

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E
ESTATÍSTICA**

Eduardo Machado Carboni

**Análise de Dados Espaço-Temporais
Gerados por Dispositivos Móveis**

**Florianópolis – SC
2011**

Eduardo Machado Carboni

Analise de Dados Espaço-Temporais Gerados por Dispositivos Móveis

Trabalho de conclusão de curso
apresentado como parte dos
requisitos para obtenção do grau
de bacharel em Sistemas de
Informação pela Universidade
Federal de Santa Catarina.

Orientadora

Vania Bogorny

Banca

Vitório Bruno Mazzola

João Bosco Sobral

Universidade Federal de Santa Catarina

1º Semestre 2011

Agradecimentos

Agradeço primeiramente a Deus por poder viver experiências boas e ruins da minha vida no período do curso de graduação de sistemas de informação da Universidade Federal de Santa Catarina, mas sempre com pensamentos positivos tendo a impressão de existirem sempre melhores momentos.

Dedico este trabalho à minha família que sempre me incentivou e apoiou nos momentos mais difíceis de minha vida, onde eu pude sempre estar para conseguir apoio e escolher caminhos a seguir. Em especial meu pai Altair Carboni, minha mãe M^a Gorete Machado Carboni e meus irmãos Fernando Machado Carboni e Julian Machado Carboni que nunca me privaram da oportunidade de estudar, que sempre me apoiaram em momentos de desânimo e que são motivos de orgulho em minha vida.

Agradeço à minha namorada Kamila Croceta Willemann pelo amor, carinho, paciência que soube ser compreensiva em momentos de provas, trabalhos e principalmente quando eu precisava de uma orientação para não me desesperar e perder o rumo de meus objetivos.

Aos meus amigos de faculdade, em especial Brian Henkels, Samuel Steckert, Leandro da Silveira, Tiago Bortoli e Rafael Bianco, que compartilhamos vários momentos de tristezas e comemorações, muitas vezes numa mesa de bar. Também ao Luis Fernando Custódio, Pedro da Silveira Grandi, João Vitor Martins e Gustavo Coan Hobold, que compartilhamos alegrias e tristezas, provas e trabalhos, e claro a vitória da conclusão deste curso.

Agradeço à professora e orientadora Vania Bogorny pelo desafio deste trabalho, dedicação e amizade dispensada. Aos professores Vitório Bruno Mazzola e João Bosco Sobral que aceitaram participar da banca de trabalho e que contribuiu de forma significativa para a realização do mesmo.

Enfim, agradeço a todos que me auxiliaram nesta caminhada tão importante para atingir essa grande etapa da minha vida.

Resumo

Os sistemas embarcados ficam a cada dia mais populares, e com eles as coleções de dados de objetos móveis vem ficando mais e mais comuns, criando assim uma crescente necessidade de análises eficientes e extração de conhecimento destes dados para aplicações de diferentes domínios. Dispositivos móveis geram dados chamados de trajetórias de objetos móveis. Contudo, os dados de trajetórias são normalmente disponibilizados como um conjunto de coordenadas geográficas associadas a um instante de tempo, sem informações semânticas, o que traz dificuldades na hora de interpretar os dados.

Diversos trabalhos existem na área de análises de trajetórias, mas poucos tem focado na análise de comportamento da trajetória. Com esta situação é proposta uma nova maneira de analisar os dados de trajetórias, através da análise da aceleração, desaceleração e mudança brusca de direção. Em resumo, o objetivo geral deste trabalho é buscar a inferência automática do comportamento de objetos móveis analisando regiões de aceleração brusca, desaceleração brusca e mudança brusca de direção.

Com este estudo espera-se um resultado que mostre trechos anormais que permitam analisar o comportamento do objeto móvel considerando os parâmetros anteriormente mencionados.

Palavras-Chave

Trajетórias de objetos móveis, Análise de dados, Comportamento anormal, Análise de velocidade e direção em trajetórias.

Sumário

| | |
|--|----|
| Siglas e Abreviações..... | 6 |
| Listagem de Figuras | 7 |
| Listagem de Tabelas | 8 |
| 1. Introdução e Motivação | 9 |
| 2. Conceitos Básicos..... | 15 |
| 2.1Trajetórias Brutas | 15 |
| 2.2Trajetórias Semânticas..... | 15 |
| 2.3 Técnicas para Adição de Semântica em Trajetórias..... | 16 |
| 2.3.1 IB-SMoT (Intersection Based Stops and Moves of Trajectories)[2]..... | 16 |
| 2.3.2 CB-SMoT (Clustering Based Stops and Moves of Trajectories) [3]..... | 17 |
| 2.3.3 DB-SMoT (Direction Based Stops and Moves of Trajectories)..... | 18 |
| 2.4 Análise de Comportamentos em Trajetórias | 20 |
| 3 Um método para Análise de Comportamento | 21 |
| 3.1 Aceleração e Desaceleração | 21 |
| 3.1.1 Acelerações..... | 22 |
| 3.1.2 Desacelerações..... | 24 |
| 3.2 Mudança de Direção..... | 26 |
| 3.3 Um algoritmo para identificar variações anormais em trajetórias com base na aceleração, desaceleração e mudança de direção..... | 28 |
| 3.4 Implementação..... | 30 |
| 4 Experimentos | 34 |
| 4.1 Experimentos com fonte de dados 01 | 41 |
| 4.2 Experimentos com fonte de dados 02..... | 34 |
| 5 Conclusões..... | 49 |
| 6 Referências | 51 |

Siglas e Abreviações

GPS: Global Positioning System

BD: Banco de Dados

Stops: Regiões importantes nas trajetórias

Moves: Sub-trajetórias não compreendidas nos Stops

SMoT: Stops and Moves of Trajectory

IB-SMoT : Intersection Based - Stops and Moves of Trajectory

CB-SMoT: Cluster Based - Stops and Moves of Trajectory

DB – SmoT: Direction Based - Stops and Moves of Trajectory

KDD LAB – Knowledge Discovery and Data Mining Laboratory

UFSC – Universidade Federal de Santa Catarina

Listagem de Figuras

| | |
|--|----|
| Figura 1 – Trajetória representada por pontos no espaço. [6] | 10 |
| Figura 2 – Trajetória representada por pontos no espaço e informações semânticas [6] | 11 |
| Figura 3– Trajetória Bruta [5] | 15 |
| Figura 4- Comparação de uma trajetória bruta com uma trajetória semântica [5] | 16 |
| Figura 5 – Trajetória de barco pescueiro [4] | 20 |
| Figura 6 – Trajetória com aceleração brusca e desaceleração brusca | 22 |
| Figura 7 - Tela inicial de conexão | 32 |
| Figura 8 - Tela de trajetórias..... | 33 |
| Figura 9– Interface desenvolvida no TCC..... | 34 |
| Figura 10 - Trajetórias dos dados reais de Santa Catarina | 36 |
| Figura 11 - Trajetória 34 com possíveis locais de comportamento anormal..... | 38 |
| Figura 12 – Trajetória 34 com parâmetros diferentes para busca de comportamentos anormais | 39 |
| Figura 13 - Regiões de mau comportamento na trajetória 34(A) e intersecção de polígonos com comportamentos anormais da fonte de dados de Santa Catarina(B) | 41 |
| Figura 14 - Trajetórias da fonte do Rio de Janeiro | 43 |
| Figura 15 – Trajetória 26 com parâmetros diferentes para busca de comportamentos anormais | 45 |
| Figura 16 - Regiões anômalas da trajetória 16 e em mais de outras 4 trajetórias | 47 |
| Figura 17 - Regiões anômalas da trajetória 49 e em mais de outras 4 trajetórias | 48 |

Listagem de Tabelas

| | |
|--|--|
| Tabela 1 – Valores de velocidades e pesos..... | Erro! Indicador não definido. |
| Tabela 2 – Valores de curvatura e velocidades aceitáveis..... | Erro! Indicador não definido. |
| Tabela 3 - Resultados do algoritmo para a fonte de dados de Santa Catarina..... | 36 |
| Tabela 4 – Resultados do algoritmo para a fonte de dados do Rio de Janeiro | Erro! Indicador não definido. |

1. Introdução e Motivação

Em todo local de qualquer sociedade, em todas as épocas, existiu a necessidade de deslocamento de um local para outro de pessoas. Com esta movimentação também são transportadas mais coisas além do objeto que se move. Também sempre existiram sistemas de trocas e vendas de mercadorias entre povos e etnias, e o transporte dessas encomendas de um ponto geográfico a outro para chegar do vendedor ao consumidor. Contudo, no trajeto dos produtos, por exemplo, podem acontecer várias coisas inesperadas que, dependendo da situação, tanto cliente como produtor, nunca ficarão sabendo.

Com o aumento da população e do transporte de produtos surge a necessidade de controlar o percurso dos meios de transporte e o comportamento destes em suas trajetórias. Uma maneira de controlar isso é através de pessoas de confiança que podem acompanhar o trajeto e anotar numa planilha o comportamento dos meios de transporte e assim deduzir o comportamento daqueles que conduzem o veículo. Hoje, porém, com o auxílio da computação e da tecnologia, é crescente o número de dispositivos de sistemas embarcados com preços reduzidos que permitem a obtenção de dados com as posições geográficas e o instante de tempo de toda a trajetória de objetos móveis, tais como GPS, celulares, entre outros, aumentando significativamente a utilização dos serviços de geo-referenciamento por todas as classes da população. Com este novo tipo de dado é possível realizar análises de percursos percorridos por um objeto no espaço em um período de tempo, conseguindo assim identificar comportamentos, podendo afirmar com bastante precisão o comportamento que um objeto apresentou durante toda a sua trajetória.

A trajetória de um objeto é formada pelos pontos coletados por seu aparelho embarcado, desde determinado local inicial a um local final, durante um intervalo de tempo. Estes dados são armazenados no banco de dados normalmente na forma (tid, x, y, t) , onde tid é o identificador de cada trajetória, x e y são as coordenadas geográficas no espaço e no tempo que é representado por t . Os aparelhos mais utilizados para coleta de dados são GPS. Normalmente os aparelhos de GPS armazenam informações num intervalo de 1 segundo, trazendo um grande volume de dados, o que pode ser considerado uma grande mina de informações para serem exploradas e interpretadas com objetivos e finalidades diferentes, desde a identificação de regiões com alto número de tráfego,

seqüências de trajetórias e até mesmo os pontos importantes das trajetórias para diferentes domínios de aplicação.

Muitos estudos focalizam a descoberta de informações entre grupos de trajetórias, enquanto neste trabalho se objetiva traçar perfis de comportamento de trajetórias individuais, de um mesmo indivíduo. Neste trabalho o foco será em comportamentos incorretos e irresponsáveis, tais como de um veículo com motorista que age realizando acelerações e desacelerações bruscas, altas velocidades em curvas com baixa amplitude angular, entre outras. Imagine a situação de um veículo que transporta frutas, por exemplo, caso fique acelerando e desacelerando bruscamente, as frutas irão sofrer algum tipo de esmagamento, deixando as mesmas com qualidade inferior. Estima-se que 30% de produção de frutas são perdidos devido ao transporte incorreto. Outro caso, por exemplo, é um veículo de transporte coletivo onde os usuários o utilizam ficando de pé, e o veículo faz curvas fortes em alta velocidade. Isso pode gerar desconforto aos passageiros.

Conhecendo a maneira como os dados são armazenados no banco de dados geográfico (tid, x, y, t), podemos agora visualizar os dados de modo a mostrar uma trajetória, conseguindo visualizar para quais direções o objeto está movendo e que de um ponto para outro ele pode percorrer uma distância maior, ou seja, alterado sua velocidade. A figura 1 ilustra um simples trecho de uma trajetória com informações simples e de pouca semântica, mas mesmo assim conseguimos identificar as posições onde o objeto estava mais rápido. Quanto maior a distancia entre os pontos, mais rápido estava a movimentação do objeto.

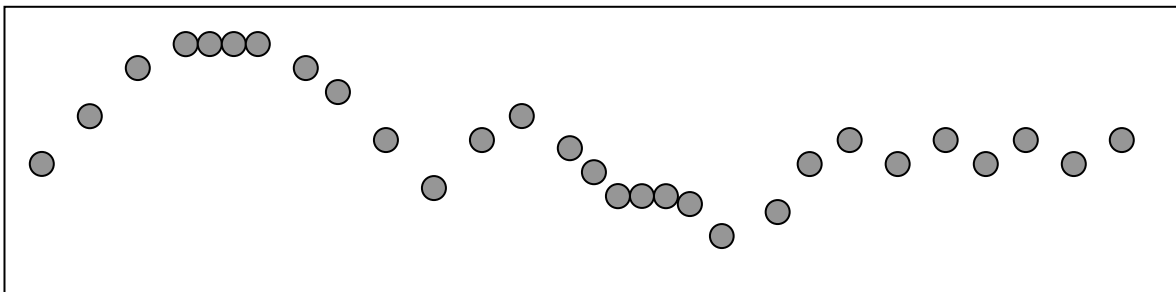


Figura 1 – Trajetória representada por pontos no espaço. [6]

Olhando para a figura 1 podemos inferir alguns fatos, tais como a variação da velocidade e da direção. Pontos que são gravados em intervalos de tempo de 1 segundo se

não variassem a velocidade estariam sempre com a mesma distância entre si, e a variação da direção podemos perceber através do posicionamento dos pontos, pois se fossem sempre na mesma direção teríamos uma reta. Informações mais importantes como os locais onde o objeto permanece por um certo tempo não se consegue, apenas o posicionamento geográfico. Nesse aspecto, surge a idéia de acrescentar informações semânticas às trajetórias, trazendo informações dos lugares que o objeto móvel visitou como, por exemplo, pontos turísticos, postos policiais, hotéis, entre outros, dependendo da aplicação. A figura 2 apresenta a mesma trajetória da figura 1, porém com algumas informações a mais com locais importantes por onde a trajetória passou.

