

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO RIO GRANDE DO NORTE

FELIPE MATEUS FREIRE PONTES

**UM MODELO ESPAÇO-TEMPORAL PARA EXPLORAR REGIÕES DENSAS
INTERESSANTES**

NATAL

2018

FELIPE MATEUS FREIRE PONTES

**UM MODELO ESPAÇO-TEMPORAL PARA EXPLORAR REGIÕES DENSAS
INTERESSANTES**

Trabalho de conclusão de curso apresentado ao Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas da Diretoria Acadêmica de Gestão e Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte como requisito parcial à obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Dr. Plácido Antônio da Silva Neto

NATAL
2018

FELIPE MATEUS FREIRE PONTES

UM MODELO ESPAÇO-TEMPORAL PARA EXPLORAR REGIÕES DENSAS
INTERESSANTES

Trabalho de conclusão de curso apresentado ao Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas da Diretoria Acadêmica de Gestão e Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte como requisito parcial à obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Trabalho de Conclusão de Curso apresentado e aprovado em ____/____/____, pela seguinte Banca Examinadora:

BANCA EXAMINADORA

Plácido Antônio da Silva Neto, D.r – Presidente
Instituto Federal de Educação, Ciência e Tecnologia do Rio
Grande do Norte

Jorgiano Márcio Bruno Vidal, D.r – Examinador
Instituto Federal de Educação, Ciência e Tecnologia do Rio
Grande do Norte

Marília Aranha Freire, D.ra – Examinador
Instituto Federal de Educação, Ciência e Tecnologia do Rio
Grande do Norte

NATAL
2018

Aos meus pais que nunca duvidaram de mim.

AGRADECIMENTOS

Ao meu orientador Plácido Antônio da Silva Neto, pelo seu empenho em sempre buscar extraír o melhor de mim.

A Behrooz Omidvar-Tehrani, pelas suas colaborações mesmo estando a um oceano de distância.

A Francisco Bento da Silva Júnior, pelas suas contribuições como amigo e colega de pesquisa.

A coisa mais autêntica sobre nós é nossa capacidade de criar, de superar, de suportar, de transformar, de amar e de sermos maiores que nosso sofrimento.

Ben Okri

RESUMO

Este trabalho propõe um modelo espaço-temporal para exploração de preferências do usuário no cenário da análise exploratória de dados espaciais. A partir da observação das dificuldades enfrentadas por analistas durante a análise exploratória de grandes conjuntos de dados, verificou-se a possibilidade de combinar métodos de exploração de preferências do usuário com análise temporal, o que resultou no modelo de dados proposto neste trabalho. Para tanto, aprimorou-se a ferramenta GeoGuide, um ambiente de exploração de dados espaciais alinhado com métodos de destacamento de informações com base nas preferências do usuário, para coletar dados não somente no contexto espacial, mas também no contexto de domínio. Este trabalho detalha o processo para capturar de maneira transparente as preferências do usuário com o objetivo de viabilizar a aplicação do modelo proposto. Também apresenta como os dados coletados podem ser analisados temporalmente a fim de encontrar padrões nas preferências do usuário. Por fim, espera-se que o modelo proposto seja explorado, adaptado e aplicado além do cenário apresentado.

Palavras-chave: Dados Espaciais. Análise Temporal. Preferência do Usuário.

ABSTRACT

This work proposes a spatial-temporal model to exploit user preferences in the scenario of an exploratory analysis of spatial data. From the observation of the difficulties faced by analysts during the exploratory analysis of large datasets, the possibility of combining methods of exploring user preferences with time analysis was verified, which resulted in the data model proposed in this work. Therefore, the GeoGuide tool, a spatial data exploration environment which provides a guidance approach using information highlighting methods based on user preferences, has been improved to collect data not only in the spatial context but also in the domain context. This work details the process to transparently capture user preferences with the objective of making the proposed model feasible. It also displays how the data collected can be analyzed temporally in order to find patterns in user preferences. With this work, it is expected that the proposed model will be explored, adapted and applied beyond the presented scenario.

Keywords: Spatial data. Temporal Analysis. User Feedback.

LISTA DE ILUSTRAÇÕES

Figura 1 – Processo de explorar estadias em Paris	24
Figura 2 – Componentes do GeoGuide	29
Figura 3 – Exemplo de pontos geográficos com seus atributos	30
Quadro 1 – Algoritmo para Descobrimento de IDR _s	34
Figura 4 – Processo de descoberta de IDR _s	35
Figura 5 – Evolução no contexto espacial	38
Figura 6 – Evolução de atributo numérico	39
Figura 7 – Evolução de atributo temporal	42

LISTA DE TABELAS

Tabela 1 – Exemplo de Índice de Relevância e Diversidade para pontos na Figura 3	30
Tabela 2 – Exemplo de atributos de um ponto que representa uma estadia	35
Tabela 3 – Exemplo de perfil dos atributos numéricos de uma IDR	36
Tabela 4 – Exemplo parcial de perfil do atributo textual de uma IDR	37
Tabela 5 – Exemplo de perfil dos atributos categóricos de uma IDR	37
Tabela 6 – Exemplo de perfil de atributos categóricos	40
Tabela 7 – Atualizando vetor de feedback	41
Tabela 8 – Exemplo parcial de perfil do atributo textual	41

LISTA DE ABREVIATURAS E SIGLAS

IDR	<i>Interesting Dense Region</i>
IFRN	Instituto Federal do Rio Grande do Norte
TADS	Tecnologia em Análise e Desenvolvimento de Sistemas

SUMÁRIO

1	INTRODUÇÃO	23
1.1	PROBLEMA	23
1.1.1	Estudo de Caso	24
1.2	OBJETIVOS	25
1.2.1	Objetivos Gerais	25
1.2.2	Objetivos Específicos	26
1.3	ORGANIZAÇÃO	26
2	CONTEXTUALIZAÇÃO	27
2.1	TRABALHOS RELACIONADOS	27
2.1.1	Exploração de Feedback	27
2.1.2	Métodos de Destacamento de Informação	28
2.1.3	Aplicações de Análise Temporal	28
2.2	GEOGUIDE	29
2.2.1	Pré-processamento	29
2.2.1.1	Relevância e Diversidade	29
2.2.2	Preferências do Usuário	31
2.2.3	Destacamento de Dados Espaciais	31
3	MODELO ESPAÇO-TEMPORAL APLICADO AO GEOGUIDE	33
3.1	REGIÕES DENSAS INTERESSANTES	33
3.1.1	Perfil	35
3.2	ANÁLISE DE VARIAÇÃO TEMPORAL	37
3.2.1	Contexto Espacial	38
3.2.1.1	Deslocamento de IDR	38
3.2.1.2	Distância entre IDRs	38
3.2.1.3	Área de IDR	39
3.2.2	Contexto de Domínio	39
4	CONSIDERAÇÕES FINAIS	43
4.1	CONTRIBUIÇÕES	43
4.2	TRABALHOS FUTUROS	43
	REFERÊNCIAS	45

1 INTRODUÇÃO

Mais do que nunca observa-se o quanto sobrecarregadas as pessoas estão com a quantidade de dados que criam a cada dia (PRADEEP; KALLIMANI, 2017). Quando compara-se quanto de informação vem sendo gerada nos últimos anos, percebe-se que está aumentando significamente. Além dessa evolução quantitativa, hoje, encontram-se os mais diversos tipos de informações, por exemplo: documentos, tuítes, fotos, vídeos, *GIFs*, *check-ins* entre vários outros.

Esse fenômeno vem sendo chamado de *Big Data* e representa uma crescente área de estudo atualmente. Como consequência, pesquisadores estão analisando e aprendendo com essas informações geradas, entretanto o crescimento contínuo da quantidade de dados dificulta as análises. Portanto pessoas estão investindo em novas técnicas e ferramentas para romper desafios como mineração de dados, *data cleaning*, visualização de dados, classificação de dados, exploração de dados e muito mais (ZHANG et al., 2015).

