DASHTrails: An Approach for Modeling and Analysis of Distribution-Adapted Sequential Hypotheses and Trails

Martin Atzmueller and Andreas Schmidt and Mark Kibanov
University of Kassel, Research Center for Information System Design
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
{atzmueller, schmidt, kibanov}@cs.uni-kassel.de

ABSTRACT

The analysis of sequential trails and patterns is a prominent research topic. However, typically only explicitly observed trails are considered. In contrast, this paper proposes the *DASHTrails* approach that enables the modeling and analysis of distribution-adapted sequential trails and hypotheses. It presents a method for deriving transition matrices given a probability distribution over certain events. We demonstrate the applicability of the proposed approach using real-world data in the mobility domain, i. e., car trajectories and spatio-temporal distributions on car accidents.

Keywords

behavioral analytics, sequential hypothesis, human trails, mobility, social media, social network analysis

1. INTRODUCTION

The analysis of human behavior is a prominent topic in web and network science, e. g., for analyzing human navigation trails on the web or for exploring movement patterns in mobile and spatio-temporal applications. The HypTrails approach [18], for example, allows the comparison of hypotheses with such trails, for identifying the hypotheses that show the largest evidence concerning the observed data. However, the approach considers *explicitly observed* trails, e. g., navigational trails in social online systems. In contrast, this paper outlines an approach for the extended modeling and analysis of sequential hypotheses and trails, i. e., by deriving according transition matrices in a distribution-adapted approach. Then, we can analyze, e. g., geo-tagged datasets or (social) network data, with a probability distribution assigned to the data points and nodes of the network, respectively. There is a vast range of possible application areas. These include, e. g., the derivation and analysis of paths in mobile applications and social software, as well as analyzing complex heterogeneous networks in industrial plants, where e. g., connections between sensors (assets), events in alarm logs, and human (operator) actions can be investigated.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890082.

Objectives. Understanding the influence factors for trail-related events is important, e. g., for predictive modeling. This paper provides a systematic approach for modeling sequential trails and hypotheses – as a set of transitions between discrete states represented as transition matrices – given a probability distribution on these states. We can then model, e. g., (human) location-based indicators, geo-tagged datasets, or time-stamped events on complex networks, using the respective distributions. Specifically, we can model both *observed* and *derived* transition matrices and analyze them in a unified framework, given by the DASHTrails approach. In this paper, we exemplify the approach presenting first experiments that focus on the real world problem of understanding trail-related effects on car accidents, which can, e. g., be beneficial for resource planning and load balancing in health care, dynamic pricing for insurance companies as well as road planning and traffic control.

Contribution. Our contribution is summarized as follows:

1. We propose a systematic method for the modeling and analysis of sequential trails and hypotheses that are *derived* from *observed data*, i. e., estimated given a probability distribution of certain events. This enables a data-driven approach for analysis, comparison and assessment of sequential trails and hypotheses.
2. We present a modeling method that applies a certain process interpretation for deriving transition matrices, e. g., based on flow-based mechanisms: The transition matrices incorporate thus a certain interpretation of the data towards a sequential representation. This enables a comprehensive analysis approach, e. g., by estimating and comparing evidence for the hypotheses induced by a set of influence factors, statistically grounded utilizing Bayesian estimation techniques.
3. We demonstrate the applicability of the proposed approach and provide first results using real-world data in the context of the Telecom Italia Big Data Challenge 2015[1]: We model human sequential trails given a spatio-temporal distribution on car accident claim data in seven different cities in Italy. Hypotheses include one given observed car trajectories.

The remainder of the paper is organized as follows: Section 2 discusses related work. After that, we introduce the proposed approach in Section 3. Section 4 presents the results using real world mobility data from the Big Data Challenge 2015. Finally, Section 5 concludes the paper with a discussion and some interesting directions for future work.