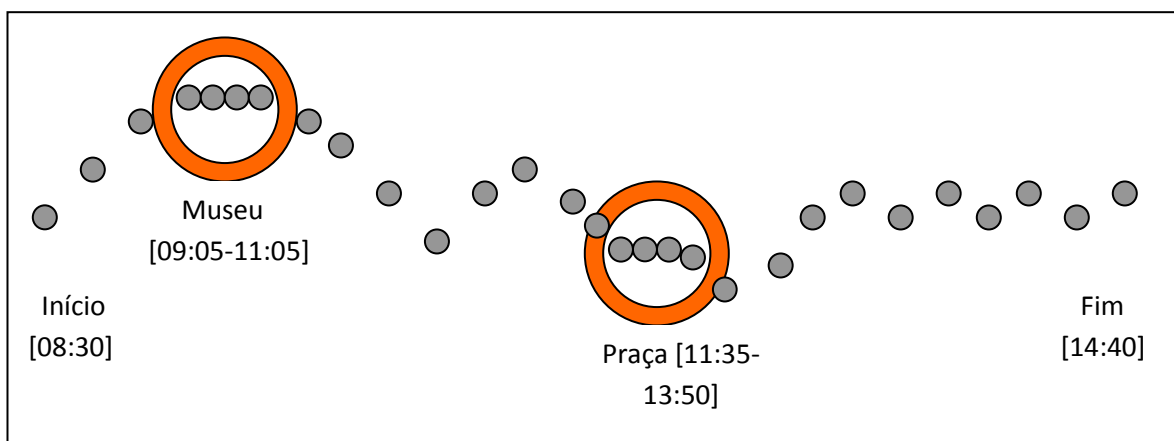


Figura 2 – Trajetória representada por pontos no espaço e informações semânticas [6]

Para encontrar pontos importantes em trajetórias foram criados alguns modelos de análise e enriquecimento semântico das trajetórias. Em 2008, Spaccapietra [1] introduziu um novo modelo de raciocínio sobre trajetórias, que nos permite uma poderosa análise semântica, chamado *stops e moves* [1]. O primeiro é um local de parada do objeto por um determinado período de tempo. Já o segundo representa os pontos da trajetória que estão entre os *stops*, que estão entre o início da trajetória e o primeiro *stop* ou estão entre o último *stop* e o fim da trajetória.

Numa trajetória, todos os *stops* precisam ser calculados, e o método varia de acordo com a aplicação e o interesse do usuário. Nosso grupo de pesquisa já desenvolveu três métodos para este fim, IB-SMoT (*Intersection Based Stops and Moves of Trajectories*) [2], CB-SMoT (*Clustering Based Stops and Moves of Trajectories*) [3] e DB-SMoT (*Direction Based Stops and Moves of Trajectories*) [4]. O primeiro calcula *stops* com base na

intersecção da trajetória com objetos geográficos de interesse. O segundo calcula *stops* em lugares onde a velocidade do objeto é baixa e o terceiro gera *stops* onde a direção da trajetória varia mais que um determinado valor. Para todos os métodos, os *stops* são gerados quando o objeto permaneceu no *stop* por um período mínimo de tempo. Podemos exemplificar a aplicação dos três modelos nas seguintes situações:

A – IB-SMoT [2] - Encontrar os hotéis que um veículo ficou parado durante uma viagem. Nesse caso seriam os pontos da trajetória que ficaram parados (interceptando) os hotéis por um período mínimo de tempo. Esse método é interessante para aplicações como turismo;

B – CB-SMoT [3] – Encontrar trechos de uma rodovia onde houve congestionamento. São encontrados os locais onde o veículo ficou abaixo de determinada velocidade por um período de tempo mínimo. Esse método é interessante para aplicações de gerenciamento de tráfego urbano;

C – DB-SMoT [4] – Encontrar locais onde um barco pesqueiro esteve pescando. Para utilizar esse modelo é necessário definir um tempo mínimo em que um barco ficou abaixo de determinada velocidade e variando uma angulação mínima de direção.

Com os métodos e modelos já propostos é possível definir pontos importantes das trajetórias, mas não o comportamento dos objetos que percorrem essas trajetórias, por isso um algoritmo foi desenvolvido para analisar a variação da direção, aceleração e desaceleração em trajetórias definindo comportamentos.

Criar perfis de pessoas e as classificar é algo que se torna cada vez mais necessário para adquirir conhecimento sobre interesses pessoais e comportamentos. Seja por perfis de compras ou por perfis de redes sociais, hoje existem grandes quantidades de dados conhecidos como “*reality mining*”, que seriam as análises dos *footprints*, ou pegadas digitais com o objetivo de traçar os perfis humanos. Nessa área existe a subárea de geo-localização, que pode nos auxiliar a descobrir quais regiões de uma cidade possuem maior poder aquisitivo. Pode-se, geograficamente, também monitorar o trânsito e extrair padrões, descobrindo, por exemplo, quais regiões são mais movimentadas e em quais horários. Também é possível, através das trajetórias, inferir os locais prováveis por onde um objeto

móvel irá passar e descobrir seu comportamento conforme suas atitudes no passar do tempo no espaço geográfico.

Em contra partida aos dados gerados e o potencial de mineração dos mesmos, está a invasão da privacidade. Temos um exemplo retirado de um trecho do site do terra [8]: “Darryl Littlejohn, segurança de uma boate, fez diversas chamadas de seu telefone enquanto dirigia de sua casa no Queens para uma rua vazia no Brooklyn, onde jogou o corpo de Imette St. Guillen, a estudante de pós-graduação por cujo assassinato ele foi condenado este mês.” (<http://tecnologia.terra.com.br>). O trecho acima cruzou informações da localização do aparelho telefônico do segurança e o local onde o corpo da vítima foi encontrado, incriminando o homem. Esse é um exemplo de quebra da privacidade, embora que por fins justificáveis. Outro bom exemplo é o que a operadora de telefonia Vivo realiza: “Pelo "Vivo Localiza Família", pais e responsáveis também podem monitorar crianças e jovens, que se transformam em pontos que piscam em mapas na tela do celular ou na Internet. Como em um "Big Brother" móvel, é possível delimitar regiões geográficas como limites de circulação: os pais recebem alertas por SMS ou e-mail quando os filhos estão dentro ou fora destes locais.”(<http://oglobo.globo.com>). Embora se tenham muitos argumentos que possam vir a inibir a mineração de dados geográficos com objetivos de definir comportamentos e perfis, ainda continuamos com o objetivo do trabalho inalterado, pois os resultados têm por finalidades apenas contribuições científicas.

Ainda para um incentivo maior, hoje se vivencia uma grande popularidade de aparelhos de coleta de dados geográficos com o barateamento de celulares e GPS, o que disponibiliza um grande volume de dados com muita diversidade de informações a serem extraídas e estudadas.

O restante deste trabalho é organizado como segue: O capítulo 2 apresenta alguns conceitos de trajetórias brutas e semânticas, juntamente com os modelos de adicionar semântica e definições de comportamentos de motoristas. A seção 3 mostra como foram realizados os métodos de análise. A quarta parte do trabalho são experimentos e por fim o capítulo 5 apresenta as conclusões do trabalho.

2. Conceitos Básicos

Para facilitar a compreensão deste trabalho vamos aprimorar os conceitos de trajetórias com e sem informações semânticas, demonstrar os três modelos de mineração de dados que utilizam os conceitos de *stops* e *moves* e por fim apresentar algumas definições de como serão tratados os assuntos e regras de comportamentos de pessoas.

2.1 Trajetórias Brutas

Na introdução deste estudo mostramos de maneira subjetiva como os pontos são representados no banco de dados geográfico na forma (tid, x, y, t) , onde tid é o identificador de cada trajetória, x e y são as coordenadas geográficas no espaço e no tempo que é representado por t e ainda são mostrados espacialmente as suas representações. Podemos definir uma trajetória como:

As trajetórias brutas são apenas seqüências de pontos espaciais ordenados pelo tempo de ocorrência de cada ponto [5], conforme ilustra a figura 3:

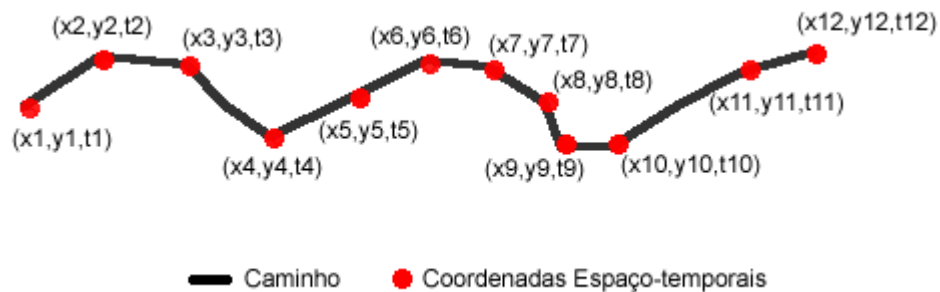


Figura 3– Trajetória Bruta [5]

2.2 Trajetórias Semânticas

Tendo uma trajetória bruta disponível, o minerador dos dados ainda não consegue responder algumas perguntas tais como em qual hotel um turista dormiu, ou em qual shopping ele esteve realizando compras, pois o que temos até então são a localização e o tempo. Com este problema proposto, foram desenvolvidas técnicas para adicionar semântica às trajetórias, como por exemplo, cruzar informações geográficas da localização dos hotéis de determinada região. Na figura 4 temos a comparação de uma trajetória bruta e uma trajetória semântica:

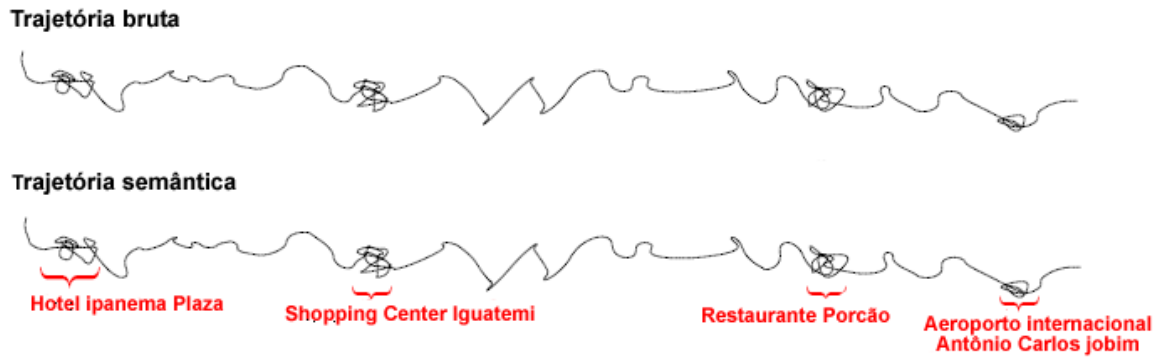


Figura 4- Comparação de uma trajetória bruta com uma trajetória semântica [5]

Como podemos observar na figura 4, os locais onde o objeto móvel ficou mais tempo foram locais onde um turista ou viajante esteve. Com os dados das trajetórias cruzados com dados de locais podemos responder perguntas do tipo: onde o indivíduo dormiu, ou onde ele almoçou e por assim seguir. Então podemos inferir a seguinte definição:

Definição (Trajetória Semântica): Uma trajetória semântica é formada pelos dados da trajetória bruta, aliados à outros dados semânticos feitos através do cruzamento das informações.

Para definir os locais importantes das trajetórias e cruzar as informações foi criado o modelo de enriquecimento semântico das trajetórias conhecido como modelo de *stops* e *moves* [1], onde o primeiro representa os locais importantes das trajetórias e o segundo são os pontos entre os *stops* ou início e fim. Para a descoberta dos *stops* e a adição de semântica nas trajetórias foram desenvolvidos, com base no modelo de *stops* e *moves*, alguns algoritmos baseados em velocidade, tempo e direção. A seguir estão detalhados estes algoritmos.

2.3 Técnicas para Adição de Semântica em Trajetórias

O objetivo dos algoritmos que foram desenvolvidos para adicionar semântica às trajetórias é basicamente descobrir os pontos importantes das trajetórias e guardá-los em

tabelas específicas para que quando se vá minerar uma trajetória ou fazer consultas sobre ela, seja possível um ganho em performance para consultas similares. Embora os modelos de criação de *stops* possam gastar algum tempo a mais em seu pré-processamento, depois de criados os *stops*, todas as minerações e consultas daquelas trajetórias serão mais rápidas e com qualidade. Nos próximos tópicos descrevemos os três modelos IB-SMoT [2], CB-SMoT [3] e DB-SMoT [4] que têm por objetivo a descoberta dos *stops* nas trajetórias.

2.3.1 IB-SMoT (Intersection-Based Stops and Moves of Trajectories)

Este modelo foi desenvolvido por Alvares [2] em 2007, e foi o primeiro modelo semântico para trajetórias. Ele é baseado na intersecção da trajetória com pontos importantes, tais como hotéis, restaurantes, hospitais por um determinado período de tempo.

