

INSTITUTO FEDERAL DO RIO GRANDE DO NORTE
CAMPUS NATAL - CENTRAL
DIRETORIA DE GESTÃO E TECNOLOGIA DA INFORMAÇÃO
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

An Spatial-Temporal Model to Explore Interesting Dense Regions over Time

Felipe Mateus Freire Pontes

Natal-RN
Mês (por extenso) e ano

Felipe Mateus Freire Pontes

An Spatial-Temporal Model to Explore Interesting Dense Regions over Time

Trabalho de conclusão de curso de graduação do curso de Tecnologia e Análise em Desenvolvimento de Sistemas da Diretoria de Gestão e Tecnologia de Informação do Instituto Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Linha de pesquisa:
Nome da linha de pesquisa

Orientador

Dr. Plácido Antônio de Souza Neto

TADS – CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS
DIATINF – DIRETORIA ACADÊMICA DE GESTÃO E TECNOLOGIA DA INFORMAÇÃO
CNAT – CAMPUS NATAL - CENTRAL
IFRN – INSTITUTO FEDERAL DO RIO GRANDE DO NORTE

Natal-RN

Mês e ano

Trabalho de Conclusão de Curso de Graduação sob o título *An Spatial-Temporal Model to Explore Interesting Dense Regions over Time* apresentada por Felipe Mateus Freire Pontes e aceita pelo Diretoria de Gestão e Tecnologia da Informação do Instituto Federal do Rio Grande do Norte, sendo aprovada por todos os membros da banca examinadora abaixo especificada:

Dr. Plácido Antônio de Souza Neto
Presidente

DIATINF – Diretoria Acadêmica de Gestão e Tecnologia da
Informação
IFRN – Instituto Federal do Rio Grande do Norte

Nome completo do examinador e titulação
Examinador
Diretoria/Departamento
Instituto

Nome completo do examinador e titulação
Examinador
Diretoria/Departamento
Universidade

Natal-RN, data da defesa (dia, mês e ano).

Homenagem que o autor presta a uma ou mais pessoas.

Agradecimentos

Agradecimentos dirigidos àqueles que contribuíram de maneira relevante à elaboração do trabalho, sejam eles pessoas ou mesmo organizações.

Citação

Autor

An Spatial-Temporal Model to Explore Interesting Dense Regions over Time

Autor: Felipe Mateus Freire Pontes

Orientador(a): Dr. Plácido Antônio de Souza Neto

RESUMO

O resumo deve apresentar de forma concisa os pontos relevantes de um texto, fornecendo uma visão rápida e clara do conteúdo e das conclusões do trabalho. O texto, redigido na forma impersonal do verbo, é constituído de uma sequência de frases concisas e objetivas e não de uma simples enumeração de tópicos, não ultrapassando 500 palavras, seguido, logo abaixo, das palavras representativas do conteúdo do trabalho, isto é, palavras-chave e/ou descritores. Por fim, deve-se evitar, na redação do resumo, o uso de parágrafos (em geral resumos são escritos em parágrafo único), bem como de fórmulas, diagramas e símbolos, optando-se, quando necessário, pela transcrição na forma extensa, além de não incluir citações bibliográficas.

Palavras-chave: Palavra-chave 1, Palavra-chave 2, Palavra-chave 3.

An Spatial-Temporal Model to Explore Interesting Dense Regions over Time

Author: Felipe Mateus Freire Pontes

Supervisor: Dr. Plácido Antônio de Souza Neto

ABSTRACT

O resumo em língua estrangeira (em inglês *Abstract*, em espanhol *Resumen*, em francês *Résumé*) é uma versão do resumo escrito na língua vernácula para idioma de divulgação internacional. Ele deve apresentar as mesmas características do anterior (incluindo as mesmas palavras, isto é, seu conteúdo não deve diferir do resumo anterior), bem como ser seguido das palavras representativas do conteúdo do trabalho, isto é, palavras-chave e/ou descritores, na língua estrangeira. Embora a especificação abaixo considere o inglês como língua estrangeira (o mais comum), não fica impedido a adoção de outras línguas (a exemplo de espanhol ou francês) para redação do resumo em língua estrangeira.

Keywords: Keyword 1, Keyword 2, Keyword 3.