[1] http://www.telecomitalia.com/tit/en/bigdatachallenge/

2. RELATED WORK

The investigation of human interactions in social networks, their dynamics, and finally sequential analysis are interesting and challenging tasks in data mining and social science. Contacts patterns, for example, and their underlying mechanisms, e. g., [13, 15] are a classic topic of social network analysis. Also, the analysis of human behavioral and mobility patterns has received increasing attention recently, covering localization, human interactions, as well as proximity and mobility patterns, cf., [23, 4, 7, 9, 6, 17, 2, 12, 3, 1]. Macek et al. [12], for example, analyzed the dynamics of participants' contact patterns, and also the connection between academic jobs and status at a conference. Furthermore, Atzmueller et al. [2] described the dynamics of community structures and roles at conferences, while Kibanov et al. [11] focused on their evolution. However, the analysis in these contexts does not transcend to sequential mobility, and modeling approaches. For that, important goals concern, e. g., analysis [9], modeling and characterization [23], and the predictability of mobility patterns [20]. In addition, there are several approaches for mining such patterns, e. g., [7, 22, 10]. Furthermore, navigational patterns, as a kind of mobility patterns in online systems, have been analyzed and modeled, e. g., [16, 19].

In contrast to those approaches, the proposed approach is not only about comparing sequential patterns. We provide a systematic approach for the analysis of sequential transition matrices derived given a probability distribution over certain events. Thus, similar to evidence networks in the context of social networks, e. g., [14], we can model according transitions assuming a certain interpretation of the data towards a sequential representation.

Singer et al. [18] propose HypTrails for comparing hypotheses on human trails. Like for DASHTrails, a trail is represented by a sequence of transitions between states. Originally, HypTrails has been applied to click data on the web and recently to geo-spatial trajectory data [5], but no work has been proposed that incorporates probability distributions for deriving transitions with a Markov chain modeling [19] interpretation so far. Hence, DASHTrails enables an extended modeling approach compared to HypTrails, while utilizing its Bayesian inference [21, 19] technique.

With respect to our human mobility example, relationships between traffic accidents and other influence factors like weather and lightning conditions have been reviewed in [8]. In contrast to that, our work is not about prediction, but on the assessment of certain (complex) hypotheses in order to identify those that explain the observed data best.

3. METHOD

In the DASHTrails approach, we integrate modeling and analysis of transition matrices given a probability distribution of certain states (events). We assume a discrete set of such states Ω (without loss of generality $\Omega = \{1, \ldots, n\}$, $n \in \mathbb{N}, |\Omega| = n$). We derive the transition matrices (modeling transitions between the respective events) using a certain *transition modeling function* $\tau : \Omega \times \Omega \rightarrow \mathbb{R}$. For that, we utilize the given probability distribution, e. g., concerning frequencies of page views, human presence in a certain location, or alarms in a sensor network. The transition modeling function τ captures a certain interpretation of the probabilities, e. g., based on flow-based mechanisms.

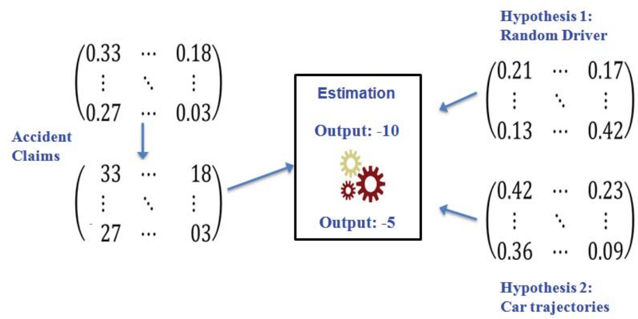

Figure 1: Our example compares a transition matrix derived from car accident claims (observed data) to two hypotheses in order to identify the one which explains the accident claims best: random driver hypothesis (null model), or car trajectory model.

Using τ, we can model *derived* matrices corresponding to the *observed data*, e. g., given densities on accidents on a spatial grid, as well as hypotheses on these trails. In our example, these correspond, e. g., to accident transition probabilities between the individual grids. For trails, we need to map the transition matrices into derived counts in relation to the data; for hypotheses we provide the (normalized) transition probabilities. For assessing a set of hypotheses that consider different transition probabilities between the respective states, we apply HypTrails [18] for comparing a set of hypotheses representing certains beliefs with respect to an (observed) transition matrix. Please note, that HypTrails can be seen as a special case of DASHTrails, if we only consider *explicitly observed trails*, since then we do not need to apply our distribution-adapted modeling approach.