O algoritmo funciona iniciando com a aquisição de todos os pontos de uma dada trajetória. Tendo-se todos os pontos, inicia a partir do primeiro até o último, verificando se o ponto intercepta algum candidato à *stop*. Um candidato à *stop* é um objeto geográfico de interesse, como um hotel, por exemplo. A partir do momento que um primeiro ponto intercepta um candidato à *stop* então se inicia um laço que vai até o último ponto interceptado. Quando os pontos param de interceptar um determinado candidato, então são feitas comparações de tempo para ver se o objeto ficou o tempo mínimo no candidato a *stop*. Se isso for respeitado, então se considera que aquele conjunto de pontos formam um *stop*. Depois de percorrida a trajetória são criadas duas tabelas, uma com *stops* e outra com os *moves* da referida trajetória.

Este algoritmo é muito útil para responder perguntas sobre, por exemplo, em quais pontos turísticos um turista visitou durante seu passeio. Podemos também descobrir hotéis onde ele passou as noites, restaurantes onde almoçou e assim por diante.

2.3.2 CB-SMoT (Clustering-Based Stops and Moves of Trajectories)

Este modelo foi desenvolvido por Palma [3] em 2008. Ele além de considerar o tempo que um objeto ficou interceptando um candidato à *stop*, ele também considera a velocidade do objeto verificando se ela respeita uma velocidade mínima. É interessante destacar também que ele procura por *clusters* ainda desconhecidos para que se tornem *stops* das trajetórias que são testadas levando em consideração um determinado tempo em que o objeto permaneceu em baixa velocidade. Com isso se conseguem encontrar novos tipos de *stops*.

Este algoritmo opera basicamente em três etapas. A primeira acontece com a seleção dos pontos que respeitam a regra de estarem a uma velocidade abaixo da definida, por um período de tempo mínimo, definido pelo usuário. A velocidade a ser respeitada é sempre calculada com o valor passado como parâmetro multiplicado pela velocidade média. Por exemplo, se considerarmos a trajetória de um automóvel onde a velocidade média na trajetória é de 60 km/h e o parâmetro de velocidade é 0,5 (50%), então o veículo deve estar abaixo de 30 km/h para que exista um *stop*. Estando o veículo abaixo da velocidade, é verificado o tempo em que ele fica nessa faixa de velocidade. Os pontos com velocidade baixa por um tempo mínimo são chamados de *clusters*.

O segundo passo do algoritmo é cruzar os *clusters* com os candidatos à *stops* que estão na região da trajetória, de maneira similar ao algoritmo anterior, onde se os *clusters* selecionados respeitarem o tempo mínimo interceptando o possível candidato a *stop*, então serão considerados *stops*. Os *clusters* encontrados que não interceptarem os candidatos à *stops* serão considerados *stops* desconhecidos (*Unknown stops*).

Por fim, uma funcionalidade a mais, a última tarefa deste algoritmo é verificar se os *clusters* desconhecidos acontecem em mais de uma trajetória e no mesmo lugar. Em caso positivo receberão o mesmo nome.

É importante destacar a importância deste modelo no aspecto de descoberta de um *stop* sem informação alguma da região, apenas dos dados da trajetória. Por exemplo, podemos imaginar os dados da trajetória de um veículo que em nenhum momento esteve

parado. Digamos que o veículo partiu do bairro *A* em direção ao bairro *B*, sendo que sua velocidade mudou frequentemente na rodovia. Com o modelo proposto por Palma [3] é possível encontrar regiões onde o automóvel passou por congestionamentos, pois nesses locais o objeto percorreu o espaço com baixa velocidade e demorando um tempo maior que o esperado. Situações como essa não são solucionadas com o algoritmo IB-SMoT.

2.3.3 DB-SMoT (Direction-Based Stops and Moves of Trajectories)

Este modelo foi criado por Manso [4] em 2010, e tem por objetivo encontrar pontos importantes em trajetórias considerando a variação da direção. Ele, da mesma maneira que o primeiro e o segundo compara o tempo de permanência em regiões que podem ser um *stop*.

Este modelo inicia com todos os pontos de uma trajetória e começa a comparar de dois em dois a variação da direção entre um ponto e outro. Caso a variação seja maior que o limite dado como parâmetro, esses pontos são comparados com os próximos dois. Encontrado um ponto que não atende a variação da direção mínima, é verificado o tempo que o objeto ficou variando a direção, e caso satisfaça a condição de ficar mais tempo que o mínimo determinado como parâmetro, então esses pontos são acrescentados numa tabela com todos os *clusters*. Cada *cluster* é a sub-trajetória com os pontos que variam de direção por um tempo mínimo.

Este modelo é interessante em aplicações onde a variação de direção é importante. É possível, por exemplo, solucionar um problema encontrado em barcos pesqueiros conforme foi demonstrado no artigo do Manso [4]. Com o método de variação de direção pode-se definir o local onde um barco se manteve pescando, pois quando ele sai do porto em direção ao local de pesca ele anda em linha reta até o local. Chegando lá ele permanece desligado variando sua direção conforme as correntes marítimas e o balançar das ondas. Após determinado tempo variando a direção, esperando os peixes, então o barco vai para outro local de pesca ou retorna ao porto quase em linha reta, conforme ilustra a figura 5.

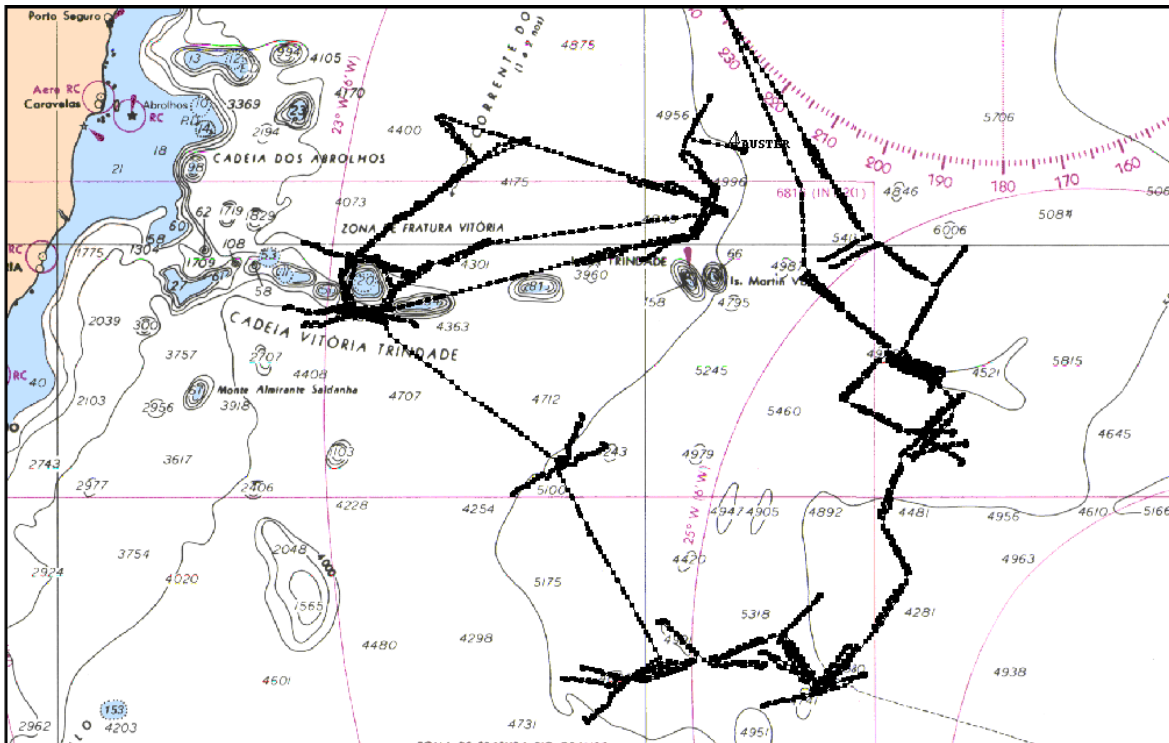


Figura 5 – Trajetória de barco pescador [4]

Em uma aplicação de pesca, com o algoritmo CB-SMoT seria encontrados *stops* que não correspondem à regiões de pesca como, por exemplo, o sair e chegar ao porto, pois nessas regiões o barco anda lentamente devido ao peso (na partida pelo peso do combustível e na chegada pelo peso dos pescados).

De maneira geral, o algoritmo DB-SMoT é mais útil em situações em que a variação da direção importa para descobrirmos regiões importantes nas trajetórias, mas caso isso não seja importante, como num congestionamento, então o modelo CB-SMoT é mais útil por conseguir um processamento mais rápido com resultados similares.

2.4 Análises de Comportamentos em Trajetórias

Sobre análise de comportamentos em trajetórias tem-se algumas pesquisas existentes na literatura, em especial, do *KDD LAB* [14] da Universidade de Pisa na Itália.

O trabalho desenvolvido por Chiara Renso [15] traça tipos de comportamentos em trajetórias através da origem, destino e locais percorridos pelas trajetórias. Esse trabalho desenvolvido no KDD Lab verifica as informações das trajetórias e as classifica em trajetórias turísticas ou de trabalho, entre outras.

O trabalho aqui proposto mostra o comportamento dos objetos móveis ao longo de suas trajetórias, sem interessar os tipos de trajetórias e com o objetivo de inferir se um condutor está ou não cumprindo com algumas regras de boa condução.

3 Um método para Análise de Comportamento em Trajetórias

Como o objetivo deste trabalho é analisar o comportamento do objeto móvel através de sua trajetória, serão consideradas três características fundamentais: aceleração brusca, desaceleração brusca e mudança brusca de direção. Estas três medidas foram definidas por acreditar-se que estas caracterizam o perfil de um mau condutor.

3.1 Aceleração e Desaceleração

A figura 6 apresenta uma trajetória, com pontos de aceleração e desaceleração, que podem ser percebidos pelas distâncias entre os pontos. Considerando neste exemplo, que entre cada ponto existe um intervalo de 1 segundo, quando os pontos vão se afastando entre si, os mesmos estão aumentando a velocidade, e quando vão se aproximando, estão reduzindo a velocidade. Se a aproximação dos pontos é muito rápida, então podemos perceber uma desaceleração brusca.

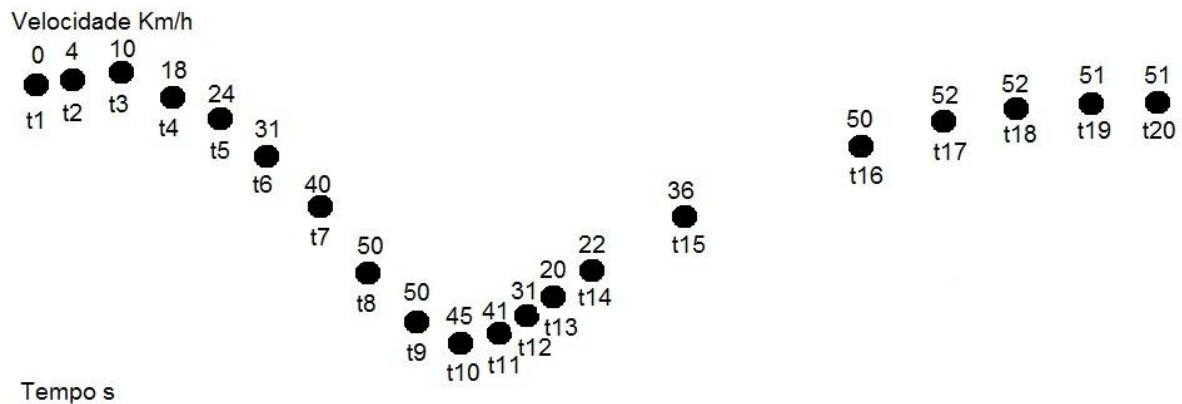


Figura 6 – Trajetória com aceleração brusca e desaceleração brusca

A figura 6 ilustra um exemplo de uma trajetória com uma desaceleração brusca dos pontos t9 à t13 e uma aceleração brusca nos pontos t14 à t16. A trajetória tem um início suave, onde os pontos se separam gradativamente. De repente, a partir do ponto no instante t9 temos uma desaceleração brusca. Aquele local é considerado um ponto importante na

trajetória do objeto, e poderíamos eventualmente considerar um conjunto de pontos como um *stop*.

Da mesma maneira que a desaceleração observada na figura 6 no intervalo de t_9 a t_{13} , é possível observar o oposto, uma aceleração brusca na parte da trajetória representada nos pontos t_{14} até t_{16} . Para entendermos a diferença entre acelerações e desacelerações bruscas, as seguintes subseções apresentarão as definições de acelerações e desacelerações normais e bruscas.

3.1.1 Acelerações

Quando há uma variação na velocidade, caracteriza-se uma aceleração ou uma desaceleração. Nas situações em que a velocidade do objeto móvel aumenta de um ponto para outro, caracteriza-se uma aceleração.

Definição 1(aceleração): A aceleração é um aumento na velocidade V de um ponto p , tal que $V_p > V_{p-1}$ e $V_{p-1} > V_{p-2}$, onde V_p é a velocidade no ponto p , V_{p-1} é a velocidade no ponto anterior ao ponto p , V_{p-2} é a velocidade no ponto anterior ao ponto $p-1$.

