Symp. on Information Visualization*, pages 115–123, 2002.

[15] H. Hofmann. Mosaic plots and their variants. *Handbook of data visualization*, pages 617–642, 2008.

[16] S. Imfeld. *Time, Points and Space-Towards a better analysis of wildlife data in GIS*. PhD thesis, Geographisches Institut der Universität Zürich, 2000.

[17] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics*, pages 993–1000, 2009.

[18] T. Kapler and W. Wright. Geotime information visualization. *Information Visualization*, 4(2):136, 2005.

[19] D. Ketchen and C. Shook. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6):441–458, 1996.

[20] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, pages 558–568, 2006.

[21] M. Kraak. The space-time cube revisited from a geovisualization perspective. In *Proc. 21st International Cartographic Conference*, pages 1988–1996, 2003.

[22] H. Kriegel, P. Kröger, and A. Zimek. Outlier detection techniques. In *Tutorial at the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2009.

[23] O. Lampe and H. Hauser. Curve density estimates. In *Computer Graphics Forum*, volume 30, pages 633–642. Wiley Online Library, 2011.

[24] A. Lex, M. Streit, C. Partl, K. Kashofer, and D. Schmalstieg. Comparative analysis of multidimensional, quantitative data. *IEEE Trans. on Visualization and Computer Graphics*, 16(6):1027–1035, 11 2010.

[25] T. May. Working with patterns in large multivariate datasets-karnaugh-veitch-maps revisited. In *11th Int. Conf. on Information Visualization*, pages 277–285. IEEE, 2007.

[26] Mei-Po and Kwan. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1-6):185 – 203, 2000.

[27] C. Panagiotakis, A. Doulamis, and G. Tziritas. Equivalent key frames selection based on iso-content principles. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(3):447–451, 2009.

[28] D. Peuquet. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3):441–461, 1994.

[29] K. Schäfer, W. Thomas, A. Peters, L. Ries, F. Obleitner, J. Schnelle-Kreis, W. Birmili, J. Diemer, W. Fricke, W. Junkermann, et al. Influences of the 2010 eyjafjallajökull volcanic plume on air quality in the northern alpine region. *Atmos. Chem. Phys*, 11:8555–8575, 2011.

[30] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996.

[31] B. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 3(1):3, 2007.

[32] E. Tufte and G. Howard. *The visual display of quantitative information*, volume 7. Graphics press Cheshire, CT, USA, 1983.

[33] C. Turkay, J. Parulek, N. Reuter, and H. Hauser. Interactive visual analysis of temporal cluster structures. In *Computer Graphics Forum*, volume 30, pages 711–720. Wiley Online Library, 2011.

[34] B. Tversky, J. Morrison, and M. Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

[35] F. van Ham, H. van de Wetering, and J. van Wijk. Visualization of state transition graphs. In *IEEE Symposium on Information Visualization*, pages 59 –66, 2001.

[36] F. B. Viegas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors studying cooperation and conflict between authors. In *Conference on Human-ComputerInteraction (CHI)*, pages 575–582, 2004.

[37] K. Vrotsou, J. Johansson, and M. Cooper. ActiviTree: Interactive visual exploration of sequences in event-based data using graph similarity. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):945 –952, nov.-dec. 2009.

[38] I. Witten, E. Frank, and M. Hall. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2011.

[39] J. Wood, D. Badawood, J. Dykes, and A. Slingsby. Ballotmaps: Detecting name bias in alphabetically ordered ballot papers. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2384–2391, 2011.

[40] J. Wood and J. Dykes. Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1348–1355, 2008.

[41] J. Zhou, S. Konecni, and G. Grinstein. Visually comparing multiple partitions of data with applications to clustering. In *Proceedings of SPIE*, volume 7243, page 72430J, 2009.

[42] H. Ziegler, M. Jenny, T. Gruse, and D. Keim. Visual market sector analysis for financial time series data. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 83 –90, 2010.

[43] H. Ziegler, T. Nietzschmann, and D. Keim. Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments. In *International Conference on Information Visualisation*, pages 287 –295, 2008.