3.1 Overview

Given data for constructing trails and a set of hypotheses, we perform three steps, exemplified for our case study as shown in Figure 1; the extension to other domains, e. g., web pages and view counts, or event counts on sensor networks is straight-forward. Below, we present a behavioral analytics example in the domain of human mobility.

1. Modeling: Determine a transition model given the respective probability distribution using a *transition modeling function* $\tau : \Omega \times \Omega \rightarrow \mathbb{R}$. Transitions between sequential states $i, j \in \Omega$ are captured by the elements m_{ij} of the transition matrix M, i. e., $m_{ij} = \tau(i, j)$.

2. Generation: Collect sequential transition matrices for the sequential trail and hypotheses. Obtain a transition matrix M specifying the sequential transitions given the respective transition model τ.

3. Estimation: Apply HypTrails, cf., [18] on the given trail transition matrix and the respective hypotheses, and return the resulting evidence.

In the estimation step, we apply the core Bayesian estimation step of HypTrails for comparing a set of hypotheses representing beliefs about transitions between states. Below, we describe how we construct the transition matrices for modeling and analysis of distribution-adapted sequential hypotheses and trails.

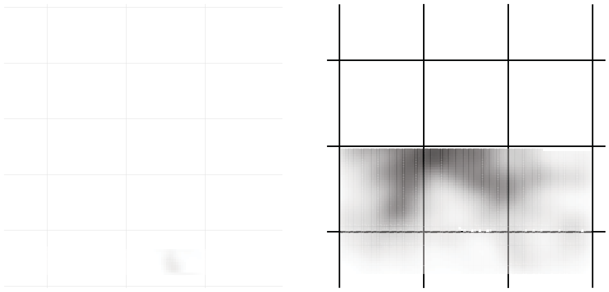

Figure 2: Example: Spatial distribution of accident claims for the city of Rome (left) and its downtown area(right). Darker regions indicate a higher density of accident claims per grid.

3.2 Modeling Approach

For modeling sequential hypotheses and trails, we consider a sequential interpretation (Markov process) with respect to the targeted transition probabilities (Markov chain). As outlined above, this results in a transition matrix M between a set of states Ω. In the case of mobility trails, these can, for example, correspond to spatially-layouted grids.

For explicitly observed trails as sequences of certain states we can simply construct transition matrices counting the transitions between the individual states, e. g., corresponding to the set of points in a grid. Then, $\tau(i,j) = |suc(i,j)|$, where $suc(i,j)$ denotes the successive sequences from state i to state j contained in the trail. In contrast, for deriving transition matrices from a probability distribution over certain events, we need to apply a more complex modeling approach.

Intuitively, when deriving the transition matrices, we can construct transitions between neighboring grids, for example, with increasing (or decreasing) density. Thus, we assume flow-based processes, such that transitions occur similarly to gradient ascent in the distributions. By intuition, probability flows along the grids towards the regions with high density. Then, we can test whether we can find evidence for explaining these transitions. Figure 3 shows an example of the spatial distribution of accident claims in the city of Rome: The figure indicates the regions with higher (and lower) densities, such that we can intuitively spot the transitions with increasing density. Figure 3 visualizes this idea using the spatial event information (accident claim frequencies): The arrows show the transitions with increasing density. Intuitively, these are the transitions that we model using the given *transition modeling function* τ utilizing the information contained in the events and their distributions.

3.3 Matrix Construction

For generating transition matrices, we utilize the given probability distribution on certain events. In the case of time-stamped events, we can aggregate data by a sliding window in order to provide different detail levels, e. g., by aggregating data per day. Without loss of generality, we can then convert non-sequential data (e. g., accident claim data) into a transition matrix by counting frequencies of accident claims within the respective sliding window position and modeling these distributions. Furthermore, our mobility example also uses a pre-specified grid of states.