Lista de figuras

- 1 The process of exploring Paris home-stays. p. 14

Lista de tabelas

Lista de abreviaturas e siglas

Sumário

1	Introduction	p. 13
1.1	Problem Definition	p. 13
1.1.1	Case Study	p. 14
1.2	Objectives	p. 15
1.2.1	General Objectives	p. 15
1.2.2	Specific Objectives	p. 16
1.3	Organization	p. 16
2	Background	p. 17
2.1	Related Work	p. 17
2.2	GeoGuide	p. 18
3	Data Model Definition	p. 19
3.1	Spatial layer	p. 19
3.2	Interesting Dense Regions	p. 19
4	Collecting feedback	p. 21
5	Applying temporal analysis	p. 22
6	Guiding the user	p. 23
7	Experiments	p. 24
7.1	Results	p. 24

8 Conclusion	p. 25
8.1 Contributions	p. 25
8.2 Restrictions	p. 25
8.3 Future work	p. 25
Referências	p. 26

1 Introduction

More than ever we are overwhelmed by an amount of data which is created every day. When we compare how data has been created over the past years, we realize that is already increasing significantly. Besides this evolution, nowadays we have the most diverse kinds of data (e.g., documents, tweets, pictures, videos, GIFs, check-ins).

This phenomenon has been called *Big Data* and represents an increasing field of study for the time being. Therefore researchers are analyzing and learning with these information we create. However the increasing amount of data make analyses hard to perform. So, people are investing in techniques and tools to tackle challenges such as data mining, data cleaning, data visualization, data classification, data exploration and so on.

One common kind of data is what we called *spatial data*, which are data that comes with geographical attributes like latitude and longitude (e.g., tweets, restaurants reviews, place check-ins). Spatial data can be very insightful, for instance, a check-in at the airport by your sister in the morning of your birthday, probably it means you will have a surprise.

As each record in spatial data represents an activity in a precise geographical location, analyzing such data enables discoveries grounded on facts. Analysts are often interested to observe spatial patterns and trends to improve their decision making process. Spatial data analysis has various applications such as smart city management, disaster management and autonomous transport (RODDICK et al., 2004; TELANG; PADMANABHAN; DESHPANDE, 2012).

1.1 Problem Definition

Spatial data analysis is often performed in *exploratory context*: the analyst does not have a precise query in mind and she explores data in iterative steps in order to find potentially interesting results. Traditionally, an exploratory analysis scenario on spatial data is described as follows: the analyst visualizes a subset of data using a query in an vi-

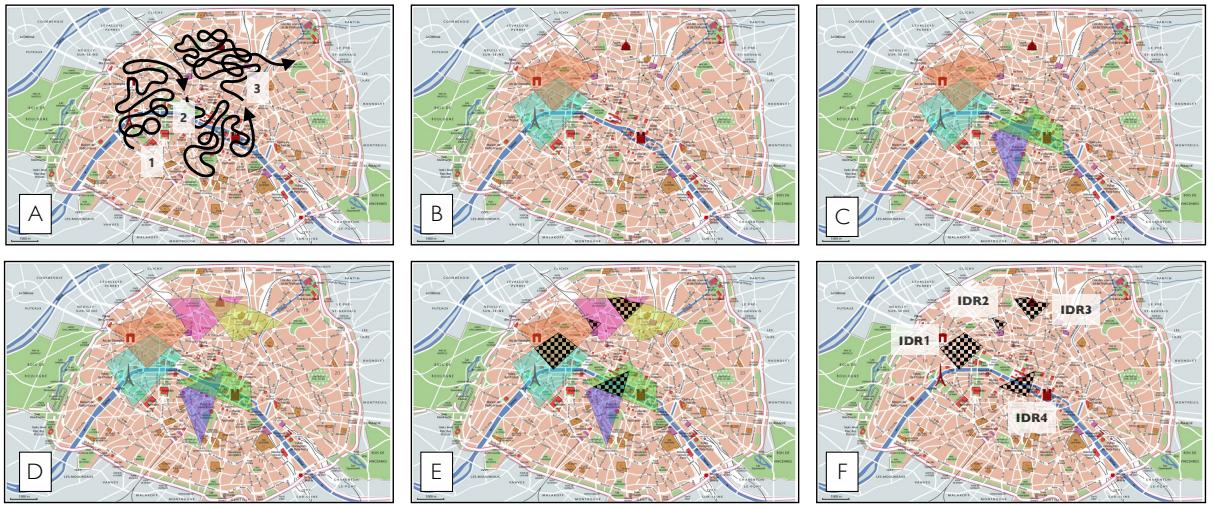


Figura 1: The process of exploring Paris home-stays.

sualization environment (e.g., Tableau¹, Exhibit², Spotfire³). The result will be illustrated on a geographical map. Then she investigates on different parts of the data by zooming in/out and panning the map in order to discover patterns and trends of interest. The analyst may iterate on this process several times by issuing different queries and focusing on different aspects of data.