Quando o aumento de velocidade entre dois pontos consecutivos é maior que um determinado percentual, então caracteriza-se uma aceleração brusca. No exemplo da figura 6, os pontos dos instantes $[t_{16}$ e $t_{17}]$ e $[t_{15}$ e $t_{16}]$ apresentam acelerações, mas se for vista a velocidade nos pontos e ter como premissa que se o aumento da velocidade for maior que 10% para ser considerada brusca, então a aceleração nos instantes t_{16} e t_{17} não é brusca, pois a velocidade destes pontos não aumenta 10%, enquanto que nos instantes t_{15} e t_{16} o aumento é superior à 10%, que seguindo a premissa seria uma aceleração brusca.

A principal diferença entre uma aceleração normal e uma aceleração brusca é o tempo. Em uma aceleração brusca o objeto vai percorrer a mesma distância que em uma aceleração normal, mas em menos tempo.

Uma aceleração normal caracteriza-se por um percentual aceitável no aumento de velocidade em um segundo, por exemplo. Quando este percentual de aumento é muito elevado, caracteriza-se uma aceleração brusca. Entretanto, o aumento percentual de

velocidade a cada segundo, quando um objeto está quase parado é muito maior do que quando o objeto já está em alta velocidade. Em outras palavras, uma aceleração é maior e mais brusca quando o objeto inicia o movimento em uma velocidade baixa. Assim sendo, para evitar que uma aceleração normal de 1 Km/h para 2 Km/h em um segundo seja considerada brusca, foi criada uma escala de velocidades com pesos diferentes.

As escalas foram definidas gerando um grau de importância no percentual de aumento da velocidade, conforme ilustra a tabela 1.

Tabela 1 – Valores de velocidades e pesos.

| Velocidade: Km/h | Peso (grau de importância): |
|------------------|-----------------------------|
| 0 até 1 | 20 |
| 1 até 3 | 10 |
| 3 até 10 | 4 |
| 10 até 30 | 3 |
| 30 até 70 | 2 |
| Maior que 70 | 1 |

O grau de importância funciona para multiplicar o percentual de aceleração, por exemplo, se for considerado um aumento de 20% e um ponto estiver com velocidade de 2 Km/h, o aumento real limite seria, ao invés de 0,2, o valor 2, pois seria $0,2 \cdot 10$. Com isso, o objeto teria de estar a mais de 4 Km/h para ser caracterizada uma aceleração brusca. Essas faixas de valores foram separadas considerando testes feitos na base de dados onde com valores menores de peso retornavam valores irrealis.

Considerando então a escala aceitável de aceleração normal apresentada na tabela 1, é possível definir uma aceleração brusca.

Definição 2 (aceleração brusca) : A aceleração brusca é um aumento na velocidade V de um ponto p , tal que $(V_p + V_p \cdot (A \cdot S)) \cdot \Delta t < V_{p+1}$ e $(V_{p+1} + (V_{p+1} \cdot (A \cdot S))) \cdot \Delta t < V_{p+2}$, onde V_p é a velocidade no ponto p , V_{p+1} é a velocidade no ponto seguinte ao ponto

p , V_{p+2} é a velocidade no ponto seguinte ao ponto $p+1$, A é o percentual de aceleração aceitável para o objeto móvel, sendo que este A varia de 0 a 1, S é o peso atribuído a velocidade de acordo com a tabela 1 e Δt é dado pela diferença de tempo entre os pontos consecutivos. Se Δt for até 99% a mais que a média total da diferença de tempo entre todos os pontos consecutivos, fica com valor 1, se for de 100% até 199% da média fica com valor 2 e assim por diante.

3.1.2 Desacelerações

Ao contrário das acelerações, as situações em que a velocidade de um objeto móvel registrada nos pontos de sua trajetória diminui de um ponto para outro, caracteriza-se como uma desaceleração.

Definição 3 (desaceleração) A desaceleração é uma diminuição na velocidade V de um ponto p , tal que: $V_p < V_{p-1}$ e $V_{p-1} < V_{p-2}$, onde V_p é a velocidade do ponto p , V_{p-1} é a velocidade do ponto anterior ao ponto p e V_{p-2} é a velocidade do ponto anterior ao ponto $p-1$.

Quando a diminuição da velocidade entre dois pontos consecutivos é maior que um determinado percentual, então caracteriza-se uma desaceleração brusca. No exemplo da figura 6, os pontos dos instantes $[t_{10}$ e $t_{11}]$ e $[t_{12}$ e $t_{13}]$ apresentam desacelerações, mas se for vista a velocidade nos pontos e ter como premissa que se a diminuição da velocidade for maior que 10% será considerada brusca, tem-se que os pontos nos instantes t_{10} e t_{11} não representam uma desaceleração brusca, pois a velocidade não diminui 10%, enquanto que nos instantes t_{12} e t_{13} a diminuição é superior à 10%, que seguindo a premissa, seria uma desaceleração brusca.

A principal diferença entre uma desaceleração normal e uma desaceleração brusca é o tempo. Em uma desaceleração brusca o objeto vai diminuir a mesma quantidade de velocidade que em uma desaceleração normal, mas em menos tempo.

Uma desaceleração normal se caracteriza por um percentual aceitável na diminuição de velocidade em um segundo. Quando este percentual de diminuição é muito elevado, caracteriza-se uma desaceleração brusca. Entretanto, da mesma maneira que nas

acelerações, a diminuição do percentual de velocidade a cada segundo quando um objeto está quase parado é muito maior do que quando o objeto já está em alta velocidade. Em outras palavras, uma desaceleração é maior e mais brusca quando o objeto está encerrando o movimento em uma velocidade baixa. Assim sendo, também para evitar que uma desaceleração normal de 2 Km/h para 1 Km/h em um segundo seja considerada brusca, foi utilizada a mesma escala de velocidades com pesos diferentes das acelerações conforme já ilustrado na tabela 1.

O funcionamento do grau de importância é o mesmo que nas acelerações, funcionando para multiplicar o percentual da desaceleração, por exemplo, se for considerado uma diminuição de 20% e um ponto estiver com velocidade de 6 Km/h, a diminuição real limite seria ao invés de 1,2, o valor 4,8, pois seria $1,2 \cdot 4$. Com isso o objeto teria de estar a menos de 1,2 Km/h para ser caracterizada uma desaceleração brusca. Foram feitos testes também para verificar se as faixas de valores se enquadravam nas situações reais, e com essas faixas de valores os resultados foram satisfatórios.

Considerando então a escala aceitável de desaceleração normal apresentada na tabela 1, é possível definir uma desaceleração brusca.

Definição 4 (desaceleração brusca) A desaceleração brusca é uma diminuição na velocidade V de um ponto p , tal que $V_p > (V_{p+1} + (V_{p+1} \cdot (A \cdot S))) \cdot \Delta t$ e $V_{p+1} > (V_{p+2} + (V_{p+2} \cdot (A \cdot S))) \cdot \Delta t$, onde V_p é a velocidade no ponto p , V_{p+1} é a velocidade no ponto seguinte ao ponto p , V_{p+2} é a velocidade no ponto seguinte ao ponto $p+1$, A é o percentual de desaceleração aceitável para o objeto móvel, sendo que este A varia de 0 a 1, S é o peso atribuído a velocidade de acordo com a tabela 1 e Δt é dado pela diferença de tempo entre os pontos consecutivos.

Se Δt for até 99% a mais que a média total da diferença de tempo entre todos os pontos consecutivos, fica com valor 1, se for de 100% até 199% da média fica com valor 2 e assim por diante.

3.2 Mudanças de Direção

Mudanças na direção da trajetória de um objeto móvel também podem caracterizar o comportamento do indivíduo. Podemos dizer que um objeto que muda a sua direção em grande escala e em alta velocidade caracteriza um mau comportamento e uma situação anormal. Nesta seção, definiremos as mudanças de direção e as mudanças bruscas de direção com alta velocidade.

Uma mudança normal de direção é uma situação em que um objeto muda de direção e sua velocidade não está alta a ponto de gerar desconforto ou que ainda venha a desviar a trajetória de seu traçado. Para definir uma mudança de direção foi utilizada a função azimuth, definida como segue em [16] e [10].

Definição 5 (Azimuth) É definido como o ângulo horizontal medido no sentido horário partido de um linha de base para o norte. [16].

Definição 6 (Azimuth Postgis): A função azimuth no Postgis recebe como entrada duas geometrias e retorna a diferença da direção media entre essas geometrias com a direção do norte geográfico. [10].

Logo, pode-se definir que a variação da direção numa trajetória utiliza a seguinte definição:

Definição 7 (Mudança de direção): A mudança de direção numa trajetória pode ser calculada por $\text{azimuth}(p1,p2) - \text{azimuth}(p5,p6)$, onde os valores de p são correspondentes aos pontos.

A mudança de direção foi restringida a se repetir em cinco pontos seguidos para caracterizar uma curva, por isso os pontos $p1$, $p2$, $p5$ e $p6$ são considerados na definição. Essa restrição foi baseada em testes realizados pilotando-se veículos e verificando o tempo de mudança das direções baseadas nos dados de GPS. Então, tendo-se cinco pontos seguidos com variação de direção entre si em sequência, acima de um limite definido para o objeto móvel e ultrapassando um valor de velocidade, serão considerados como sendo uma região que possui uma variação de direção anormal. Se um objeto possui muitos pontos de variação de direção em altas velocidades podemos inferir que seu comportamento é anormal.

Para diferenciar as situações onde existe uma mudança brusca de direção, ou seja, uma grande mudança de direção com velocidades altas deve-se considerar além da mudança de direção entre os pontos, também a velocidade nos pontos. Caso se caracterize uma mudança de direção maior que certo valor e a uma velocidade acima de certo percentual, é considerado uma mudança brusca de direção.

Quanto maior a curvatura do trajeto, menor deverá ser a velocidade nos pontos do trajeto, e por isso foi criada uma tabela com os valores de curvaturas e velocidades permitidas para curvas normais, conforme ilustra a tabela 2. Esses valores foram obtidos pilotando um veículo de passeio por curvas e verificando as velocidades e as variações de direção.

Tabela 2 – Valores de curvatura e velocidades aceitáveis

| Ângulo: Graus | Limite de Velocidade: Km/h |
|---------------|----------------------------|
| 0 até 90 | 90 |
| 90 até 180 | 60 |
| 180 até 270 | 40 |
| Maior que 270 | 30 |

Com essa tabela conseguimos dizer que um ângulo de 190° entre o primeiro e o quinto ponto com velocidade de 50KM/h estaria com uma velocidade muito alta para a situação, enquanto que um objeto a uma velocidade de 30 Km/h estaria com velocidade dentro dos padrões para esta mudança de direção. Por fim o cálculo da mudança brusca de direção com velocidades altas ficou da seguinte forma:

Definição 8 (mudança brusca de direção): Uma mudança brusca de direção é caracterizada por $\Delta d > d$ e $d > Dd$ e $Vp > Vd$, onde Δd é a diferença simples de direção calculada pela diferença de duas funções *azimuth* conforme mostrado no conceito de mudança de direção, d é o ângulo de mudança de direção permitido, Dd é a faixa angular que o ponto se classifica, conforme definido na tabela 2, Vp é a velocidade do ponto e o

Vd é a velocidade máxima definida na tabela 2 para a classificação da mudança de direção do ponto.

3.3 Um algoritmo para identificar variações anormais em trajetórias com base na aceleração, desaceleração e mudança de direção

Para a realização deste trabalho foi desenvolvido um algoritmo para identificar comportamentos anômalos em trajetórias, identificando-se sub-trajetórias com acelerações bruscas, desacelerações bruscas e curvas fechadas feitas em alta velocidade.

O algoritmo implementado foi incorporado à ferramenta Weka[11] que já possui um módulo para análise de trajetórias de objetos móveis desenvolvido pelo grupo de banco de dados sob orientação de Vania Bogorny e chamado de Weka-STPM[13].

O algoritmo está definido na listagem 1. A entrada do algoritmo é um conjunto de trajetórias T , um percentual aceitável de aceleração a , um percentual aceitável de desaceleração s e um limite de mudança de direção d .