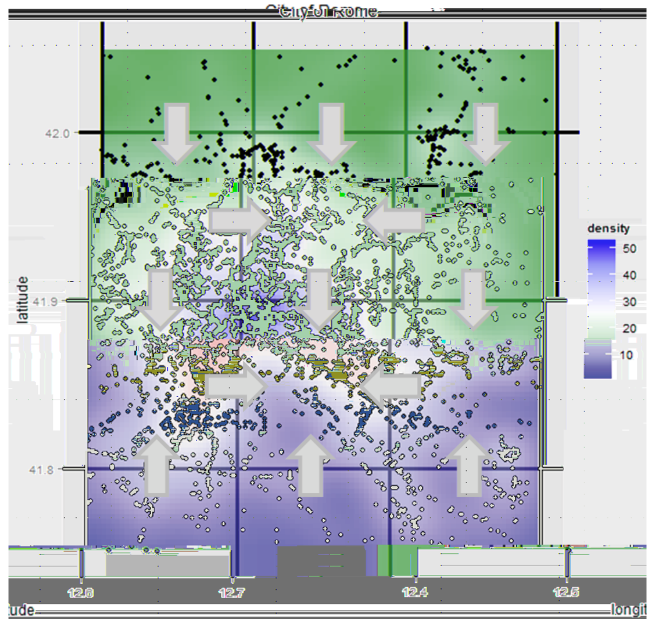

Figure 3: Example for modeling spatial (grid-based) transitions using accident frequencies (for Rome).

3.3.1 *Transition Modeling*

Formally, let $G = \{g_1, g_2, ..., g_n\}$ be a set of grids (and accordingly $\Omega = \{1, 2, \ldots, n\}$, assuming a suitable bijection) with a neighborhood defined by the function:

$$\mathrm{adj}(g_i, g_j) = \begin{cases} 1 & \text{if grid } g_i \text{ is adjacent to } g_j \\ 0 & \text{otherwise} \end{cases}$$

Furthermore we define the functions $\mathrm{freq}(g_i)$ to count the number of data points within a grid (in a time-based analysis: relative to the current sliding window position). Using τ as described below, we construct the $|G| \times |G|$ transition matrix M with the entries m_{ij} for transitions between grids:

$$m_{ij} = \mathrm{adj}(g_i, g_j) \cdot \delta(g_i, g_j)\,,$$

with

$$\delta(g_i, g_j) = \frac{\mathrm{freq}(g_j) - \mathrm{freq}(g_i)}{\sum_{g_k \in G} \mathrm{freq}(g_k)}\,.$$

We will refer to this case as a linear transition. Intuitively, m_{ij} will be 1 when we have a transition from grid g_i to g_j with maximum increasing frequency, i. e.,

$$\mathrm{freq}(g_i) = 0, \text{ and } \mathrm{freq}(g_j) = \sum_{g_k \in G} \mathrm{freq}(g_k)\,.$$

In contrast m_{ij} will be -1 when we have a transition from grid g_i to g_j with maximum decreasing frequency, i. e.,

$$\mathrm{freq}(g_i) = \sum_{g_k \in G} \mathrm{freq}(g_k), \text{ and } \mathrm{freq}(g_j) = 0\,.$$

For the linear function we observed that transitions could have a value near zero because of small variance in the frequency distributions of the grids. To account for this problem, we also investigated other transitions by transforming m_{ij} with a non-linear function.