Um tipo comum de dado é chamado de dado espacial, o qual possui atributos geográficos como latitude e longitude como, por exemplo, tuítes, avaliação de restaurantes, *check-ins* em estabelecimentos. Dados espaciais podem ser muito significativos, por exemplo, um *check-in* no aeroporto por sua irmã na manhã do seu aniversário, provavelmente significa que você terá uma surpresa.

Cada registro de dados espaciais representa uma atividade em uma precisa localização geográfica, em outras palavras, a análise desse tipo de dado permite realizar descobertas baseadas em fatos. Analistas estão frequentemente interessados em observar padrões espaciais e tendências para melhorar seus processos de tomada de decisão. Análise de dados espaciais tem várias aplicações como gerenciamento de cidades inteligentes, gerenciamento de desastres e transporte autônomo (RODDICK et al., 2004; TELANG; PADMANABHAN; DESHPANDE, 2012).

1.1 PROBLEMA

A análise de dados espaciais geralmente é realizada num contexto exploratório: o analista não tem uma consulta precisa em mente e ele explora os dados em passos iterativos a fim de encontrar resultados potencialmente interessantes. Tradicionalmente, um cenário de análise exploratória é descrito na seguinte maneira: o analista visualiza um subconjunto de dados usando uma consulta em ambiente de visualização (por exemplo: Tableau¹, Exhibit², Spotfire³); o resultado será ilustrado em um mapa geográfico; então o analista investiga diferentes partes do conjunto de dados movendo ou focando uma região do mapa a fim de encontrar padrões ou tendências de interesse. O analista pode iterar por esse processo várias vezes realizando consultas diferentes e focando em diferentes aspectos.

Contudo, o vasto tamanho do conjunto de dados espaciais faz com que o analista se sinta

¹ <http://www.tableau.com>

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

³ <http://spotfire.tibco.com>

Figura 1 – Processo de explorar estadias em Paris



Fonte: Elaboração própria em 2018

perdido durante a exploração. É possível ter milhares de pontos geográficos em cada bairro de uma cidade, por exemplo. Analistas precisam ter acesso apenas a algumas opções, chamadas de “*highlights*”, que ajam como uma direção e assim permitam que ele foque no que lhe interessa na análise. No cenário perfeito, essas opções não são aleatoriamente escolhidas e representam o que o analista se mostrou interessado em iterações passadas.

Este trabalho formula um modelo para permitir o “realçamento de dados usando feedback coletado ao longo do tempo”. Em outras palavras, busca-se realçar alguns pontos geográficos baseado nos interesses do analista com o intuito de guiá-lo na direção ao que ele deve se concentrar nas iterações seguintes do processo de análise.

1.1.1 Estudo de Caso

Nesta subseção, apresenta-se um estudo de caso com o objetivo de demonstrar a funcionalidade da abordagem proposta na prática.

Exemplo: Benício está planejando passar alguns dias em Paris, França. Sua apreciação pela cultura francesa faz com que ele tenha interesse em novas experiências na cidade. Ele decidiu por alugar uma estadia pelo Airbnb⁴. Ele gosta de descobrir a cidade, portanto ele é aberto a qualquer tipo de estadia em qualquer região com um leve interesse em ficar perto do centro da cidade. O sistema retorna 4000 opções diferentes. Como ele não tem outras

⁴ <http://www.airbnb.com>

preferências, é necessário uma investigação exaustiva para avaliar cada região da cidade independentemente, o que é quase impossível. Enquanto estava avaliando algumas opções, ele demonstrou interesse na região de “Champ de Mars” (próximo à Torre Eiffel), mas esqueceu ou não achou necessário clicar num ponto nesta região. Coletando o feedback do seus movimentos com o mouse no mapa de estadias em Paris, a abordagem proposta consegue de maneira transparente detectar o interesse dele na região supracitada e apresentar uma quantidade pequena de opções recomendadas para Benício.

O exemplo acima é usado para descrever como feedback implícito é coletado na prática. Imagem 1 mostra os passos de Benício para explorar estadias em Paris. Imagem 1.A mostra os movimentos do mouse dele em diferentes intervalos de tempo. Nesse exemplo, o feedback de Benício é coletado em 3 diferentes intervalos de tempo (evoluindo das Imagens 1.B até 1.D). Isso mostra que Benício começou sua busca perto da Torre Eiffel e *Arc de Triomphe* (Imagen 1.B) e gradualmente mostrou também interesse no sul (Imagen 1.C) e norte (Imagen 1.D). Todas as interseções entre essas regiões são descobertas (regiões tachadas na Imagem 1.E), o que representa um conjunto de regiões onde o interesse de Benício está direcionado e onde, provavelmente, ele vai decidir ficar durante sua visita à Paris.

E se Benício quiser voltar para Paris próximo ano? Ele teria que repetir a mesma análise exploratória, a não ser que ele lembre a localização exata das estadias interessantes a ele no ano passado. Na abordagem proposta, ele não precisaria lembrar, porque suas preferências foram coletadas e poderiam ser usadas para realçar um subconjunto similar ao do ano anterior.

No contexto da análise exploratória, o analista talvez mude suas preferências entre as sessões (por exemplo, no inverno, Benício talvez queira ficar próximo à Torre Eiffel, mas no verão, ele talvez queira ficar mais distante dos pontos turísticos). A fim de enfrentar esse desafio, é proposto um modelo que permite uma análise de variação temporal para identificar padrões em como as preferências dos analistas mudam entre as sessões, o que permite aprimorar o método de realçamento para ser mais preciso e consistente com o interesse do analista mesmo em momentos diferentes do ano, por exemplo.

1.2 OBJETIVOS

Nesta seção, define-se os objetivos gerais e específicos do trabalho.

1.2.1 Objetivos Gerais

- a) Propor um modelo de análise espaço-temporal para orientação na exploração de dados espaciais;
- b) Elaborar como análise temporal pode ser efetivamente aplicada na exploração de dados espaciais.

1.2.2 Objetivos Específicos

- a) Descrever o conceito de Regiões Densas Interessantes usado para captura de feedback;
- b) Apresentar como o modelo de dados pode ser aplicado no contexto da análise exploratória;
- c) Descrever o modelo de dados usado para análise temporal;
- d) Investigar como o modelo proposto pode ser explorado no contexto espacial e no contexto de domínio.

1.3 ORGANIZAÇÃO

Os próximos capítulos estão organizados na seguinte maneira: no Capítulo 2 é discutido o estado da arte por trás desse trabalho; Capítulo 3 define o modelo de dados, apresenta como é feito a coleta de feedback durante a análise exploratória e demonstra como a análise de variação temporal pode ser aplicada. Capítulo 4 conclui e propõe futuros trabalhos.

2 CONTEXTUALIZAÇÃO

Este capítulo dá uma visão geral sobre os trabalhos relacionados acerca de exploração de feedback, métodos de destaqueamento de informações e aplicação de análises temporais. Também é apresentado o sistema que está sendo estendido neste trabalho.

2.1 TRABALHOS RELACIONADOS

A literatura na análise de dados espaciais possui um foco na eficiência das iterações exploratórias. A abordagem comum é projetar índices pré-processados, os quais permitem a consulta eficiente de dados espaciais (LINS; KLOSOWSKI; SCHEIDEGGER, 2013). No entanto, também é preciso direcionar a atenção para o *valor* dos dados espaciais, porque é muito comum encontrar um analista se perdendo numa enorme quantidade de pontos geográficos. Para solucionar esse problema, ambientes de visualização, como, por exemplo, Tableau¹, Exhibit², Spotfire³, oferecem funcionalidades para manipular os dados como filtros, consultas agregadoras, entre outras. Entretanto essas funcionalidades não se mostram eficazes, visto que nessas ferramentas o pesquisador precisa saber exatamente o que procura. Este trabalho combina a exploração de feedback, métodos de destaqueamento de informações e análise temporal a fim de otimizar a análise exploratória.