The large size of spatial data make the analyst feel lost during the exploration. There could be thousands of points in each neighborhood of a city, for example. Analysts require to obtain only few options (so-called "highlights") to act as a direction and be able to focus on. In the perfect scenario, these options are not randomly chosen and represent what they showed to be interested in the previous iterations.

In this work, we formulate a solution for "information highlighting using feedback collected over time", i.e., highlight few geographical points based on implicit interests of the analyst in order to guide the user towards what he should concentrate on in consecutive iterations of the analysis process.

1.1.1 Case Study

Now, we will present a case study in order to show the functionality of our approach in practice.

¹<http://www.tableau.com>

²<http://www.simile-widgets.org/exhibit/>

³<http://spotfire.tibco.com>

Example. Lucas is planning to spend few days in Paris, France. His appreciation of French culture makes him interested in new experiences in the city. He decides to rent a home-stay from Airbnb website⁴. He likes to discover the city, hence he is open to any type of lodging in any region with an interest to stay in the city center. The website returns 4000 different locations. As he has no other preferences, an exhaustive investigation needs scanning each location independently which is nearly infeasible. While he is scanning few first options, he shows interest in the region of “Champ de Mars” (near Eiffel Tower), but he forgets or doesn’t feel necessary to click a point there. By collecting feedback on his mouse moves over the home-stays in Paris, our system can quickly detect his interest in the region and short-list a small subset of locations (i.e., highlights) accordingly to be recommended to Lucas.

We follow the above example to describe how implicit feedback is collected in action. Figure 1 shows Lucas’ steps to explore home-stays in Paris. Figure 1.A shows his mouse movements in different time stages. In this example, we consider $g = 3$ and capture Lucas’ feedback in three different time segments (progressing from Figures 1.B to 1.D). It shows that Lucas started his search around Eiffel Tower and Arc de Triomphe (Figure 1.B) and gradually showed interest in south (Figure 1.C) and north (Figure 1.D) as well. All intersections between those regions are discovered (hatching regions in Figure 1.E) which will constitute the set of *Interesting Dense Regions* (Figure 1.F), i.e., IDR1 to IDR4.

What if Lucas wanted to come back to Paris, France next year? He will have to repeat the same exploratory analyse, unless he remember the exact location of home-stays he showed interest last year. Using our system, he won’t need to remember, because his preferences were collected and can be used to highlight a subset of similar home-stays.

1.2 Objectives

In this section, we define the general and specific objectives of our work.

1.2.1 General Objectives

- Introduce a time-aware guidance approach for spatial data exploration...
- Elaborate how temporal analyses can be effectively applied in data exploration...
- Present the results of our guidance approach...

⁴<http://www.airbnb.com>

1.2.2 Specific Objectives

- Present our proposed guidance approach...
- Describe our data model used for temporal analyses...
- Describe our concept of IDR used for collecting feedback...

1.3 Organization

The next sections is as follow: in the Chapter 2 we discuss the background of this work. Chapter 3 defines the data model. Chapter 4 presents how the feedback is collected during exploration. Chapter 5 presents how temporal analysis is applied. Chapter 6 presents how highlight interesting points in order to guide the user using collected feedback and results from temporal analysis. Chapter 7 shows experiments and its results. Chapter 8 presents some conclusions and future directions.

2 Background

This section gives an overview of related work in literature about feedback exploitation, information-highlighting methods and temporal analysis applications. We also present the system we are extending.

2.1 Related Work

The literature in spatial data analysis has a focus on *efficiency* of exploratory interactions. The common approach is to design pre-computed index with enable efficient retrieval of spatial data (e.g., (LINS; KŁOSOWSKI; SCHEIDEGGER, 2013)). However, we should also put attention in the *value* of spatial data, because is very common to see an analyst getting lost in the huge amount of geographical points. In order to overcome this challenge, visualization environments (e.g., Tableau¹, Exhibit², Spotfire³) offer features to manipulate data (e.g., filters, aggregate queries, etc).

Our proposed visualization environment leverage the spatial data analysis by exploiting collected feedback during the analyst exploration to highlight subsets of geographical points. In the literature, are several instances of feedback exploitation to guide the analysts in further analysis steps (e.g., (BOLEY et al., 2013)).