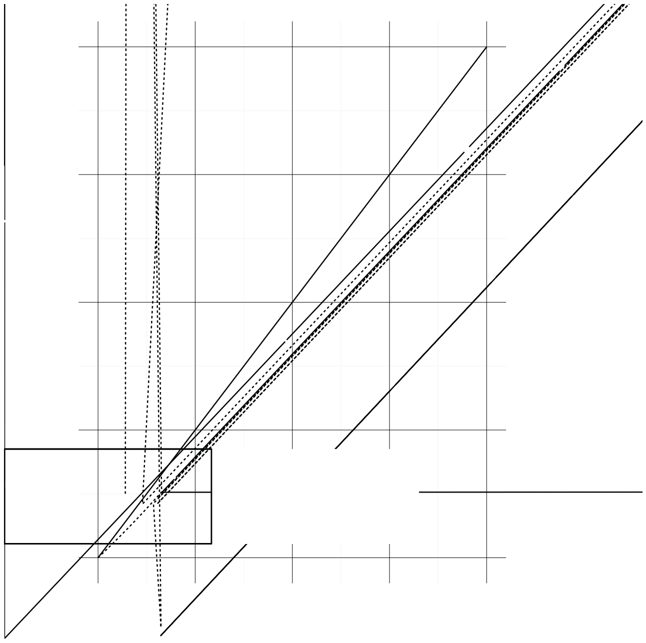

Figure 4: Exemplification of different logarithmic transformation options.

We can perform, e. g., a logarithmic transformation to have a larger "spread" of the values across the domain range, i. e., giving smaller differences in the frequencies a higher impact. The transformation for a given base b is given as:

$$m'_{ij}(b) = \text{sign}(m_{ij}) \cdot \frac{\log(|m_{ij}|+1)}{\log(b)}$$

Different transformation options are shown in Figure 4, demonstrating the effects of parameter b in order to have a larger weight also for smaller differences between two distributional points.

Finally, we map the values v of the matrix entries m_{ij} (and $m'_{ij}(b)$, respectively) from the range $[-1, 1]$ to $[0, 1]$. This can be done by applying the function

$$\text{standardize}(v) = \frac{1}{2} \cdot (v+1)\,,$$

or by strictly taking the negative $\text{neg}(v)$ or positive $\text{pos}(v)$ values. The latter allows us to interpret our gradient ascent idea rather strictly, by focusing on increasing frequencies. After that, we normalize M as outlined above.

$$\text{pos}(v) = \begin{cases} v & \text{if v} > 0 \\ 0 & \text{otherwise} \end{cases}, \ \text{neg}(v) = \begin{cases} -v & \text{if v} < 0 \\ 0 & \text{otherwise} \end{cases}$$

3.3.2 Null Model

In our experiments we also used a random model (e.g. random driver with equal distributed weight) as a (lower bound) base line. We define the $|G| \times |G|$ random transition matrix R by computing:

$$r_{ij} = \begin{cases} \Big(\sum\limits_{g_p, g_q \in G} \text{adj}(g_p, g_q)\Big)^{-1} & \text{if adj}(g_p, g_q) > 0 \\ 0 & \text{otherwise} \end{cases}$$

4. CASE STUDY

In the following, we describe first results of an example case of the proposed approach using a real-world dataset in the mobility domain. Below, we first describe the dataset and the applied data preprocessing steps. After that, we describe and discuss the results in detail.

4.1 Dataset

We utilized data from the Telecom Italia Big Data Challenge 2015, including about 56 GB of floating car data in seven different cities (Bari, Milano, Napoli, Palermo, Roma, Torino, Venezia), i. e., trajectories (ordered sequences of locations/points) of different vehicles in those cities arranged in different trips per car. Additionally, raw car accident claim data with a total of about 1 MB was provided denoting when and where an accident had occurred. In addition, the dataset contained also information about the individual layout of the grids for the different cities.

After preprocessing, we obtained an extended accident claim dataset of about 77 MB, and a floating car dataset of about 4.3 GB modeling the car trajectories. Both datasets were anonymized and contained the information per day of the week as temporal context information. Table 1 provides an overview on characteristics of the data.

Table 1: Description of the used data of the Telecom Italia Big Data Challenge 2015: For each city, the table shows the grid size, the number of car accident claims and the number of trips corresponding to the car trajectories.