2.1.1 Exploração de Feedback

O modelo espaço-temporal proposto aprimora o processo de análise de dados espaciais destacando subconjuntos de pontos geográficos com base no feedback coletado durante a exploração do analista. Na literatura, há vários trabalhos sobre exploração de feedback para orientar o analista nas futuras iterações da análise como, por exemplo, Boley et al. (2013). A abordagem comum é uma metodologia *top-k* para reduzir o escopo da consulta baseado no feedback explícito e recomendar um pequeno subconjunto de resultados interessantes de tamanho k . Uma clara distinção do presente trabalho é que não busca-se reduzir o escopo, mas alavancar o conjunto de dados com resultados potencialmente interessantes que o analista talvez não tenha notado devido ao enorme volume de dados espaciais. Enquanto as escolhas do analista são limitadas por k em algoritmos de *top-k* processamento, a abordagem proposta oferece a liberdade de escolha ao mesmo tempo que pontos geográficos vão sendo transparentemente destacadados com base nas novas escolhas do analista.

¹ <http://www.tableau.com>

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

³ <http://spotfire.tibco.com>

2.1.2 Métodos de Destacamento de Informação

Há trabalhos na literatura sobre métodos de destacamento de informações, por exemplo: Liang e Huang (2010), Robinson (2011), Wongsuphasawat et al. (2016), Willett, Heer e Agrawala (2007). Entretanto todos esses métodos são *objetivos* e não são aplicáveis para o contexto de orientação espacial onde o feedback do usuário é envolvido. Em termos de recomendação, algumas abordagens focam na dimensão espacial (BAO et al., 2015; LEVANDOSKI et al., 2012) enquanto o contexto e a diversidade do resultado é deixado de lado.

2.1.3 Aplicações de Análise Temporal

Existem várias instâncias na literatura que combinam análise temporal com dados espaciais, como Baculo et al. (2017), Balahadia e Trillanes (2017), Chidean et al. (2018), Ghahramani, Zhou e Hon (2018), Kamath e Caverlee (2013), Lopes-Teixeira, Batista e Ribeiro (2018), Ma et al. (2017), Mijović et al. (2016), Tomoki e Keiji (2010), Nara e Torrens (2007), Zhan et al. (2017), Zheng et al. (2018). Essas aplicações de análise temporal são em contextos específicos, os quais não envolvem feedback do usuário, mas representam como análise temporal pode ser perspicaz e proveitosa.

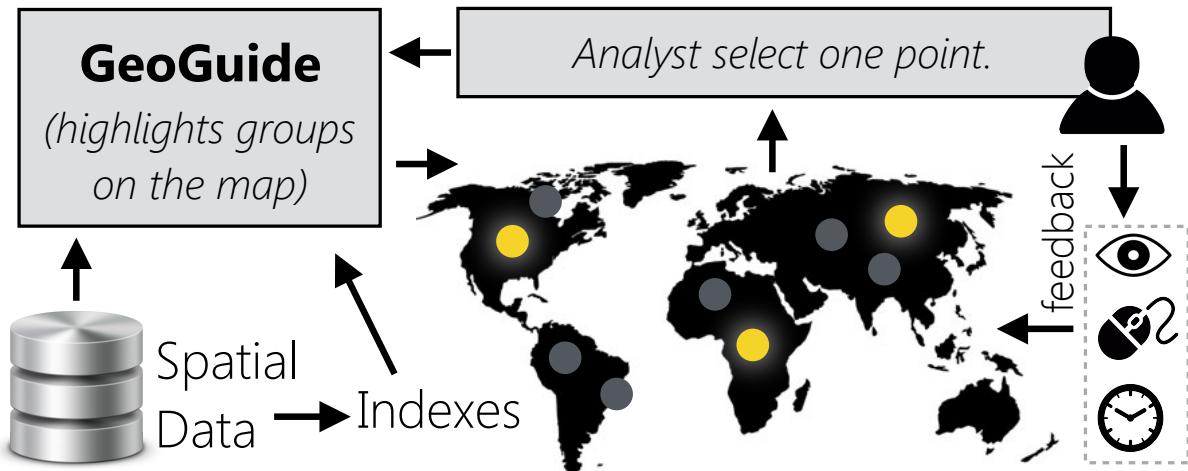
Baculo et al. (2017) e Balahadia e Trillanes (2017) fazem uso de dados públicos de Manila, capital das Filipinas, combinando dados espaciais, análise temporal e modelos preditivos e mostrando resultados que podem ser utilizados para preparação de um plano de gestão pública eficaz. Ma et al. (2017) e Zheng et al. (2018) também fazem análises realistas de como eventos, como protestos, impactam nas trajetórias de táxis, cujos resultados podem auxiliar no controle de tráfego urbano e nos planos de serviços de transporte da cidade. Ambos realizam ricas análises, as quais contribuíram como inspiração neste trabalho.

Chidean et al. (2018) apresenta como detectar padrões espaço-temporais no contexto do uso de energia eólica na Península Ibérica usando o algoritmo *Second-Order Data-Coupled Clustering*. Apesar do estudo detalhado, esse trabalho não contempla um contexto de análise exploratória.

Ghahramani, Zhou e Hon (2018), Lopes-Teixeira, Batista e Ribeiro (2018) e Zhan et al. (2017) demonstram como análise temporal pode ser aplicada no contexto geográfico. Zhan et al. (2017) vai além gerando uma árvore de clusterização hierárquica. Apesar dos métodos e resultados serem bem detalhados nos trabalhos, essas contribuições não se aplicam ao assunto em questão.

Kamath e Caverlee (2013) propõe uma abordagem de *reinforcement learning* para prever eventos (adoção de *memes*) num contexto espaço-temporal. Nara e Torrens (2007) introduz um modelo de visualização 3D para dados espaço-temporais que ajuda a analisar qualitativa e quantitativamente os padrões e tendências espaço-temporais. Ambos os trabalhos contribuem para representar como o modelo proposto pode ser combinado com diversas técnicas.

Figura 2 – Componentes do GeoGuide



Fonte: Omidvar-Tehrani et al. (2017)

2.2 GEOGUIDE

GeoGuide (OMIDVAR-TEHRANI et al., 2017) é fruto de um projeto de pesquisa realizado por alunos do curso de Tecnologia em Análise e Desenvolvimento de Sistemas (TADS) no Instituto Federal do Rio Grande do Norte (IFRN) em colaboração com a Universidade de Grenoble. Esse projeto se trata de um ambiente de visualização de dados espaciais que coleta as preferências do usuário durante a exploração para destacar subconjuntos de pontos geográficos que podem ser interessantes ao analista. Figura 2 ilustra os principais componentes da arquitetura do GeoGuide explorados nas próximas subseções.

Neste trabalho, o GeoGuide é potencializado à dois novos conceitos: *i.* regiões densas interessantes e *ii.* análise temporal das preferências do usuário. Esses dois conceitos serão explorados no próximo capítulo.

