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 Tecnologo 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 impessoal 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 vernícula 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 linguas (a exemplo de espanhol ou francês) para redação do resumo em língua estrangeira.

Keywords: Keyword 1, Keyword 2, Keyword 3.

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

Sumário

1 Introduction

More than ever we are overwhelmed by the amount of information has been created. When we compare how much data has been created over the past years, we realize it is already increasing exponentially. Besides this quantitative evolution, nowadays we have the most diverse kinds of data (e.g. documents, tweets, pictures, videos, GIF, check-ins).

This phenomenon has been called *Big Data* and represents an increasing field of study for the time being. Therefore researchers all around the world are analyzing and learning with these information we create everyday. However the increasing amount of data is making analyses a way more difficult. So we are investing in techniques and tools to tackle problems such as data mining, data cleaning, data visualization, data classification, data exploration and so on.

Right in the middle of this scenario we may find data that comes along with latitude and longitude (tweets and check-ins are good examples). We categorize those data as *spatial data*. 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.

1.1 Problem Definition

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. 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 this work, we formulate the problem of "information highlighting using implicit feedback collected over time", i.e., highlight few spatial points based on implicit interests of the analysts in order to guide her towards what she should concentrate on in consecutive iterations of the analysis process.

1.1.1 Case Study

We discuss a real-world example to show the functionality of our approach in practice.

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 website*[1]*. 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 discovering IDRs 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 temporal analyses can be effectively applied in the highlighting of interesting information based on previous experience in the next sections.

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...
- Elaborate how temporal analyses can be effectively applied in data exploration...
- Present the results of our guidance approach...

1.2.2 Specific Objectives

- Present our proposed guidance approach...
- Describe our data model used for temporal analyses...

[1] *http://www.airbnb.com*

- Describe our concept of IDR used for collecting feedback...
- TODO...

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; KLOSOWSKI; 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., Tableau[1], Exhibit[2], Spotfire[3]) 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 (LIANG; 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 solutions for applying temporal analysis...

[1] *http://www.tableau.com*
[2] *http://www.simile-widgets.org/exhibit/*
[3] *http://spotfire.tibco.com*

2.2 GeoGuide

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

GeoGuide used both explicit and implicit feedback. Explicit feedback is when the user are 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 movement, 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

We consider two layers: spatial layer and feedback layer.

In the 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 ($dom(p)$). For instance,

In the feedback layer: we have IDRs per iteration/session where implicid 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 $dom(p)$, 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 $dom(p)$, we evaluate the most used terms in order to create a word rank (KUMAR; KAUR, 2017).
- A profile has a map M between the $< name, value >$ of categoricals attributes and its relevance in $dom(p)$.
- 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

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.

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.; KLOSOWSKI, 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.

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>.

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.