

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 for Highlighting Interesting Data over Time

Felipe Mateus Freire Pontes

Natal-RN

Mês (por extenso) e ano

Felipe Mateus Freire Pontes

An Spatial-Temporal Model for Highlighting Interesting Data 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 for Highlighting Interesting Data 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:

Nome completo do orientador e titulação

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 for Highlighting Interesting Data 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 for Highlighting Interesting Data 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	p. 13
1.1	Problem Definition	p. 13
1.1.1	Case Study	p. 14
1.2	Goals	p. 14
1.2.1	Specific goals	p. 14
1.3	Organization	p. 15
2	Background	p. 16
2.1	Related Work	p. 16
2.2	GeoGuide	p. 16
3	Data Model Definition	p. 17
4	Collecting feedback	p. 18
5	Applying temporal analysis	p. 19
6	Guiding the user	p. 20
7	Experiments	p. 21
7.1	Results	p. 21
8	Conclusion	p. 22
8.1	Contributions	p. 22
8.2	Restrictions	p. 22

8.3	Future work	p. 22
	Referências	p. 23

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. *George is a software engineer with a very good taste for food. He likes to go to restaurants every month, try new restaurants and he already knows almost every place in his city, Parnamirim. So now George is looking for places in the nearby city, Natal. Before GeoGuide, George founds himself losing a lot of time choosing restaurant instead of actually going to them. Another interesting fact about him is he changes his preferences with the seasons, therefore it represent another challenge for him during the Yelp exploration. For example, in the winter George prefers places close to his home or his work and in the summer, he prefers places which offers sea food. By using GeoGuide during the past year to get to know almost every place in his city, when George starts to looking for places in the nearby city, GeoGuide already understands the behavior of how George's preferences changes over the seasons and highlight suggestions of places most likely to be interesting for him according to the moment of the year.*

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 Goals

We iam for the this work to be...

- Super duper goal 1...
- Awesome goal 2...

1.2.1 Specific goals

List of specific goals of this work:...

- Something 1...
- Another 2...

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

2.1 Related Work

There are currently several solutions for applying temporal analysis...

2.2 GeoGuide

GeoGuide is an awesome solution for exploring spatial data...

GeoGuide is about making easy to the researcher explore, visualize and learn with spatial datasets. In order to accomplish a better user experience, we provide a guidance approach based on his preferences and explicit feedback. GeoGuide continuously improves its guidance algorithm while researcher is exploring the dataset. The framework captures implicit feedback and try to understand what the user wants by analysing how the researcher preferences has evolved 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 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 $dom(p)$, 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 $\langle name, value \rangle$ of categorical 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

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.