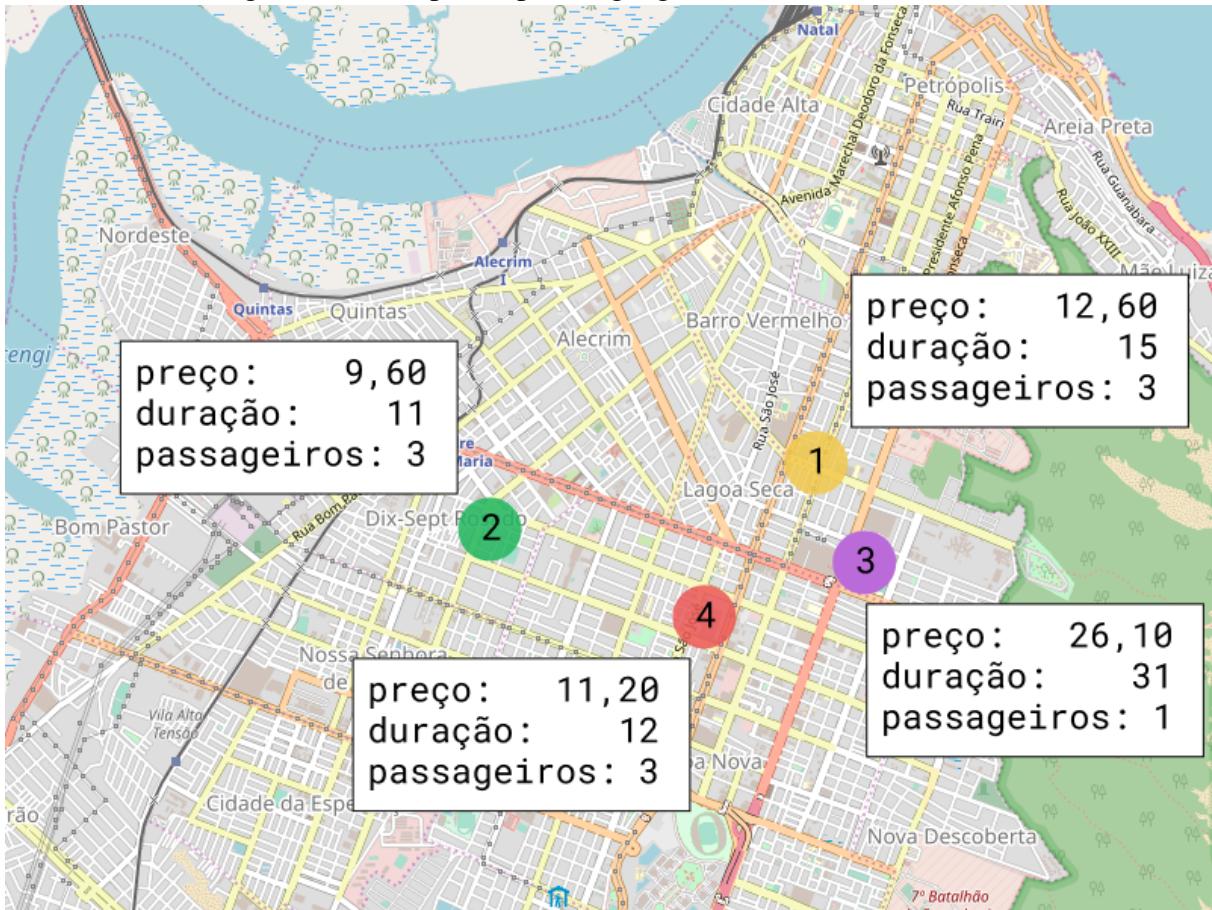
2.2.1 Pré-processamento

GeoGuide realiza um passo de pré-processamento para criar os índices que serão usados durante a fase de destacamento. O índice é uma tabela comparativa entre todos os pontos usando duas métricas de qualidade: relevância e diversidade. Os valores calculados são normalizados no intervalo de 0.0 à 1.0.

2.2.1.1 Relevância e Diversidade

Relevância representa o quanto similar é o ponto *a* com o ponto *b* num conjunto de dados. GeoGuide usa a relevância para destacar pontos similares ao feedback do analista. Diversidade representa quanto distante o ponto *a* está localizado do ponto *b*. GeoGuide usa a diversidade para permitir ao analista explorar diferentes regiões, mas ainda assim trabalhar com pontos relevantes ao seu interesse.

Figura 3 – Exemplo de pontos geográficos com seus atributos



Fonte: Elaboração própria em 2018

Tabela 1 – Exemplo de Índice de Relevância e Diversidade para pontos na Figura 3

Ponto A	Ponto B	Relevância	Diversidade
1	2	0.8	0.9
1	3	0.25	0.2
1	4	0.9	0.45
2	3	0.2	1.0
2	4	1.0	0.48
3	4	0.2	0.3

Fonte: Elaboração própria em 2018

Na Figura 3, tem-se, por exemplo, pontos geográficos que representam viagens de táxi. Cada ponto (1, 2, 3 e 4), possui seus atributos: preço da viagem, duração da viagem e a quantidade de passageiros; e sua localização geográfica. Na tabela 1, ilustra-se o exemplo do índice calculado diante dos pontos apresentados na Figura 3. Percebe-se que os pontos mais similares entre si são os pontos 2 e 4, enquanto que 2 e 3 são os mais distantes, ou seja, possuem o maior valor de diversidade.

2.2.2 Preferências do Usuário

Para coletar as preferências do usuário, GeoGuide usa ambos feedback implícito e explícito. Feedback explícito é quando o usuário está analisando os atributos de um ponto, por exemplo a descrição de uma casa no Airbnb, e explicitamente pede para explorar pontos similares ao selecionado. Feedback implícito é coletado através da captura dos movimentos do mouse e métricas como “quanto tempo o usuário passou analisando o perfil de um ponto”.

2.2.3 Destacamento de Dados Espaciais

GeoGuide combina o índice pré-processado e o feedback coletado para destacar subconjuntos de dados espaciais de acordo com as preferências do analista. O processo de destacamento provou ser eficiente em termos de “quantos passos o analista leva até completar a tarefa de encontrar um ponto com um determinado perfil”. Usando GeoGuide, analistas foram capazes de completar a tarefa usando, em média, 10.7 passos, enquanto que usando Tableau, foram necessários 43 passos (OMIDVAR-TEHRANI et al., 2017).

3 MODELO ESPAÇO-TEMPORAL APLICADO AO GEOGUIDE

Neste capítulo, explica-se o que são as chamadas Regiões Densas Interessantes e define-se o modelo espaço-temporal e sua aplicabilidade.

3.1 REGIÕES DENSAS INTERESSANTES

Uma Região Densa Interessante (do inglês *Interesting Dense Region*, IDR) é uma região espacial com uma alta probabilidade de conter pontos de interesse do analista. IDRs são coletadas e definidas durante o processamento do feedback do usuário. Diferentemente da literatura que predominantemente foca em interações explícitas, como clicar no botão, esta proposta investiga o feedback implícito.

Durante a exploração iterativa de dados espaciais, é comum o caso que o analista avalia algumas regiões de interesse, mas esquece de dar um feedback explícito sobre aquela região. O ato do usuário olhar para essa região pode ser capturado através do rastreio dos movimentos oculares e, como Arapakis et al. (2014) comprova, esse método possui uma forte relação com a atenção do usuário.

Entretanto, o rastreamento dos movimentos oculares fere várias questões de privacidade. Assim sendo, foi adotado a alternativa de rastrear os deslocamentos do cursor do mouse. Arapakis, Lalmas e Valkanas (2014) argumenta que esse método possui uma forte relação com o engajamento do usuário. Intuitivamente, um ponto espacial recebe um feedback positivo se o cursor do mouse se desloca próximo a ele frequentemente.

O objetivo de descobrir IDRs é para obter as preferências do analista que nunca foram expressadas explicitamente. Para tal, leva-se em consideração dois conceitos:

- **Conceito 1:** uma região é mais interessante ao analista, se for densa, ou seja, o analista movimenta seu mouse naquela região várias vezes.
- **Conceito 2:** é possível que o analista movimente seu mouse em qualquer lugar do mapa. Isso não deve significar que todo lugar no mapa tem a mesma relevância.

Com base nesses conceitos, o processo de descobrimento das IDRs pode ser descrito pelo Algoritmo 1. Pontos são adicionados em \mathcal{M} a cada $200ms$ para evitar pontos redundantes. De acordo com Conceito 1 e com o propósito de identificar comportamentos recorrentes, o algoritmo começa particionando \mathcal{M} em g segmentos consecutivos $\mathcal{M}_1, \dots, \mathcal{M}_g$. O primeiro segmento começa no momento zero (quando a iteração iniciou) e o último termina em t_c , ou seja, no tempo atual que o algoritmo é executado. De acordo com o Conceito 2, formam-se os clusters em cada segmento de \mathcal{M} usando uma variação da abordagem DBSCAN (ESTER et al., 1996). Por fim, as interseções desses clusters são encontradas e retornadas como IDRs.