Entradas:

- (1) *T// Conjunto de trajetórias*
- (2) *a// Percentual aceitável de aceleração*
- (3) *s// Percetual aceitável de desaceleração*
- (4) *d// Ângulo limite de mudança de direção*
- (5) *t//Tabela com os pesos*

Método:

- (6) *Integer total = n° total de trajetórias; // Total de trajetórias.*
 - (7) *For(i =0 ; i<total; i++){ // Laço que vai percorrer todas as trajetórias uma a uma.*
 - (8) *buscaComportamentos(){ //Início do método que busca os pontos que ultrapassam os limites estabelecidos na entrada.*
 - (9) *For(p=0, p< tamanho da trajetória - 2, p++){ //Laço que percorre todos os pontos de uma trajetória, de zero até o último ponto.*
 - (10) *Busca ponto p, p+1 e p+2*
 - (11) *if((Vp + Vp*(A*S)) * Δt < Vp+1 & (Vp+1 + Vp+1*(A*S)) * Δt < Vp+2){ // Fórmula de comparação dos três pontos de aceleração*
 - (12) *insere os dois pontos iniciais da comparação na tabela acceleration*
 - (13) *}*
 - (14) *if(Vp > (Vp+1 + Vp+1*(A*S)) * Δt & Vp+1 > (Vp+2 + Vp+2*(A*S)) * Δt){ // Fórmula de comparação dos três pontos de desaceleração*
 - (15) *insere os dois pontos iniciais da comparação na tabela slowdown*
 - (16) *}*
 - (17) *Busca pontos de ponto p até p+5*
 - (18) *if (> d e d > Dd e Vp > Vd){ //Fórmula de comparação da diferença das direções dos pontos da trajetória*
 - (19) *insere o intervalo de pontos da comparação na tabela direction*
 - (20) *}*
 - (21) *}*
 - (22) *}*
 - (23) *}*
 - (24) *Cria polígonos nas sub-trajetórias de acelerações bruscas usando buffer*
 - (25) *Cria polígonos nas sub-trajetórias de desacelerações bruscas usando buffer*
 - (26) *Cria polígonos nas sub-trajetórias de mudanças bruscas de direção usando buffer*
- Saída:*
- (27) *Sa; // Conjunto de sub-trajetórias com acelerações bruscas*
 - (28) *Ss; // Conjunto de sub-trajetórias com desacelerações bruscas*
 - (29) *Sd; // Conjunto de sub-trajetórias com mudanças bruscas de direção*

Com os valores de entrada definidos, o algoritmo inicia com a busca do número total de trajetórias (linha 6). Para cada trajetória (linha 7), enquanto existirem pontos para a trajetória (linha 9), o algoritmo considera 3 pontos (linha 10) e faz a comparação, se os 3 pontos estão acelerando bruscamente (linha 11), e caso estejam, insere no banco de dados (linha 12).

Continuando, o algoritmo compara os 3 pontos para ver se estão com desaceleração brusca (linha 14), e se estiverem, insere no banco de dados (linha 15). Depois são buscados os 5 pontos (linha 17) para a comparação da mudança de direção (linha 18), e caso a mudança seja brusca, os pontos são inseridos no banco de dados (linha 19).

Ao final do teste com todas as trajetórias são executadas consultas no banco de dados que retornam as sub-trajetórias (linhas) de acelerações, desacelerações e mudanças de direção bruscas. Sobre estes resultados é aplicada a operação *buffer* para a criação de regiões (polígonos) com características de aceleração brusca, desaceleração brusca e mudança brusca de direção.

Depois de executado o algoritmo, os subconjuntos de trajetórias são retornados.

3.4 Implementação

Considerando agora o entendimento de como funciona o algoritmo e seus objetivos, será mostrado o seu funcionamento. Este algoritmo foi implementado no módulo Weka-STPM[13]. O Weka é um pacote desenvolvido pela Universidade de Waikato, em 1993, com o intuito de agregar algoritmos para mineração de dados na área de Inteligência Artificial. Já o Weka-STPM é uma extensão do Weka para a mineração e visualização de dados semânticos de trajetórias, que foi incorporado ao Weka em 2011.

Da mesma maneira como é feito no Weka, o usuário do software deve executá-lo, clicar no botão abrir banco de dados, que abrirá a interface apresentada na figura 7, onde o mesmo fará a conexão ao banco de dados informando os dados de conexão, tais como usuário, senha e nome da base.

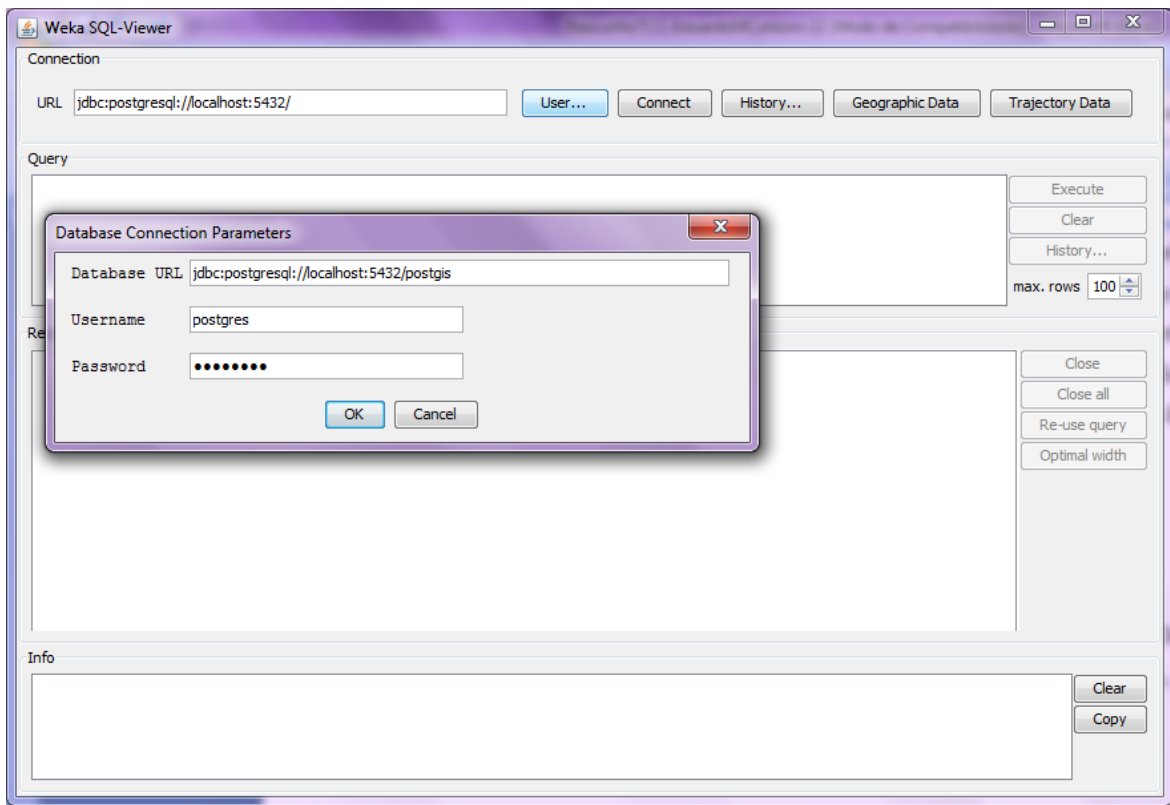


Figura 7 - Tela inicial de conexão

O usuário deverá seguir pelo botão *Trajectory Data* indo em direção à janela de trajetórias ilustrado na figura 8. Nesta interface o usuário deve selecionar um esquema, carregá-lo clicando no botão *load* e selecionar uma tabela de trajetórias no combo *Trajectory Table*.

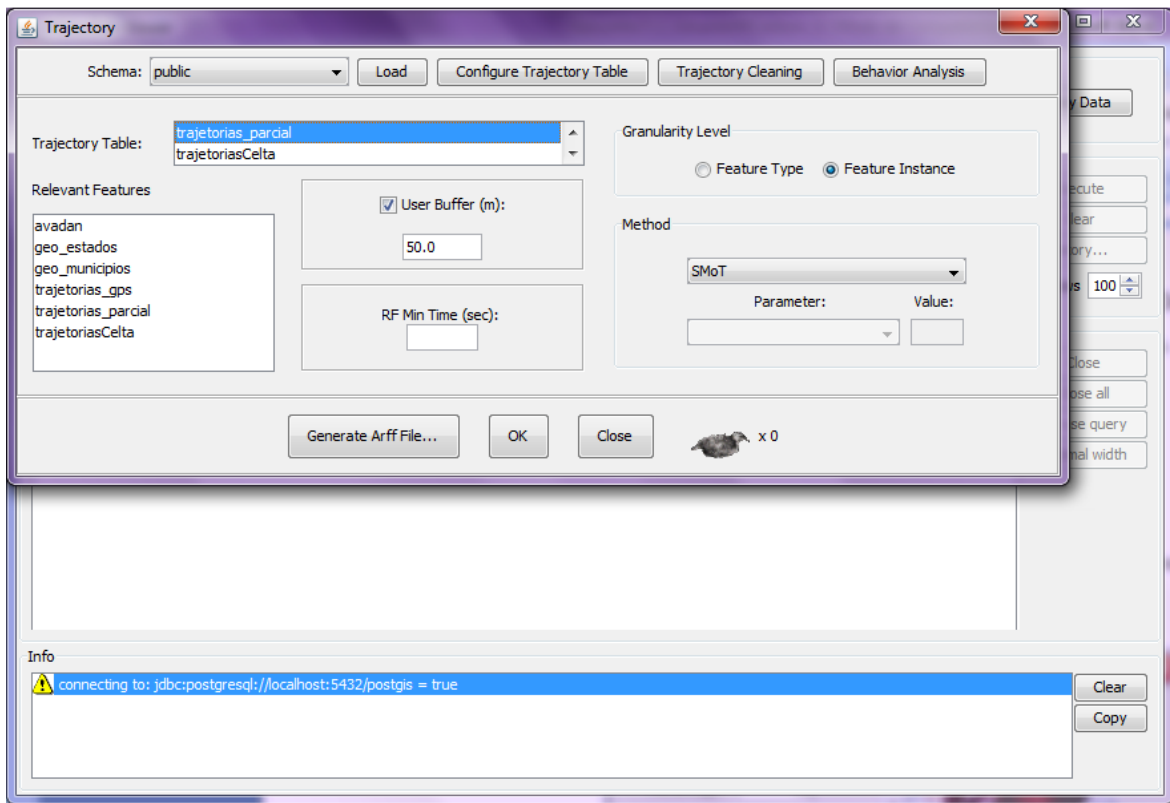


Figura 8 - Tela de trajetórias

Feito isso, o usuário pode então fazer a análise de comportamento das trajetórias pressionando o botão *Behavior Analysis*. Este botão abrirá a interface desenvolvida neste trabalho, ilustrada na figura 9.

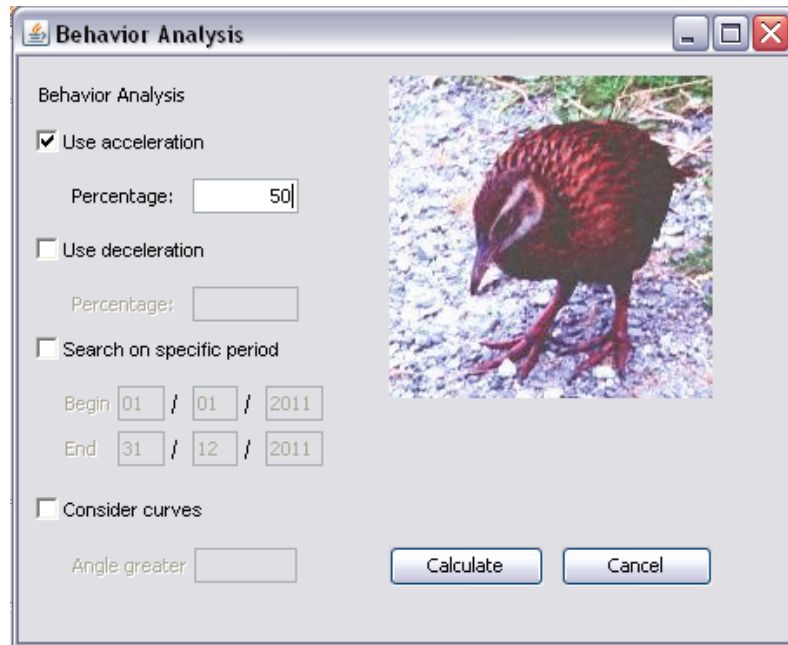


Figura 9– Interface Behavior Analysis.

Nesta interface o usuário poderá escolher análises individuais ou todas ao mesmo tempo, que são:

1 – Buscar por acelerações: É necessário clicar no *check box Use Acceleration* e então inserir o percentual de aceleração máximo permitido no espaço em branco;

2 – Buscar por desacelerações: É necessário clicar *check box Use Deceleration* e então inserir o percentual de desaceleração máximo permitido no espaço em branco;

3 – Fazer a análise em trajetórias que foram geradas em um período: É necessário que seja clicado no *check box Search on specific period* e então deve inserir os valores para data inicial e final;

4 – Buscar por curvas em alta velocidade: É necessário clicar no *check box Consider Curves* e então deve inserir um valor de ângulo máximo de uma curva para ser considerado no cálculo.

4 Experimentos

Neste capítulo apresentamos os resultados de alguns experimentos realizados sobre duas fontes de dados distintas. Uma foi coletada em sua grande maioria em auto estradas no Rio de Janeiro, enquanto que outra foi coletada em rodovias no estado de Santa Catarina.

É importante destacar os softwares utilizados para buscar regiões de trajetórias com acelerações bruscas, desacelerações bruscas e curvas bruscas. Os dados são armazenados no SGBD Postgres extensão Postgis [11], o algoritmo foi implementado em Java e inserido no Weka-STPM [13] e a visualização das informações é realizada através do Quantum-Gis [12], que conecta no BD Postgis.

4.1 Experimentos com dados reais de Santa Catarina

O primeiro conjunto de dados é composto por pontos coletados especialmente para este trabalho, no estado de Santa Catarina. Para a coleta das informações foi utilizado um automóvel e um GPS para obter novos dados. As trajetórias são sempre entre os municípios de Florianópolis (cidade de estudo, trabalho e residência do estudante) a cidade de Orleans e a cidade de Jaguaruna. Na passagem dos pontos para a base de dados foram consideradas as informações *gid*, *tid*, *x*, *y*, *time*, *time_char*, *velocidade* e *the_geom*, onde se tem identificadores (*gid*, *tid*), coordenadas (*x*, *y*), informações de tempo (*time*, *time_char*), informações geográficas (*the_geom*) e *velocidade* do objeto. É possível ter um melhor entendimento desses dados visualizando a figura 10, que mostra o mapa com as cidades do estado de Santa Catarina com as trajetórias sobrepostas.