	Grid Size	No. Claims	No. Trips
Bari	144x144	2836	4787547
Milano	1419x1419	4739	6259796
Napoli	535x535	6657	11822672
Palermo	165x165	1799	3250460
Roma	928x928	7031	7229498
Torino	571x571	3147	3273889
Venezia	496x496	881	1365384

4.2 Modeling Transition Matrices

We applied the process described above for transition modeling and matrix generation using the spatial information about the grid layout as given in the dataset.

4.2.1 Car Accident Claims Data

Using the accident claims data, we applied the transition modeling step outlined above, and obtained transition matrices for the different cities accordingly relative to the respective claims distributions. Also, since a trail should be explained best by an according hypothesis, we considered a claims hypothesis, i. e., that transitions between grids are distributed by insurance claims from car accidents.

4.2.2 Car Trajectory Data

For the car trajectory hypothesis, we made use of the given sequential car trajectories as sequences of grids.

Let

$$S_i = [s_1, s_2, ..., s_m]$$

with $s_i \in G$ be a sequence of grids (e.g. for a floating car trip) and $S_i(n)$ denote the n-th element of the sequence.

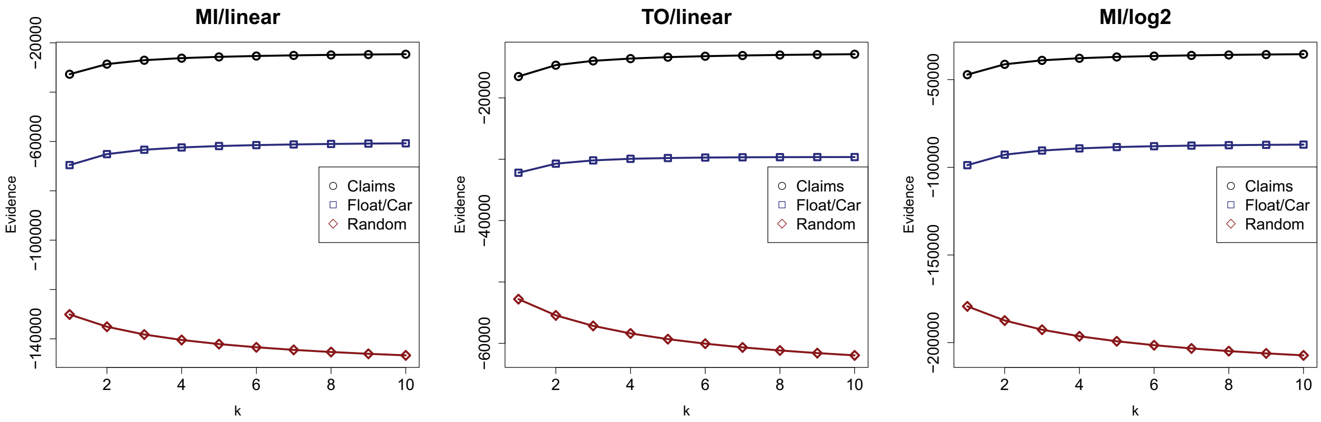

Figure 5: Exemplary results for Milano (MI) and Torino (TO). The modeled car accident claims transitions (Claims) are much better explained by the car trajectories (Float/Car) than by the null model (Random) – for increasing degree of belief (parameter k) On the right, we observe the effect of the log2 transformation.

Furthermore, let

$$S = \{S_1, S_2, ..., S_n\}$$

be the set of all sequences (in a time-based analysis: within the respective sliding window position, e. g., all trips within the morning rush hour). We define the following function to count successive sequence elements:

$$\text{suc}(S_v, g_i, g_j) = |\{e \,|\, g_i = S_v(e) \wedge g_j = S_v(e+1)\}|$$

The $|G| \times |G|$ transition matrix F for trajectory data is then computed by:

$$f_{ij} = \frac{\sum_{S_v \in S} \text{suc}(S_v, g_i, g_j)}{\arg\max_{g_j \in G} \sum_{S_v \in S} \text{suc}(S_v, g_i, g_j)}$$

4.2.3 Null Model

We constructed a null model using the random driver hypothesis described above, utilizing the grid information, such that transitions between grids are uniformly distributed.