There exist few instances of information-highlighting methods in the literature (LI-ANG; HUANG, 2010; ROBINSON, 2011; WONGSUPHASAWAT et al., 2016; WILLETT; HEER; AGRAWALA, 2007). All these methods are *objective* and do not apply to the context of spatial guidance where user feedback is involved. In terms of recommendation, few approaches focus on spatial dimension (BAO et al., 2015; LEVANDOSKI et al., 2012) while the context and result diversification are missing.

There are currently several instances which combine temporal analysis with spatial

¹<http://www.tableau.com>

²<http://www.simile-widgets.org/exhibit/>

³<http://spotfire.tibco.com>

data in the literature (BACULO et al., 2017; BALAHADIA; TRILLANES, 2017; CHIDEAN et al., 2018; GHAHRAMANI; ZHOU; HON, 2018; KAMATH; CAVERLEE, 2013; LOPES-TEIXEIRA; BATISTA; RIBEIRO, 2018; MA et al., 2017; MIJOVIć et al., 2016; TOMOKI; KEIJI, ; NARA; TORRENS, 2007; ZHAN et al., 2017; ZHENG et al., 2018). Those are applications of temporal analysis in specific context, which does not involve user feedback, but represent how temporal analysis could be insightful.

2.2 GeoGuide

GeoGuide (OMIDVAR-TEHRANI et al., 2017) is a spatial data visualization environment which keep track of user preferences during exploration in order to use this collected data to highlight subsets of geographical points that may be interesting to the analyst.

GeoGuide use both explicit and implicit feedback. Explicit feedback is when the user is analyzing the attributes of a point (e.g., the house description in a AirBnb context) and explicitly ask for similar (yet diverse) points to the current selected one. Implicit feedback is tracked using the mouse movements, eye/gaze tracking etc.

In this work, we will leverage GeoGuide into two new concepts: *i.* interesting dense regions and *ii.* understanding how the user preferences change over time.

3 Data Model Definition

3.1 Spatial layer

Each point in a dataset ($p \in \mathcal{P}$) is described using its coordinates (latitude and longitude) and also associated with a set of attributes ($\text{dom}(p)$). For instance, TODO

3.2 Interesting Dense Regions

TODO

We have IDRs per iteration/session where implicit feedback is captured such mouse moves (or eye gaze). In the beginning, each IDRs is a group of raw points described using its coordinates (latitude and longitude) and a timestamp (the unix timestamp it was captured). These raw points once captured will enter the clustering (for now, ST-DBSCAN) phase to generate the IDR itself with a profile. The profile is built based on the spatial layer and it should represent a summary of its contained points from the spatial layer.

- A profile has summary of its spatial points number attributes. For each number attribute in $\text{dom}(p)$, we calculate the average, median and standard deviation based on the points contained in the IDR.
- A profile has a word rank R of the terms in the text attributes of its spatial points. For each text attribute in $\text{dom}(p)$, we evaluate the most used terms in order to create a word rank (KUMAR; KAUR, 2017).
- A profile has a map M between the $\langle name, value \rangle$ of categoricals attributes and its relevance in $\text{dom}(p)$.
- TODO: datetime attributes

- A profile has a meta property with values such the count of points in the IDR.

4 Collecting feedback

TODO

5 Applying temporal analysis

TODO

6 Guiding the user

TODO

7 Experiments

TODO

7.1 Results

TODO

8 Conclusion

To our knowledge...

8.1 Contributions

TODO

8.2 Restrictions

TODO

8.3 Future work

TODO

Referências

- BACULO, M. J. C. et al. Geospatial-temporal analysis and classification of criminal data in manila. In: *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*. [S.l.: s.n.], 2017. p. 6–11.
- BALAHADIA, F. F.; TRILLANES, A. O. Improving fire services using spatio-temporal analysis: Fire incidents in manila. In: *2017 IEEE Region 10 Symposium (TENSYMP)*. [S.l.: s.n.], 2017. p. 1–5.
- BAO, J. et al. Recommendations in location-based social networks: a survey. *GeoInformatica*, v. 19, n. 3, p. 525–565, 2015. Disponível em: <<http://dx.doi.org/10.1007/s10707-014-0220-8>>.
- BOLEY, M. et al. One click mining: Interactive local pattern discovery through implicit preference and performance learning. In: ACM. *Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics*. [S.l.], 2013. p. 27–35.
- CHIDEAN, M. I. et al. Spatio-temporal analysis of wind resource in the iberian peninsula with data-coupled clustering. *Renewable and Sustainable Energy Reviews*, v. 81, p. 2684 – 2694, 2018. ISSN 1364-0321. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1364032117310080>>.
- GHAHRAMANI, M.; ZHOU, M.; HON, C. T. Spatio-temporal analysis of mobile phone data for interaction recognition. In: *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. [S.l.: s.n.], 2018. p. 1–6.
- KAMATH, K. Y.; CAVERLEE, J. Spatio-temporal meme prediction: Learning what hashtags will be popular where. In: *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*. New York, NY, USA: ACM, 2013. (CIKM '13), p. 1341–1350. ISBN 978-1-4503-2263-8. Disponível em: <<http://doi.acm.org/10.1145/2505515.2505579>>.
- KUMAR, H.; KAUR, H. Clustering and ranking social media users based on temporal analysis. In: *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*. [S.l.: s.n.], 2017. p. 271–275.
- LEVANDOSKI, J. J. et al. Lars: A location-aware recommender system. In: *ICDE*. [s.n.], 2012. p. 450–461. ISBN 978-0-7695-4747-3. Disponível em: <<http://dx.doi.org/10.1109/ICDE.2012.54>>.
- LIANG, J.; HUANG, M. L. Highlighting in information visualization: A survey. In: *2010 14th International Conference Information Visualisation*. [S.l.: s.n.], 2010. ISSN 1550-6037.