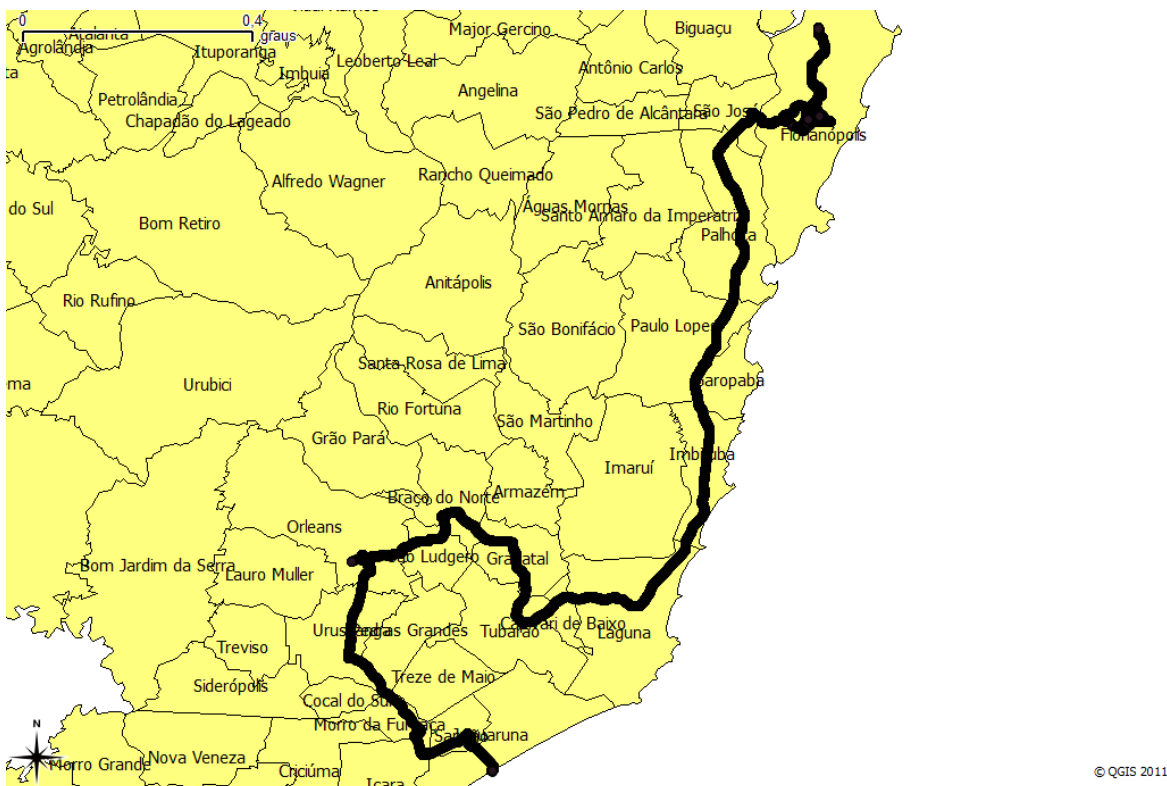


Figura 10 - Trajetórias dos dados reais de Santa Catarina

Os experimentos dessa fonte de dados foram feitos conforme descrito a seguir na tabela 3.

Tabela 3 - Resultados do algoritmo para a fonte de dados de Santa Catarina

| | Acelerações | | | Desacelerações | | | Direções | | |
|---------|--------------------------|--------------------|-------------------|-----------------------------|--------------------|-------------------|------------------|--------------------|-------------------|
| | Percentual Aceleração(%) | Regiões Retornadas | Nº de Trajetórias | Percentual Desaceleração(%) | Regiões Retornadas | Nº de Trajetórias | Ângulo Limite(°) | Regiões Retornadas | Nº de Trajetórias |
| Linha 1 | 10 | 237 | 23 | 20 | 257 | 25 | 45 | 95 | 22 |
| Linha 2 | 20 | 9 | 5 | 30 | 165 | 24 | 180 | 90 | 21 |
| Linha 3 | 30 | 0 | 0 | 50 | 92 | 19 | 270 | 84 | 21 |

Conforme mostrado na tabela 3, foram realizados três experimentos com valores diferentes para os três parâmetros, pois cada um é independente entre si. Primeiramente foi iniciada a execução do algoritmo com valores de aceleração em 10%, desaceleração 20% e

direção com 45°. Estes valores foram escolhidos considerando os seguintes aspectos: acelerações bruscas com um valor baixo, desacelerações bruscas com um valor baixo, onde um valor de desaceleração brusca baixo é maior que uma aceleração brusca, pois um veículo perde velocidade mais rápido do que adquire e o valor de mudança de direção brusca foi considerado como um valor também baixo. Os resultados mostrados na tabela 3, foram acelerações bruscas em 237 sub-trajetórias de 23 trajetórias diferentes, desacelerações bruscas em 247 regiões de 25 trajetórias distintas e mudanças bruscas de direção 95 regiões de 22 trajetórias diferentes.

Depois de vistos os resultados obtidos com as situações anteriores, foram alterados os valores de todos os parâmetros, para novas situações. Os novos valores dos parâmetros foram passados com 20% para acelerações bruscas, 30% para desacelerações bruscas e 180° para as mudanças bruscas de direção. Neste experimento, obteve-se um valor de retorno bem diferente para as acelerações bruscas, que passou anteriormente de 237 para 9 regiões e 5 trajetórias diferentes com 20%, ou seja, houve uma redução significativa no número de regiões com aceleração brusca. As desacelerações bruscas tiveram uma queda de aproximadamente 100 regiões, passado de 257 para 165 regiões e 24 trajetórias distintas. No caso da mudança de direção brusca, pouco se alterou nos resultados, ficando agora com 90 regiões e 21 trajetórias diferentes, conforme ilustrado na tabela 3.

Como ainda não existe uma conclusão de um melhor valor, os testes foram continuados. Foi utilizado para o percentual da aceleração o valor de 30%, para a desaceleração 50% e para a mudança de direção 270°. Esses valores foram escolhidos para demonstrar valores maiores para os parâmetros. No módulo das acelerações o resultado foi nulo, ou seja, não houveram regiões com acelerações bruscas. Nas desacelerações houve uma redução de aproximadamente a metade se comparado ao experimento anterior, ficando nessa vez com 92 regiões e 19 trajetórias diferentes de desacelerações bruscas. Já o número de regiões retornadas no nível de mudanças bruscas de direção reduziu pouco, passando para 84 regiões e continuou em 21 trajetórias, conforme visto na tabela 3.

Analisando os resultados podemos verificar a ocorrência de um grande número de desacelerações bruscas. Com os resultados descritos acima, buscou-se uma trajetória que estivesse no cotidiano dos envolvidos no desenvolvimento deste trabalho para poder

comparar os resultados do trabalho com os locais que deveriam ter comportamento anormal ao longo do percurso, e com isso, foi escolhida a trajetória com identificador 34, que está localizada na cidade de Florianópolis, mais especificamente entre os bairros Trindade e Saco Grande. A figura 11 ilustra a trajetória 34 com possíveis locais onde sabe-se que devem haver movimentos bruscos.

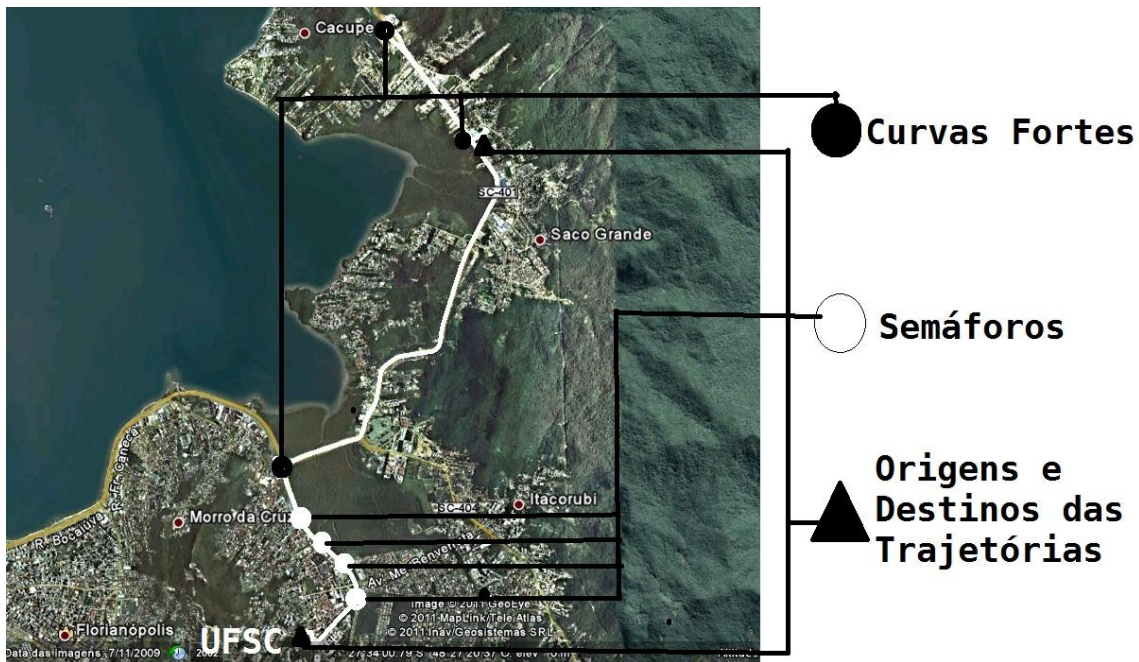


Figura 11 - Trajetória 34 com possíveis locais de comportamento anormal

Tendo conhecida a região e os possíveis locais onde existem comportamentos anormais é apresentada a figura 12, que ilustra a mesma trajetória com as sub-trajetórias que apresentaram comportamento anormal identificadas pelo algoritmo. As regiões de polígonos em preto correspondem às acelerações bruscas, as regiões de polígonos em branco representam a desacelerações bruscas e as de polígonos em cinza mudanças bruscas de direção. Ainda na mesma figura 12, existe a separação por experimento A, B e C, onde a primeira corresponde aos testes pertencentes à linha 1 da tabela 3, a segunda região corresponde à linha 2 da tabela 3 e a terceira região à linha 3 da tabela 3. Para as representações foram criados *buffers* ao redor de todas as regiões de acelerações bruscas, desacelerações bruscas e mudanças bruscas de direção. Criar um *buffer* de 20 metros ao redor de uma região é o mesmo que aumentar sua área num raio de 20 metros. O *buffer* de um ponto criaria um círculo ao redor do ponto com um raio no valor do *buffer*. Para este

estudo, as acelerações bruscas foram *bufferizadas* com valor 45, as desacelerações bruscas com valor 30 e as mudanças bruscas de direção com valor 15. Esses valores foram escolhidos para ter uma visualização clara das regiões com diferente semântica, pois como elas se sobrepõem em alguns momentos, é necessário que algumas sejam maiores que as outras para diferenciá-las.

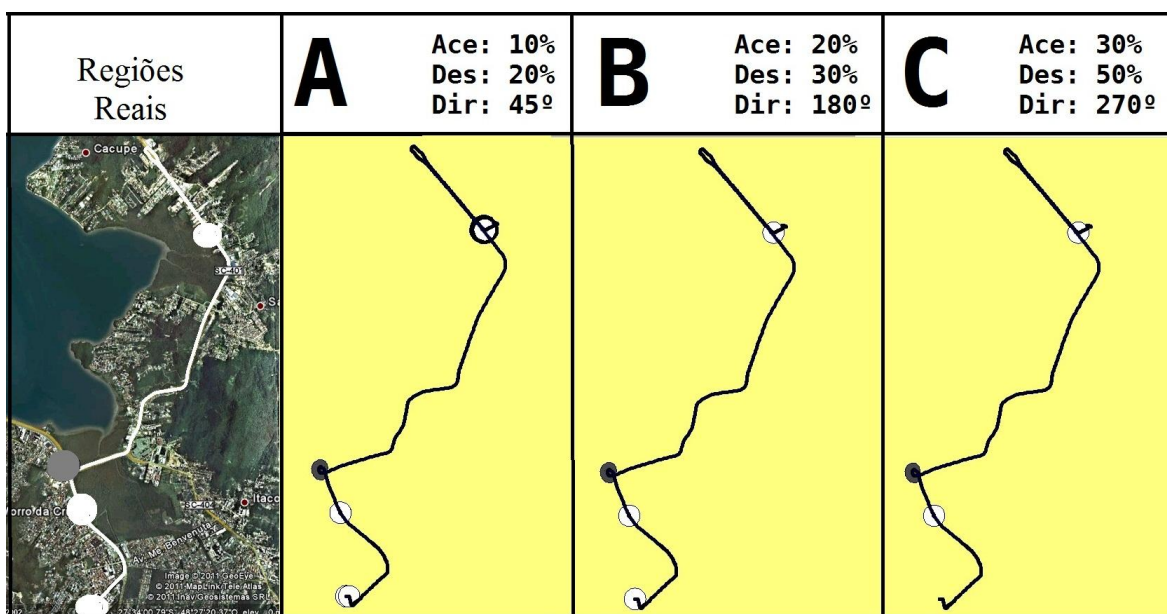





Figura 12 – Trajetória 34 com parâmetros diferentes para busca de comportamentos anormais

-  Aceleração Brusca
-  Desaceleração Brusca
-  Mudança Brusca de Direção

Para o estudo, foi produzida a trajetória 34 com quatro regiões de comportamento anormal. Na imagem 12, temos na lateral esquerda representados os locais onde o algoritmo deve retornar os comportamentos anormais, pois foram nessas regiões que o condutor simulou desacelerações e curvas fortes. Também conseguimos perceber que conforme o aumento do valor dos parâmetros, menor é o número de regiões com comportamento anormal identificado pelo algoritmo. Analisando os valores mostrados na tabela 3 e as visualizações da imagem 12, consegue-se perceber que o experimento A possui uma região de aceleração brusca para a trajetória 34, porém como não deveriam ter acelerações bruscas, com limite de aceleração imposto de 10%, a região retornada possui

uma baixa aceleração, passando de 3 Km/h para 5 Km/h em um segundo. O experimento *C* exclui de seus resultados um dos destinos com desaceleração brusca. Logo, cruzando as informações vistas na figura 11 com os resultados na figura 12, escolheram-se os parâmetros do experimento *B*, pois eles correspondem a totalidade dos locais com comportamentos anormais presentes na figura 12, que neste caso são semáforos, curvas fortes, origens e destinos da trajetória.