Para clusterizar os pontos em cada segmento de tempo (linha 5 do algoritmo apresentado no Quadro 1), é utilizado o ST-DBSCAN, uma variação do DB-SCAN que leva em consideração

Quadro 1 – Algoritmo para Descobrimento de IDRs

Entrada: Tempo atual t_c , pontos dos movimentos do mouse \mathcal{M}
Salida: IDRs encontradas \mathcal{S}
1 $\mathcal{S} \leftarrow \emptyset$
2 $g \leftarrow$ número de segmentos
3 para $i \in [0, g]$ hacer
4 $\mathcal{M}_i \leftarrow \{m = \langle x, y, t \rangle (\frac{t_c}{g} \times i) \leq t \leq (\frac{t_c}{g} \times (i + 1))\}$
5 $\mathcal{C}_i \leftarrow \text{mine_clusters}(\mathcal{M}_i)$
6 $\mathcal{O}_i \leftarrow \text{find_polygons}(\mathcal{C}_i)$
7 para $\mathcal{O}_i, \mathcal{O}_j$ onde $i, j \in [0, g]$ e $i \neq j$ fazer $\mathcal{S}.append(\text{intersect}(\mathcal{O}_i, \mathcal{O}_j))$
8 devolver \mathcal{S}

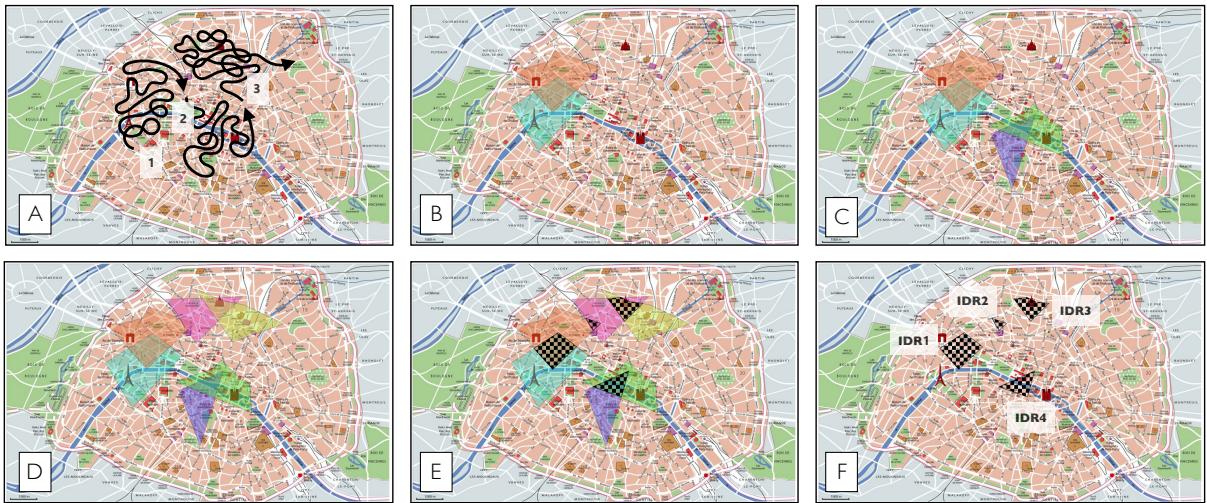
Fonte: Elaboração própria em 2018

atributos espaciais e temporais para clusterizar pontos com base na densidade (BIRANT; KUT, 2007). Para cada subconjunto de pontos dos movimentos do mouse \mathcal{M}_i , $i \in [0, g]$, ST-DBSCAN começa com um ponto aleatório $m_0 \in \mathcal{M}_i$ e coleta todos os pontos alcançáveis (de acordo com a densidade) de m_0 usando uma métrica de distância. Como os pontos dos movimentos do mouse ocorrem em um espaço bidimensional (monitor do computador), a distância euclidiana foi selecionada como métrica. Se m_0 for identificado como objeto *core*, uma cluster será gerado. Caso contrário, se m_0 é uma objeto de borda, nenhum ponto é alcançável de m_0 e o algoritmo buscará outro ponto aleatório em \mathcal{M}_i . Esse processo é repetido até que todos os pontos sejam processados.

Uma vez que os clusters são formados para cada subconjunto de \mathcal{M} , encontra-se suas interseções para determinar as regiões recorrentes. Para obter as interseções é preciso definir nitidamente as fronteiras dos clusters (linha 6). Portanto para cada cluster, é construído seu respectivo polígono que engloba todos os pontos. Para isso, é utilizado o algoritmo Quickhull, um método com abordagem similar ao Quicksort, que calcula o invólucro convexo para um determinado conjunto de pontos num plano bidimensional (BARBER; DOBKIN; HUHDANPAA, 1996).

Esse algoritmo de descoberta de IDR é executado a cada t intervalo de tempo durante uma análise exploratória. Dessa forma, enquanto o analista está explorando o conjunto de dados, suas preferências estão sendo coletadas continuamente. Na Figura 4, observa-se uma iteração de coleta e execução desse algoritmo onde $g = 3$, ou seja, o feedback é coletado em 3 segmentos diferentes de tempo. Os movimentos do mouse realizados pelo analista nesses 3 segmentos são demonstrados na Figura 4.A. Para cada segmento, os pontos coletados dos movimentos do mouse são clusterizados e convertidos em polígonos convexos (Figura 4.B-D, respectivamente). Por fim, os polígonos são sobrepostos (Figura 4.E) e as interseções encontradas são definidas como IDRs (IDR1-4 na Figura 4.F).

Figura 4 – Processo de descoberta de IDR's



Fonte: Elaboração própria em 2018

3.1.1 Perfil

Uma IDR descreve uma região espacial no conjunto de dados do analista coletada num determinado tempo t . Cada região representa as preferências do analista no momento t no contexto espacial, ou seja, *onde* o analista está interessado. Para entender, *em que* o analista está interessado, em outras palavras, as preferências sobre o contexto de domínio, é definido o perfil de cada região.

Tabela 2 – Exemplo de atributos de um ponto que representa uma estadia

id	15
latitude	48.88880
longitude	2.320465
name	Nice apartment in Batignolles
host_name	Daniele
neighbourhood	Batignolles-Monceau
room_type	Entire home/apt
price	65
minimum_nights	1
stars	4
number_of_reviews	25
reviews_per_month	1.74
availability_365	35
last_review	2015-11-14T09:07:02+00:00

Fonte: Elaboração própria em 2018

O perfil é a summarização dos atributos dos pontos contidos numa IDR. Cada ponto em um conjunto de dados ($p \in \mathcal{P}$) é descrito por suas coordenadas espaciais e uma série de atributos de domínio $dom(p)$. A Tabela 2 apresenta um exemplo dos atributos de um ponto no conjunto de dados do Airbnb.

Na Tabela 2, os atributos *latitude* e *longitude* são as coordenadas espaciais, enquanto que todos os outros atributos caracterizam os atributos de domínio. Dentre os atributos de domínio pode-se encontrar 4 tipos de dados:

1. Numéricos: são atributos que representam valores numéricos, como, por exemplo, os atributos *price*, *minimum_nights*, *number_of_reviews*, *reviews_per_month* e *availability_365*.
2. Textuais: são atributos que representam valores textuais, como, por exemplo, os atributos *name* e *host_name*.
3. Categóricos: são atributos que representam valores tanto numéricos quanto textuais, mas que se repetem pouco em todo o conjunto. Por exemplo, o atributo *stars* possui apenas cinco valores em todo o conjunto. Os atributos *room_type* e *neighbourhood* também possuem um número finito de opções, por isso são classificados como categóricos.
4. Temporais: são atributos que representam data e/ou hora de um determinado evento, como, por exemplo, o atributo *last_review* que representa quando foi publicado a última avaliação.

Esses 4 tipos de dados são comumente encontrados em conjuntos de dados espaciais, portanto o perfil da IDR precisa representar de maneira resumida esses atributos em relação aos pontos nela contidos. Assim sendo, cada tipo é tratado de acordo com sua especificidade.