4.3 Results and Discussion

In our experiments, we investigated how good the claim data fit with the car trajectory data. Thus, we modeled a derived transition matrix concerning the accident claim distribution, and according transition matrices for the car trajectory data as hypotheses, for each city. A trail should be explained best by an according hypothesis, therefore we also included an according accident claims hypothesis (using the derived claims transition matrix) for comparison. In addition, we applied the null model given by the random driver. Given the observed (claims) data, we expected the floating car hypothesis to be between both hypotheses (claims, random driver) – modeling transitions between grids according to car trajectories. Intuitively, the closer the evidence of the floating car hypothesis to the claims hypothesis, the better. This provides us with an indication how good the insurance claims can be explained by transitions in the car trajectories.

Results can be observed in Figure 5, exemplarily shown for the cities Milano (MI) and Torino (TO) - the results for the other five remaining cities yielded similar trends.

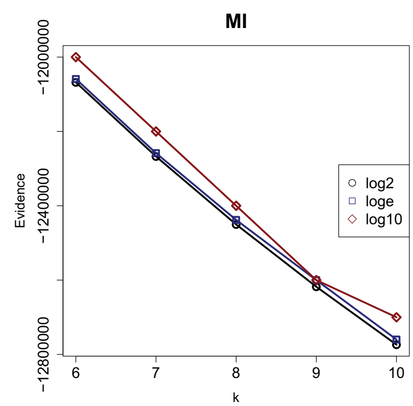

Figure 6: Example of the logarithmic transformation for Milano (MI): Three transformed car accident hypotheses compared to the car trajectory trails, for increasing belief (k).

Overall, the obtained evidence values indicate a reasonable fit of the floating car hypothesis and the accident claims' trails for the different cities. The floating car hypotheses derived from the car trajectory data scores better (in terms of evidence) than the random driver, indicating that the accident claims can indeed be explained by floating car data to a certain extent. Furthermore, we observe the derived claims hypothesis as an upper bound compared to the floating car hypothesis which also supports the validity of our modeling approach. In addition, the (logarithmic) transformation options can be used for targeted adjustments with respect to the transitions – with the effect of improving the relative evidence compared to the linear transformation. Figure 6 also shows an example of three transformed hypotheses.

These results already indicate the potential of DASHTrails for distribution-adapted modeling and analysis: As a data-driven approach, DASHTrails provides a powerful tool – grounded on Bayesian estimation [18] – for assessing influence factors with respect to derived and observed data.

5. CONCLUSIONS

In this paper, we presented a systematic approach for modeling and analysis of distribution-adapted sequential hypotheses and trails. Using the generated transition matrices, evidence for different hypotheses (e. g., based on potential influence factors) can be estimated and analyzed in a data-driven approach that is statistically grounded. Furthermore, we demonstrated the applicability of the proposed approach by first results in the context of real-world urban mobility data from the Telecom Italia Big Data Challenge 2015.

We are currently extending the approach into diverse scenarios, e. g., for modeling and analyzing complex heterogeneous networks (as complex graph structures) in industrial production contexts, where e. g., connections between sensors (assets), alarms, and human (operator) actions are analyzed. Here, the fit of certain hypotheses from dynamic and static behavior (represented by according hypothesis) in the corresponding log data can be modeled and analyzed using DASHTrails.

In future work, we aim to investigate further transition modeling adaptations in more detail, e. g., for the application and extension of the approach towards network-structured and complex graph data and according interpretation mechanisms. Furthermore, we also aim to extend the approach for analyzing combinations of hypotheses and temporal (propagation) processes, and their introspection.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) project FEE under grant number 01IS14006E. We wish to thank Andreas Hotho and Florian Lemmerich for comments and discussions, also especially concerning Florian's implementation of HypTrails[2].

6. REFERENCES

[1] M. Atzmueller, M. Becker, M. Kibanov, C. Scholz, S. Doerfel, A. Hotho, B.-E. Macek, F. Mitzlaff, J. Mueller, and G. Stumme. Ubicon and its Applications for Ubiquitous Social Computing. *New Rev Hypermedia M*, 20(1):53–77, 2014.