Depois de vistas as regiões onde houveram acelerações bruscas, desacelerações bruscas e mudanças bruscas de direção para a trajetória 34, é feita uma análise em todas as trajetórias para encontrar as regiões onde ocorreram mais de um desses comportamentos em trajetórias diferentes, para depois poder-se afirmar se os locais em que a trajetória 34 apresentou mau comportamento acontecem na maioria das trajetórias.

A figura 13 ilustra a mesma trajetória da figura 12, mas agora tendo ao lado esquerdo (*A*) a trajetória sobreposta por suas regiões com comportamento anômalo, e ao lado direito (*B*) as regiões onde aconteceram as situações de mau comportamento mais de uma vez em trajetórias diferentes. Na figura 13, as regiões em branco representam as sub-trajetórias com desaceleração brusca e as regiões em preto representam as sub-trajetórias com mudanças bruscas de direção. Na trajetória 34, não houve aceleração brusca, e neste trajeto não houve esse comportamento em mais de uma trajetória.

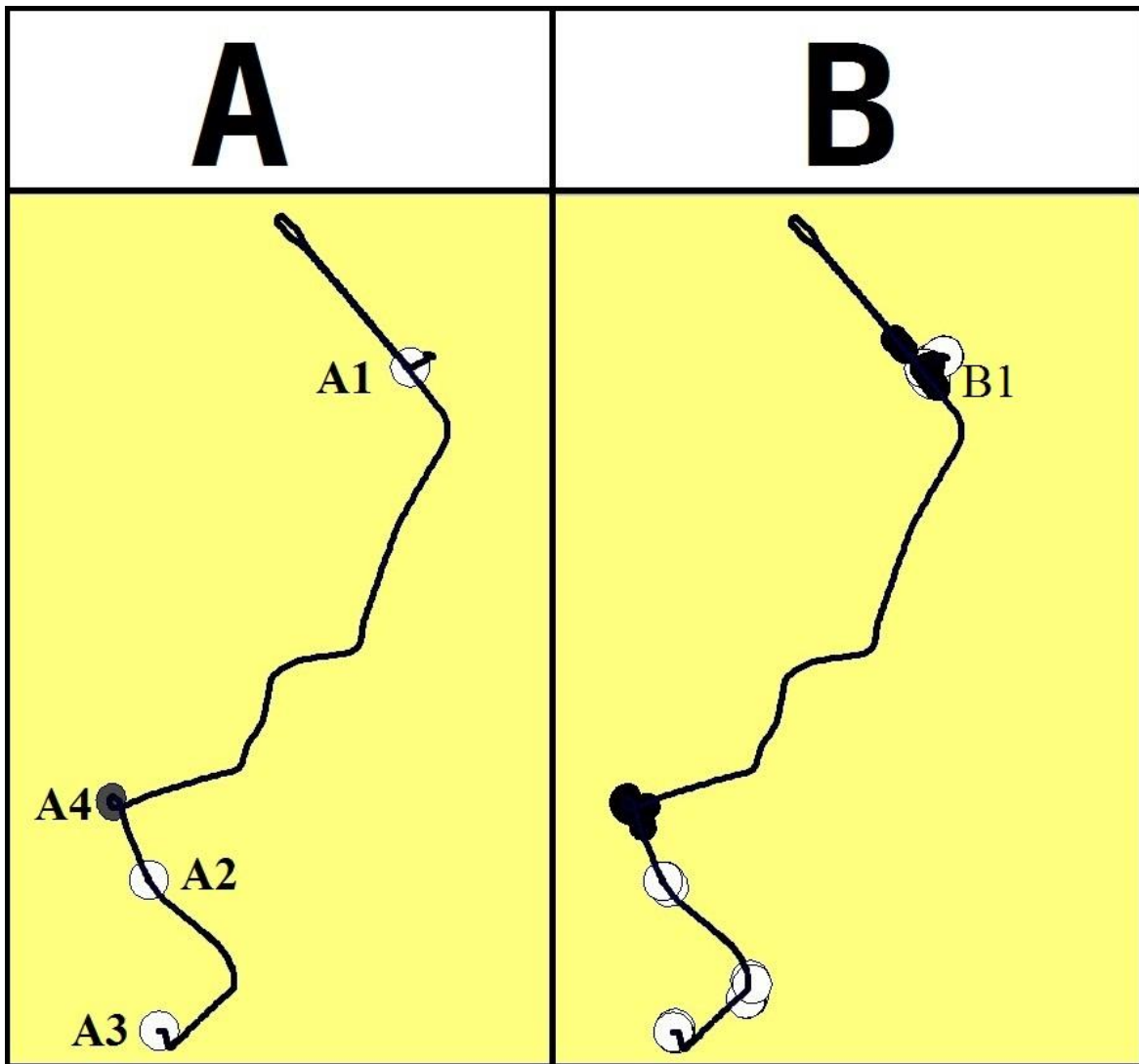


Figura 13 - Regiões de mau comportamento na trajetória 34(A) e intersecção de polígonos com comportamentos anormais da fonte de dados de Santa Catarina(B)

○ Desaceleração Brusca
● Mudança Brusca de Direção

Tendo conhecimento da região da trajetória, é possível afirmar que na figura 13 parte A, a região de desaceleração brusca mais ao sul(A3) corresponde ao destino da trajetória, a região de desaceleração (A2) corresponde ao local de um semáforo, a região de desaceleração brusca mais ao norte (A1) corresponde ao momento em que o veículo reduz a velocidade para esperar os carros passarem e entrar na rodovia e a região de mudança brusca de direção presente na imagem está localizada num local conhecido localmente como “elevado do CIC”. Vistas as regiões onde existiram mais de uma trajetória com

comportamento anormal na figura 13 parte *B*, é possível afirmar que as regiões onde acontecem desacelerações bruscas são sempre próximas à semáforos e destino das trajetórias, pois sempre acontecem paradas bruscas nessas situações. Com isso, podemos concluir que o grande motivo das desacelerações bruscas é o perfil do condutor do objeto que tende a reduzir bruscamente à velocidade em suas paradas em semáforos e destinos, exceto na entrada da rodovia, que caso esteja vindo algum carro, força o condutor a reduzir bruscamente a velocidade para evitar uma colisão.

Também é possível afirmar que numa determinada curva, várias trajetórias apresentaram uma mudança brusca de direção, neste caso o “elevado do CIC”, tanto na ida, quanto na volta do percurso, e também numa região mais ao norte (B1) existem mais de uma trajetória com mudança brusca de direção, que é o destino do objeto móvel, que neste caso ele sai da rodovia e adentra ao destino em alta velocidade.

Por fim, nos experimentos desta fonte de dados, a trajetória 34 não possui um mau comportamento em relação às outras trajetórias da mesma amostra de dados, pois todas as regiões com comportamento anormal desta trajetória aparecem em outras trajetórias. Conseguimos ver também nas trajetórias que os maiores motivos das regiões onde existem desacelerações bruscas são os semáforos e em determinadas curvas o motorista costuma percorrer em alta velocidade.

4.1 Experimentos com a fonte de dados do Rio de Janeiro

A segunda coleção de dados é composta por pontos coletados através de veículos governamentais na cidade do Rio de Janeiro. A estrutura dos dados possui as seguintes informações: *gid*, *tipo*, *e*, *n*, *data*, *hora*, *velocidade*, *direção*, *time*, *tid*, e *the_geom*, onde as colunas se referem a identificadores (*gid*, *tid*), coordenadas (*e*, *n*), informações de tempo (*data*, *hora*, *time*), localização geográfica (*the_geom*) e outras informações do objeto (*velocidade*, *direção*, *tipo*). Algumas particularidades dessa fonte de dados: é composta por 100 trajetórias, os dados foram coletados em sua maioria em auto-estradas e as velocidades tem valores mais altos no início e no fim dos dias, tendo, provavelmente, como motivo à ida e a volta à algum destino. Podem-se visualizar as trajetórias desse

conjunto de dados na figura 14 que possui o mapa do Rio de Janeiro, sobreposto pelas ruas e por cima todos os pontos coletados gerando as trajetórias.

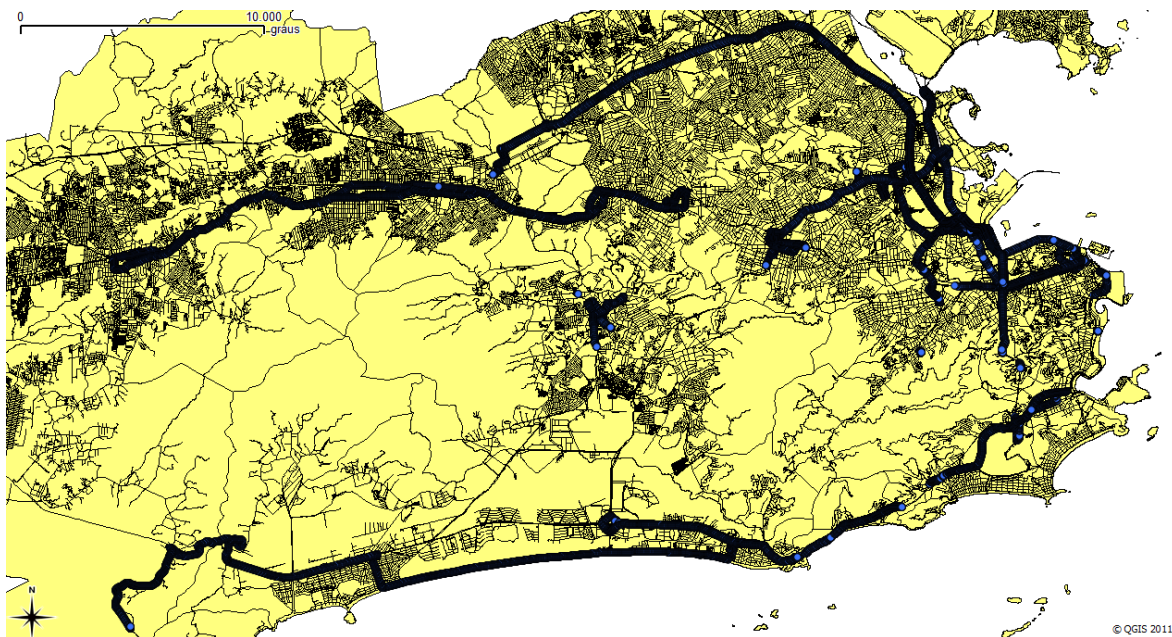


Figura 14 - Trajetórias da fonte do Rio de Janeiro

Os primeiros experimentos que foram realizados com esta fonte de dados estão na tabela 4.

Tabela 4 – Resultados do algoritmo para a fonte de dados do Rio de Janeiro

| | Acelerações | | | Desacelerações | | | Direções | | |
|---------|--------------------------|--------------------|-------------------|-----------------------------|--------------------|-------------------|------------------|--------------------|-------------------|
| | Percentual Aceleração(%) | Regiões Retornadas | Nº de Trajetórias | Percentual Desaceleração(%) | Regiões Retornadas | Nº de Trajetórias | Ângulo Limite(°) | Regiões Retornadas | Nº de Trajetórias |
| Linha 1 | 10 | 1426 | 89 | 20 | 2143 | 89 | 45 | 257 | 59 |
| Linha 2 | 20 | 100 | 39 | 30 | 1455 | 86 | 180 | 255 | 57 |
| Linha 3 | 30 | 02 | 01 | 50 | 798 | 81 | 270 | 243 | 54 |

Conforme mostrado na tabela 4, foram feitos três experimentos com valores diferentes para os três parâmetros da mesma maneira que para a fonte de dados anterior. Primeiramente foi iniciada a execução do algoritmo com valores de aceleração em 10%, desaceleração 20% e direção com 45°. Os resultados, mostrados na tabela 4, foram acelerações bruscas com 1426 sub-trajetórias de 89 trajetórias diferentes, desacelerações

bruscas em 2143 regiões de trajetórias distintas e mudanças bruscas de direção em 257 regiões de 59 trajetórias diferentes.