O atributos classificados como numéricos encontrados em uma IDR são summarizados usando funções estatísticas como média, desvio padrão, mínimo e máximo. Observa-se um exemplo do resultado dessa summarização na Tabela 3.

Tabela 3 – Exemplo de perfil dos atributos numéricos de uma IDR

Atributo	Total	Média	σ	Mínimo	Máximo
price	73	106,68	62,62	40	401
minimum_nights	73	2,36	1,60	1	7
number_of_reviews	73	17,86	27,85	0	120
reviews_per_month	55	1,64	1,53	0,08	7,27
calculated_host_listings_count	73	4,05	10,83	1	56
availability_365	73	179,64	154,48	0	365

Fonte: Elaboração própria em 2018

Os atributos textuais são tratados individualmente. Para cada atributo é feito o levantamento dos termos mais usados. Essa abordagem é utilizada por Kumar e Kaur (2017) para encontrar os termos mais frequentes no contexto temporal. A Tabela 4 apresenta os 5 primeiros resultados para o atributo *name* em uma IDR.

Os atributos categóricos são contabilizados num dicionário C onde as chaves são uma tupla com o nome do atributo e o valor da categoria. Quanto maior a quantidade de repetições de uma categoria, mais relevante essa categoria é no contexto de uma IDR. A Tabela 5 mostra o resultado num conjunto de dados para os atributos *neighbourhood*, *room_type* e *stars*.

Tabela 4 – Exemplo parcial de perfil do atributo textual de uma IDR

Atributo: name	
Termo	Ocorrências
champs	29
elysées	18
studio	14
near	13
flat	11

Fonte: Elaboração própria em 2018

Os atributos temporais são normalizados para uma escala linear (a partir do ano novo para datas e a partir da meia noite para horas) objetivando contabilizar as ocorrências. Por exemplo, um atributo com a data *2018-12-21* é normalizado para *2018, 12, 355*, isto é, numa linha temporal esse evento ocorreu no ano 2018, no 12º mês e no 355º dia. As horas são normalizadas de maneira similar, a hora *08:33:22* é normalizada para *8, 513, 30802*, isto é, esse evento ocorreu na 8º hora, 513º minuto e 30802º segundo. Com os valores normalizadas, pode-se encontrar os dias e os momentos do dia mais recorrentes para determinado evento em uma IDR.

3.2 ANÁLISE DE VARIAÇÃO TEMPORAL

Cada IDR é coletada num momento *t* e seu perfil definido através dos atributos de domínio dos pontos contidos na região delimitada. Apesar do perfil ser definido para cada IDR, pode também ser definido para um conjunto de IDRs, ou seja, se num momento *t* houver 3 IDRs, é possível combinar os pontos contidos nessas regiões e definir um perfil para esse momento *t*.

Em outras palavras, é possível realizar a análise das preferências do analista em qualquer espaço de tempo (seja minutos ou dias, semanas ou meses), uma vez que o sistema coleta as preferências do usuário de maneira transparente enquanto o mesmo realiza a análise exploratória do conjunto de dados. Dessa forma, para cada momento *t* em qualquer espaço de tempo, pode-se detectar as preferências do usuário tanto no contexto espacial, quanto no contexto de domínio.

Tabela 5 – Exemplo de perfil dos atributos categóricos de uma IDR

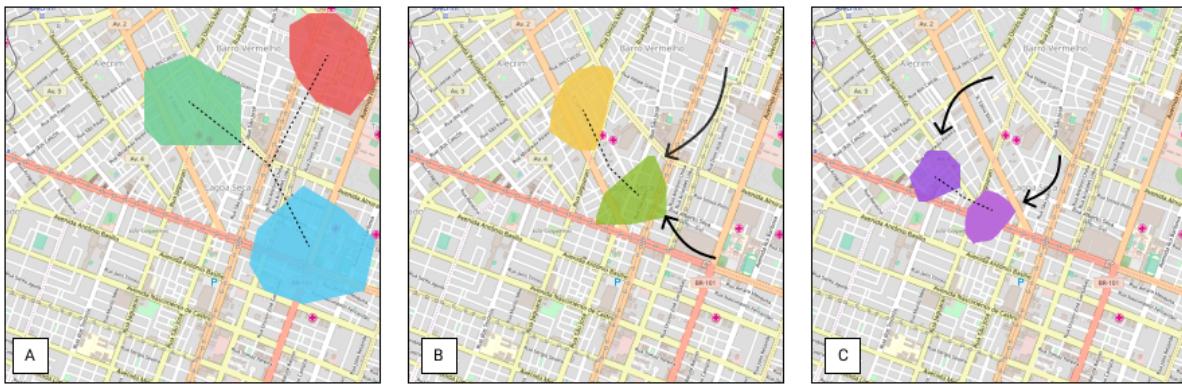
Categoria	Quantidade
<neighbourhood, "Batignolles-Monceau">	73
<room_type, "Entire home/apt">	69
<room_type, "Private room">	4
<stars, 1>	6
<stars, 2>	7
<stars, 3>	29
<stars, 4>	20
<stars, 5>	11

Fonte: Elaboração própria em 2018

3.2.1 Contexto Espacial

O contexto espacial se refere aos aspectos geográficos como, por exemplo, em que parte da cidade o analista se interessa. A Figura 5 apresenta as regiões encontradas em 3 momentos sequenciais, sendo 5.C o mais recente. É possível abordar o contexto espacial em, ao menos, 3 aspectos: deslocamento, distância e área.

Figura 5 – Evolução no contexto espacial



Fonte: Elaboração própria em 2018

3.2.1.1 Deslocamento de IDR

Identificar padrões no deslocamento das IDRs pode ser eficiente para entender o processo que o analista realizou para resolver determinada tarefa. Com esse conhecimento, o sistema pode otimizar a análise exploratória para o analista numa tarefa similar ou para outros analistas que estejam realizando a mesma tarefa. Por exemplo, na Figura 5, os padrões de deslocamento estão indicados pelas setas curvilíneas. No primeiro momento (5.A), naturalmente não é possível identificar uma movimentação, mas as regiões encontradas nesse momento servem como ponto de partida da análise. No segundo momento (5.B), percebe-se um deslocamento: o analista, que antes explorou regiões ao norte, norte e oeste, agora tem sua atenção voltada para o norte e centro-oeste. Já no terceiro momento (5.B), pode-se identificar um padrão de deslocamento para região central.

3.2.1.2 Distância entre IDRs

A distância entre as IDRs, indicadas na Figura 5 pelas retas tracejadas do centro de cada região até o centro do conjunto, revela o quanto distribuído está a análise do usuário. Na Figura 5.A, ponto de partida da análise, o analista não sabe exatamente onde procurar, então ele avalia várias regiões na cidade, resultando em conjunto de IDRs bem distribuído. Com o passar das iterações, o espaço de procura do analista tende a diminuir, ou seja, as IDRs identificadas tendem a serem mais próximas como pode ser observado na sequência 5.B-C.