[2] M. Atzmueller, S. Doerfel, A. Hotho, F. Mitzlaff, and G. Stumme. Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In *Modeling and Mining Ubiquitous Social Media*, volume 7472 of *LNAI*. Springer Verlag, Berlin, 2012.

[3] M. Atzmueller and K. Hilgenberg. Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection. In *Proc. 4th International Workshop on Modeling Social Media (MSM 2013), Hypertext 2013*, New York, NY, USA, 2013. ACM Press.

[4] A.-L. Barabasi. *Linked the New Science of Networks.* Perseus Pub., Cambridge, Mass., 2002.

[5] M. Becker, P. Singer, F. Lemmerich, A. Hotho, D. Helic, and M. Strohmaier. VizTrails: An Information Visualization Tool for Exploring Geographic Movement Trajectories. In *Proc. HT*, pages 319–320, New York, NY, USA, 2015. ACM.

[6] C. Cattuto, W. V. den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, 5(7), 07 2010.

[7] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory Pattern Mining. In *Proc. ACM SIGKDD*, pages 330–339. ACM, 2007.

[8] T. F. Golob and W. W. Recker. Relationships among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions. *J Transp Eng*, 129(4), 2003.

[9] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453(7196):779–782, 2008.

[10] Y. Huang, L. Zhang, and P. Zhang. A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets. *IEEE TKDE*, 20(4):433–448, 2008.

[11] M. Kibanov, M. Atzmueller, C. Scholz, and G. Stumme. Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. *Sci China Inf Sci*, 57:1–17, 2014.

[12] B.-E. Macek, C. Scholz, M. Atzmueller, and G. Stumme. Anatomy of a Conference. In *Proc. HT*, pages 245–254, New York, NY, USA, 2012. ACM.

[13] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.*, 27(1):415–444, 2001.

[14] F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme. Community Assessment using Evidence Networks. In *Analysis of Social Media and Ubiquitous Data*, volume 6904 of *LNAI*, 2011.

[15] F. Mitzlaff, M. Atzmueller, A. Hotho, and G. Stumme. The Social Distributional Hypothesis. *Journal of Social Network Analysis and Mining*, 4(216), 2014.

[16] P. L. Pirolli and J. E. Pitkow. Distributions of Surfers' Paths Through the World Wide Web: Empirical Characterizations. *WWW*, 2(1-2):29–45, 1999.

[17] C. Scholz, S. Doerfel, M. Atzmueller, A. Hotho, and G. Stumme. Resource-Aware On-Line RFID Localization Using Proximity Data. In *Proc. ECML/PKDD*, pages 129–144, Berlin, 2011. Springer.

[18] P. Singer, D. Helic, A. Hotho, and M. Strohmaier. Hyptrails: A Bayesian Approach for Comparing Hypotheses about Human Trails. In *Proc. WWW*, Firenze, Italy, 2015. ACM, ACM Press.

[19] P. Singer, D. Helic, B. Taraghi, and M. Strohmaier. Memory and Structure in Human Navigation Patterns. *PLoS ONE*, 9(7), 2014.

[20] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of Predictability in Human Mobility. *Science*, 327(5968):1018–1021, 2010.

[21] C. C. Strelioff, J. P. Crutchfield, and A. W. Hübler. Inferring Markov Chains: Bayesian Estimation, Model Comparison, Entropy Rate, and Out-of-Class Modeling. *Physical Review E*, 76(1):011106, 2007.

[22] J. Yang and M. Hu. Trajpattern: Mining Sequential Patterns from Imprecise Trajectories of Mobile Objects. In *Advances in Database Technology*, pages 664–681. Springer, 2006.

[23] M. M. Zonoozi and P. Dassanayake. User Mobility Modeling and Characterization of Mobility Patterns. *IEEE J Sel Area Comm*, 15(7):1239–1252, 1997.

[2] `https://bitbucket.org/florian_lemmerich/hyptrails4j`