- LINS, L.; KŁOSOWSKI, J. T.; SCHEIDEGGER, C. Nanocubes for real-time exploration of spatiotemporal datasets. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 19, n. 12, p. 2456–2465, 2013.
- LOPES-TEIXEIRA, D.; BATISTA, F.; RIBEIRO, R. Spatio-temporal analysis of brand interest using social networks. In: *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*. [S.l.: s.n.], 2018. p. 1–6.
- MA, J. W. et al. Spatio-temporal factor analysis of characterizing mass protest events using taxi trajectory in seoul, korea. In: *Proceedings of the 1st ACM SIGSPATIAL Workshop on Analytics for Local Events and News*. New York, NY, USA: ACM, 2017. (LENS'17), p. 6:1–6:7. ISBN 978-1-4503-5500-1. Disponível em: <<http://doi.acm.org/10.1145/3148044.3148050>>.
- MIJOVIĆ, V. et al. Exploratory spatio-temporal analysis of linked statistical data. *Web Semantics: Science, Services and Agents on the World Wide Web*, v. 41, p. 1 – 8, 2016. ISSN 1570-8268. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1570826816300488>>.
- NARA, A.; TORRENS, P. M. Spatial and temporal analysis of pedestrian egress behavior and efficiency. In: *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems*. New York, NY, USA: ACM, 2007. (GIS '07), p. 59:1–59:4. ISBN 978-1-59593-914-2. Disponível em: <<http://doi.acm.org/10.1145/1341012.1341083>>.
- OMIDVAR-TEHRANI, B. et al. Geoguide: An interactive guidance approach for spatial data. In: *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Exeter, United Kingdom, June 21-23, 2017*. [s.n.], 2017. p. 1112–1117. Disponível em: <<https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.170>>.
- ROBINSON, A. C. Highlighting in geovisualization. *Cartography and Geographic Information Science*, v. 38, n. 4, p. 373–383, 2011. Disponível em: <<http://dx.doi.org/10.1559/15230406384373>>.
- RODDICK, J. F. et al. Spatial, temporal and spatio-temporal databases - hot issues and directions for phd research. *SIGMOD Record*, v. 33, n. 2, p. 126–131, 2004. Disponível em: <<http://doi.acm.org/10.1145/1024694.1024724>>.
- TELANG, A.; PADMANABHAN, D.; DESHPANDE, P. Spatio-temporal indexing: Current scenario, challenges and approaches. In: *Proceedings of the 18th International Conference on Management of Data*. Mumbai, India, India: Computer Society of India, 2012. (COMAD '12), p. 9–11. Disponível em: <<http://dl.acm.org/citation.cfm?id=2694443.2694449>>.
- TOMOKI, N.; KEIJI, Y. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, v. 14, n. 3, p. 223–239. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2010.01194.x>>.

- WILLETT, W.; HEER, J.; AGRAWALA, M. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 13, n. 6, p. 1129–1136, 2007.
- WONGSUPHASAWAT, K. et al. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *TVCG*, IEEE, v. 22, n. 1, 2016.
- ZHAN, X. et al. Spatial-temporal analysis on bird habitat discovery in china. In: *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. [S.l.: s.n.], 2017. p. 573–578.
- ZHENG, L. et al. Spatial-temporal travel pattern mining using massive taxi trajectory data. *Physica A: Statistical Mechanics and its Applications*, v. 501, p. 24 – 41, 2018. ISSN 0378-4371. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378437118301419>>.