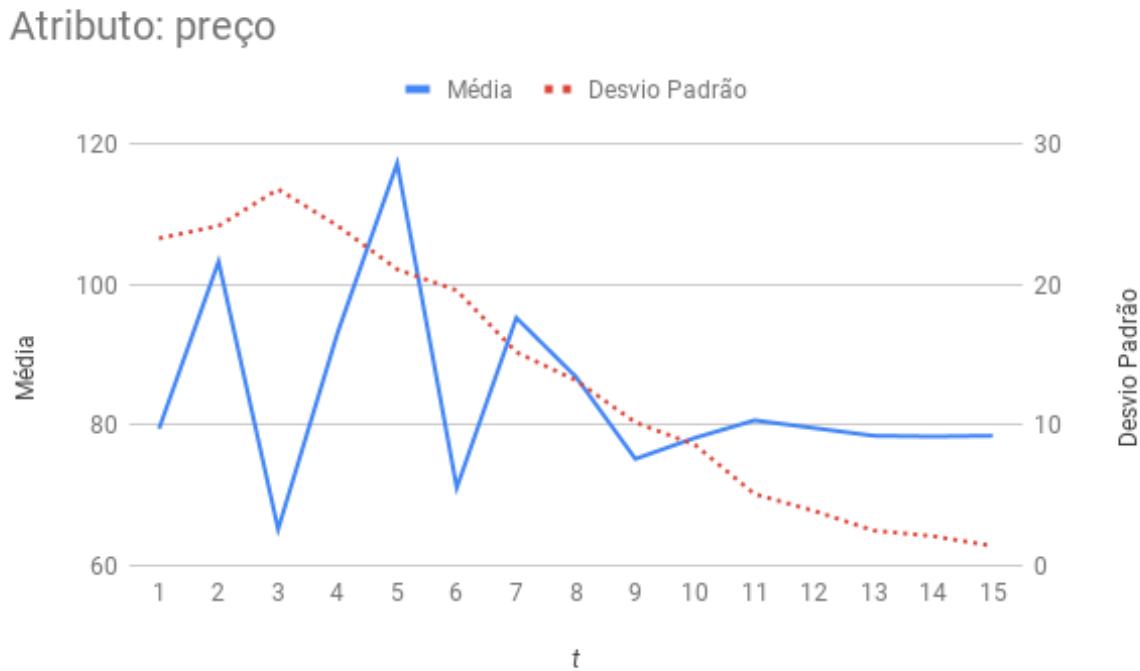
3.2.1.3 Área de IDR

A área de cada IDR, indicadas na Figura 5 pelos polígonos definidos, indica o quanto concentrado está a análise do usuário. No primeiro momento (5.A), as áreas das IDRs são consideravelmente maiores, o que pode indicar um comportamento de reconhecimento, ou seja, o analista está realizando um processo de reconhecimento do conjunto de dados a fim de se direcionar para próxima iteração. No segundo momento (5.B), as áreas das regiões diminuíram, o que pode apontar que a análise está mais precisa e direcionada. No momento seguinte (5.C), é notável a evolução do analista em direção ao objetivo da sua tarefa, visto que seu interesse, representado pelas IDRs, está mais concentrado indicando que o analista sabe onde e o quê está procurando.

3.2.2 Contexto de Domínio

O contexto de domínio se refere aos aspectos de domínio como, por exemplo, se o analista tem mais interesse em casas com varanda ou apartamentos. A análise é feita considerando os perfis das regiões, ou seja, como os atributos, de cada tipo, são relevantes ao analista ao decorrer das iterações.

Figura 6 – Evolução de atributo numérico



Fonte: Elaboração própria em 2018

Os atributos numéricos se apresentam relevantes quando a variação do valor médio tende a zero, assim como seu desvio padrão com o passar das iterações. Por exemplo, na Figura 6, observa-se que após 11 iterações (t), o atributo *preço* se mostrou relevante ao analista, uma

vez que seu valor médio se tornou linear, em outras palavras, com diferenças mínimas entre as iterações, e seu desvio padrão quase 0.

Diferentemente dos atributos numéricos que podem ser analisados pelas diferenças nas suas estatísticas, os atributos textuais e categóricos precisam ser analisados através de sua evolução no decorrer das iterações. Para tal, constrói-se um vetor de feedback F que é atualizado à cada iteração com base nos perfis das IDRs descobertas.

Considera-se um *valor de incremento* δ para atualizar F . Se v_1 para atributo a_1 é encontrado x vezes na IDR, aumenta-se por $x\delta$ o valor de $\langle a_1, v_1 \rangle$ em F . Durante a atualização, os valores em F só podem ser incrementados, ou seja, nunca são diretamente diminuídos.

Tabela 6 – Exemplo de perfil de atributos categóricos

Categoría	Quantidade
$\langle \text{Beds}, 1 \rangle$	3
$\langle \text{Beds}, 2 \rangle$	1
$\langle \text{Beds}, +2 \rangle$	0
$\langle \text{Balcony, Yes} \rangle$	4
$\langle \text{Balcony, No} \rangle$	0
$\langle \text{Air-cond., Yes} \rangle$	1
$\langle \text{Air-cond., No} \rangle$	3
$\langle \text{Rating, 1} \rangle$	0
$\langle \text{Rating, 2} \rangle$	0
$\langle \text{Rating, 3} \rangle$	0
$\langle \text{Rating, 4} \rangle$	1
$\langle \text{Rating, 5} \rangle$	3

Fonte: Elaboração própria em 2018

Usando como exemplo o perfil de atributos categóricos apresentado na Tabela 6 de um conjunto de dados sobre estadias, o vetor de feedback atualizado na primeira iteração pode ser observado na Tabela 7. Como o atributo “Beds” recebe o valor “1” três vezes, o valor da célula $\langle \text{Beds}, 1 \rangle$ é incrementado três vezes por δ . O mesmo processo é repetido para todos atributo-valores dos perfis. É importante notar que nem todas as células de F são necessariamente atualizadas durante o processo, como nesse exemplo, 5 células de 12 permaneceram sem atualização.

Por especificar um valor de incremento, é possível atualizar e normalizar o vetor de feedback usando uma função Softmax. F é sempre normalizado de maneira que seus valores somados sejam 1.0. Sendo $\delta = 1.0$, os valores normalizados do vetor F são ilustrados na terceira coluna da Tabela 7. Os maiores valores indicam os atributos mais relevantes.

Os valores normalizados do vetor F representam as preferências do analista. Por exemplo, os valores na Tabela 7 depois da atualização mostram que o analista tem bastante interesse em ter uma varanda (*balcony*) em sua estadia, uma vez que o valor $\langle \text{Balcony, Yes} \rangle$ é 0.25, isto é, o maior. Apesar de apenas considerar diferenças positivas, a função Softmax reduz valores das células que não foram relevantes a cada iteração ao passo que outras células são incrementadas.

É importante frisar que valores “0” no vetor F não significa a menor relevância, mas *irrelevância*. Por exemplo, considerando a célula $\langle Rating, 2 \rangle$ na Tabela 7, o valor “0” para essa célula indica que o analista não demonstrou interesse nessa característica. É possível que em iterações futuras, o analista demonstre interesse em estadias de 2 estrelas (potencialmente por causa do preço), nesse caso o valor dessa célula deixaria de ser 0. Entretanto, células que representam relevâncias menores são identificados com valores que tendem a 0. Por exemplo, o valor 0.06 para célula $\langle Rating, 4 \rangle$ mostra uma preferência por uma estadia de 4 estrelas inferior se comparado com as de 5 estrelas.

Tabela 7 – Atualizando vetor de feedback

Atributo-valor	Atualização	Normalizado
$\langle Beds, 1 \rangle$	$+3\delta$	0.19
$\langle Beds, 2 \rangle$	$+\delta$	0.06
$\langle Beds, +2 \rangle$	(sem atualização)	0.00
$\langle Balcony, Yes \rangle$	$+4\delta$	0.25
$\langle Balcony, No \rangle$	(sem atualização)	0.00
$\langle Air-cond., Yes \rangle$	$+\delta$	0.06
$\langle Air-cond., No \rangle$	$+3\delta$	0.19
$\langle Rating, 1 \rangle$	(sem atualização)	0.00
$\langle Rating, 2 \rangle$	(sem atualização)	0.00
$\langle Rating, 3 \rangle$	(sem atualização)	0.00
$\langle Rating, 4 \rangle$	$+\delta$	0.06
$\langle Rating, 5 \rangle$	$+3\delta$	0.19

Fonte: Elaboração própria em 2018

O vetor de feedback para atributos textuais difere apenas na sua inicialização. Como os perfis para atributos textuais são construídos separadamente, é preciso combinar os valores de cada atributo no único vetor. Por exemplo, se a frequência das palavras no atributo textual *name* é representado na Tabela 4 e do atributo *description* na Tabela 8, no vetor de feedback haveria células como $\langle confort, 33 \rangle$, $\langle mobilier, 32 \rangle$ e $\langle champs, 29 \rangle$, ou seja, as células presentes no vetor contemplam todos os atributos textuais.

Tabela 8 – Exemplo parcial de perfil do atributo textual

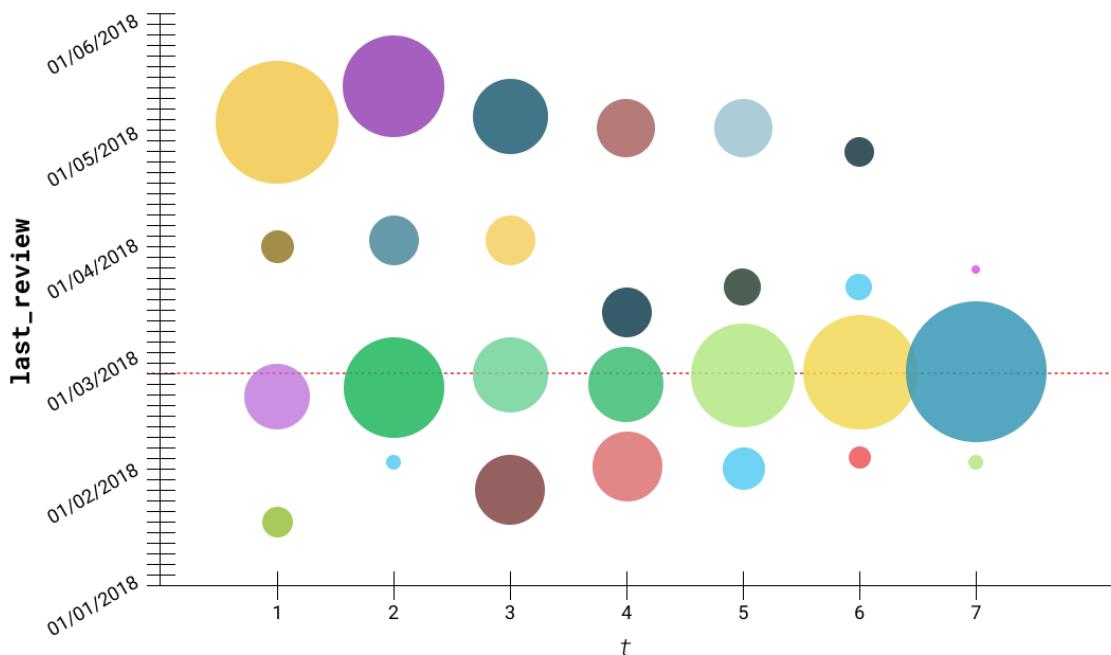
Atributo: description	
Termo	Ocorrências
confort	33
mobilier	32
appartement	22
décoré	21
adorable	18

Fonte: Elaboração própria em 2018

No caso dos atributos temporais, detectar padrões de preferências é mais complicado visto a natureza do dado temporal. A primeira abordagem é interpretar os perfis temporais ao

longo das iterações como na Figura 7, ou seja, como um gráfico de dispersão onde o eixo x são as iterações t , o eixo y é a linha temporal dos eventos para determinado atributo e o tamanho do ponto é a frequência do evento. A Figura 7 apresenta os resultados para o atributo *last_review* em um conjunto de dados de estadias após 7 iterações (t).

Figura 7 – Evolução de atributo temporal



Fonte: Elaboração própria em 2018

Nas primeiras iterações ($t \leq 4$), pode-se perceber que o analista não demonstra nenhum preferência específica para o atributo *last_review*, visto que os dados estão bem distribuídos. A partir da 5^º iteração, é notável a concentração do interesse do analista em avaliações (*reviews*) enviadas entre final de fevereiro e começo de março.

A segunda abordagem a ser utilizada para identificar padrões de preferências do analista sobre atributos temporais é considerar os valores máximos e mínimos para construir uma janela de interesse. Por exemplo, na Figura 7, pode-se dizer que a janela de interesse do usuário no evento *last_review* a partir de 7^º iteração é 15 de fevereiro de 2018 até 15 de março de 2018, portanto eventos que ocorreram entre essas datas são mais relevantes ao analista.

É importante notar no exemplo da Figura 7 que a granularidade da linha temporal do atributo em questão é de dias. A granularidade durante a análise pode ser variável, ou seja, se adequar a janela de interesse do usuário, ou fixa, como nesse exemplo. Isso é possível, pois os atributos temporais são normalizados durante a construção do perfil em 6 granularidade: anos, meses, dias, horas, minutos e segundos.

4 CONSIDERAÇÕES FINAIS

Neste trabalho é proposto um modelo de análise espaço-temporal para identificação de padrões nas preferências coletadas pela ferramenta GeoGuide enquanto um analista realiza a análise exploratória de dados espaciais. O modelo proposto mostra-se viável para ser utilizado a fim de aprimorar o uso da captura de feedback implícito para orientar o usuário nas próximas iterações através do método de destaqueamento de pontos geográficos.

4.1 CONTRIBUIÇÕES

O modelo propõe 2 novos conceitos: *i.* regiões densas interessantes e *ii.* análise temporal das preferências do usuário. As regiões densas interessantes (IDR) representam as preferências do analista no contexto espacial no determinado momento. Cada IDR possui também um perfil que representa as preferências do analista no contexto de domínio.

As regiões e seus perfis permitem a análise temporal das preferências do usuário. Os métodos para análise em ambos contextos espaciais e de domínio são investigados e apresentados com o propósito de identificar padrões nas iterações da análise exploratória.

4.2 TRABALHOS FUTUROS

Os dados gerados pelo modelo proposto podem futuramente ser explorados em diversos cenários. Por exemplo, os dados gerados podem ser utilizados para entender as preferências de um grupo de analistas a fim de potencializar o descobrimento de objetivos comuns entre eles.

No que diz respeito aos ambientes de exploração de dados espaciais, os dados coletados podem ser combinados com algoritmos preditivos para responder questionamentos como “onde o analista deve estar interessado na próxima iteração?” ou até mesmo “será que o analista está interessado neste apartamento com varanda?”.

REFERÊNCIAS

- ARAPAKIS, I. et al. User engagement in online news: Under the scope of sentiment, interest, affect, and gaze. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 65, n. 10, p. 1988–2005, 2014.
- ARAPAKIS, I.; LALMAS, M.; VALKANAS, G. Understanding within-content engagement through pattern analysis of mouse gestures. In: ACM. **Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management**. [S.I.], 2014. p. 1439–1448.
- 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>>.
- BARBER, C. B.; DOBKIN, D. P.; HUHDANPAA, H. The quickhull algorithm for convex hulls. **ACM Trans. Math. Softw.**, ACM, New York, NY, USA, v. 22, n. 4, p. 469–483, dez. 1996. ISSN 0098-3500. Disponível em: <<http://doi.acm.org/10.1145/235815.235821>>.
- BIRANT, D.; KUT, A. St-dbscan: An algorithm for clustering spatial-temporal data. **Data Knowl. Eng.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 60, n. 1, p. 208–221, jan. 2007. ISSN 0169-023X. Disponível em: <<http://dx.doi.org/10.1016/j.datak.2006.01.013>>.
- 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.I.], 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>>.
- ESTER, M. et al. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: **Proceedings of the Second International Conference on Knowledge Discovery and Data Mining**. AAAI Press, 1996. (KDD'96), p. 226–231. Disponível em: <<http://dl.acm.org/citation.cfm?id=3001460.3001507>>.
- 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>>.
- PRADEEP, S.; KALLIMANI, J. S. A survey on various challenges and aspects in handling big data. In: **2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)**. IEEE, 2017. Disponível em: <<https://doi.org/10.1109/iceeccot.2017.8284606>>.
- 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, 2010. 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.

ZHANG, J. et al. A survey of recent technologies and challenges in big data utilizations. In: **2015 International Conference on Information and Communication Technology Convergence (ICTC)**. IEEE, 2015. Disponível em: <<https://doi.org/10.1109/ictc.2015.7354594>>.

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>>.