Depois de vistos os resultados obtidos com as situações anteriores, foram alterados os valores de todos os parâmetros, para novas situações. Os novos valores dos parâmetros foram passados com 20% para acelerações bruscas, 30% para desacelerações bruscas e 180° para as mudanças bruscas de direção. Neste experimento se obteve um valor de retorno bem diferente para as acelerações bruscas, que passou anteriormente de 1426 para 100 regiões de 39 trajetórias diferentes com 20%, ou seja, houve uma redução significativa no número de regiões com aceleração brusca, pois este aumento do percentual de aceleração caracteriza movimentos mais bruscos. As desacelerações bruscas tiveram uma queda de aproximadamente um terço, passado de 2143 para 1455 regiões pertencentes à 86 trajetórias distintas. No caso da mudança de direção brusca, pouco se alterou nos resultados, ficando agora com 255 regiões de 57 trajetórias diferentes, conforme ilustrado na tabela 4.

Os resultados parecem seguir o mesmo que aconteceu em Santa Catarina. Mas como ainda não existe uma conclusão sobre o melhor valor para os parâmetros, os testes foram continuados. Foi utilizado para o percentual da aceleração o valor de 30%, para a desaceleração 50% e para a mudança de direção 270°. No módulo das acelerações o resultado foi minúsculo, retornando apenas 2 regiões de apenas 1 trajetória. Nas desacelerações houve uma redução de aproximadamente a metade das regiões se comparado ao teste anterior, ficando desta vez com 798 regiões de desacelerações bruscas pertencentes à 81 trajetórias distintas. Já o número de regiões retornadas no nível de mudanças bruscas de direção reduziu pouco, passando para 243 regiões de 54 trajetórias diferentes, conforme visto na tabela 4.

Com os resultados descritos acima, buscou-se a trajetória que apresentava mais sub-trajetórias com anomalias, que neste caso é a de identificador 26, para visualizarmos no Quantum Gis. A figura 15 ilustra parte da trajetória 26, com as sub-trajetórias que apresentaram comportamento anormal, onde as regiões de polígonos em branco correspondem às acelerações bruscas, as regiões de polígonos em cinza representam a desacelerações bruscas e as de polígonos pretos as mudanças bruscas de direção. Ainda na

mesma figura existe a separação por experimento A, B e C, onde o primeiro corresponde aos testes pertencentes à linha 1 da tabela 4, o segundo experimento corresponde à linha 2 da tabela 4 e a terceiro experimento à linha 3 da tabela 4. Nesta fonte de dados as regiões também foram *bufferizadas*, sendo que as acelerações bruscas foram *bufferizadas* com valor 45, as desacelerações bruscas com valor 30 e as mudanças bruscas de direção com valor 15.

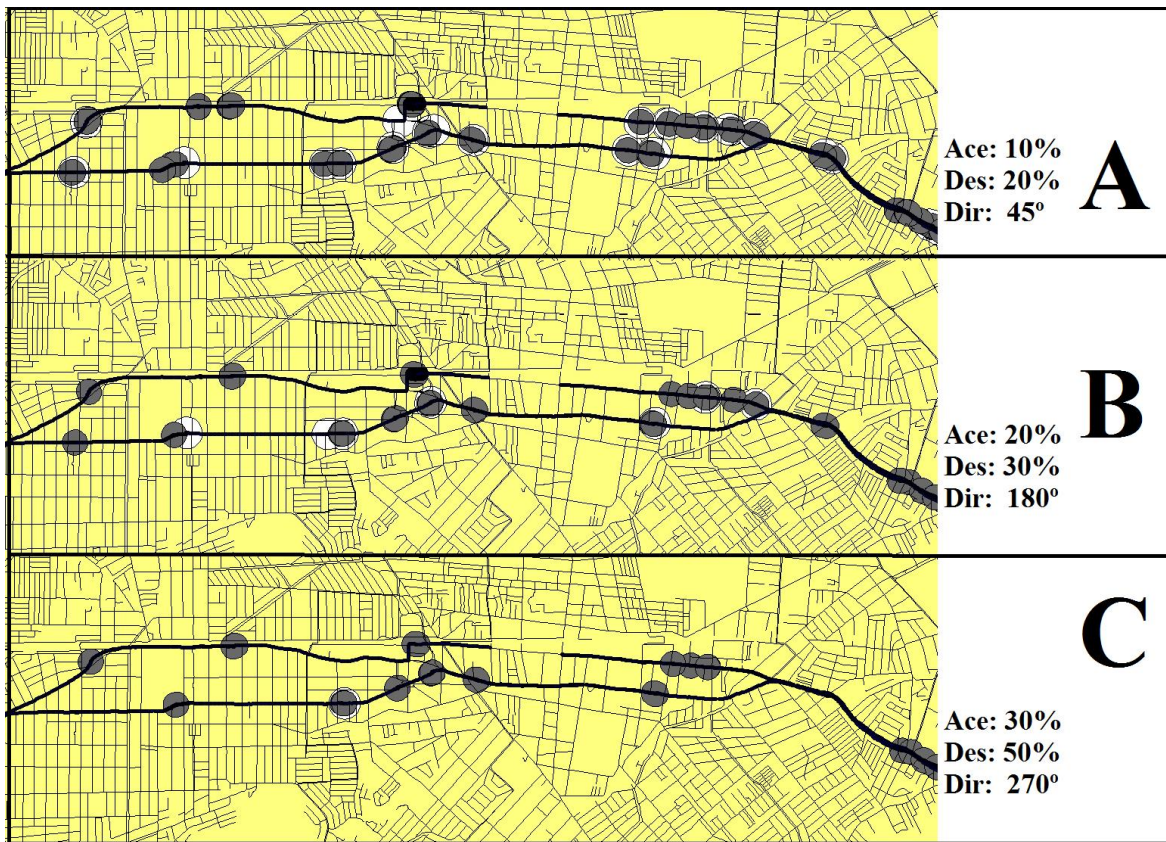





Figura 15 – Trajetória 26 com parâmetros diferentes para busca de comportamentos anormais

-  Aceleração Brusca
-  Desaceleração Brusca
-  Mudança Brusca de Direção

Na imagem 15, conseguimos perceber que conforme o aumento do valor dos parâmetros, menor é o número de regiões com comportamento anormal, pois mais brusco será o comportamento. Considerando que a região mostrada na figura 15 não é conhecida, não é possível cruzar os resultados dos experimentos com as informações semânticas dos locais. Portanto, para a seqüência dos experimentos foram utilizados os parâmetros do experimento *B*, que é 20% para acelerações bruscas, 30% para desacelerações bruscas e 180° para mudanças bruscas de direção. Estes valores foram escolhidos por serem os valores que na fonte de dados de Santa Catarina caracterizaram os locais reais de comportamento anormal.

Para analisar algumas trajetórias e descobrir seu comportamento, foram buscadas 3 trajetórias que apresentavam os 3 comportamentos anormais, e a partir disso verificado quantas regiões com comportamento anormal de cada trajetória acontece em outras trajetórias.

A primeira é a com identificador 16, que possui 20 regiões com comportamento anormal. Ela foi escolhida por apresentar poucas regiões anormais, mas ser uma das que apresenta as 3 situações de mau comportamento. Das 20 regiões com comportamento anormal da trajetória 16, 15 delas se repetem em mais outras 4 trajetórias no mínimo, portanto, podemos classificar que este motorista apresenta um provável comportamento anormal em virtude da rodovia, pois a maioria dos locais que ele possui comportamentos de mau condutor são onde outros condutores também possuem. Isso pode ser visto na figura 16, que apresenta na parte superior as regiões de mau comportamento da trajetória 16 e na parte inferior as regiões que se repetiram em mais de 4 trajetórias distintas.



Figura 16 - Regiões anômalas da trajetória 16 e em mais de outras 4 trajetórias

○ Região com comportamento anormal

A segunda trajetória para análise de comportamento é a de identificador 49, que apresenta 32 regiões de mau comportamento. Ela foi escolhida dentre as outras por apresentar um número intermediário de regiões de mau comportamento, porém tendo as três situações de comportamento anormal. Das 32 regiões com comportamento anormal da trajetória 49, 30 delas se repetem em mais outras 4 trajetórias no mínimo, portanto também podemos classificar que este motorista apresenta um comportamento anômalo forçado pela rodovia, pois a maioria dos locais que ele possui comportamentos de mau condutor são onde outros condutores também possuem. Isso pode ser visto na figura 17, que apresenta

ao lado esquerdo as regiões de mau comportamento da trajetória 49 e ao lado direito as regiões que se repetiram em mais de 4 trajetórias distintas.

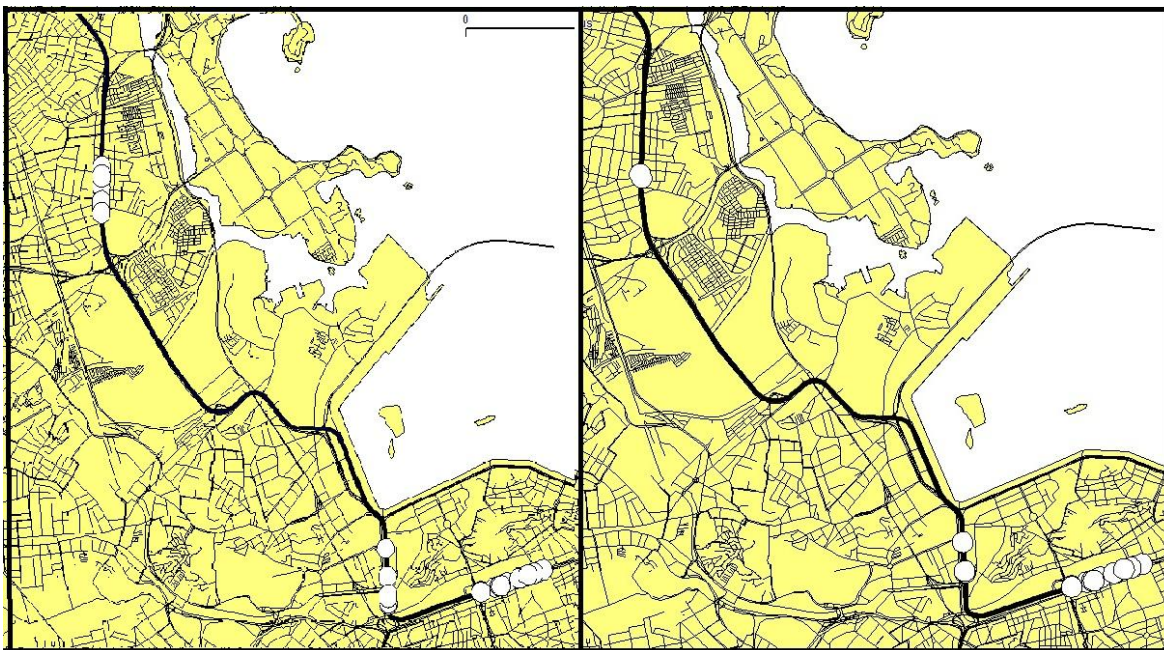


Figura 17 - Regiões anômalas da trajetória 49 e em mais de outras 4 trajetórias

○ Região com comportamento anormal

Por fim foi analisada a trajetória com identificador 26, que possui 137 regiões com comportamento anormal. Ela foi escolhida por ser uma das que apresenta as 3 situações de mau comportamento e a que mais possui esses comportamentos. Das 137 regiões com comportamento anormal da trajetória 26, nenhuma delas se repetem em mais outras 4 trajetórias, no mínimo. O fato acontece porque o trajeto da trajetória 26 não possui mais do que uma trajetória no mesmo local, apenas a trajetória 27 percorre o mesmo local, ou seja, todas as outras 98 trajetórias são percorridas em locais diferentes. Logo a trajetória 26 pode ser considerada uma trajetória anormal, pois o comportamento que ela apresenta não se repete em outras trajetórias.

Com a execução do algoritmo para as duas fontes de dados foi possível concluir que os valores de retorno são válidos, pois na fonte de dados de Santa Catarina, onde se conhecem os trajetos e os comportamentos anômalos, os resultados coincidiram com locais que tem tendências para acontecer comportamentos anormais. Já a segunda fonte de dados

foram vistas trajetórias separadas e definidos seus comportamentos individuais em relação ao total de trajetórias, tendo situações com comportamento regular e com comportamento anormal.

5 Conclusões

Com a popularização e a acessibilidade de aparelhos de coleta de informações espaço-temporais como aparelhos de telefone celular, GPS ou qualquer outro dispositivo móvel, aumentou significativamente o volume de dados espaço-temporais, chamados de trajetórias de objetos móveis. Diferentes métodos foram desenvolvidos para encontrar pontos importantes nas trajetórias, tais como IB-SMoT, CB-SMoT e DB-SMoT.

Também foram desenvolvidos métodos que através da análise de comportamento de trajetórias tentam classificar seus tipos como, por exemplo, turísticas, de trabalho, entre outras, através da identificação da origem, destino e locais por onde passaram. Mas nada foi proposto com o objetivo de caracterizar o comportamento dos objetos móveis.

Esse trabalho propôs classificar o comportamento de objetos móveis em normais ou anormais, com base em acelerações bruscas, desacelerações bruscas e mudanças bruscas de direção ao longo das trajetórias. Foi criado um algoritmo que retorna os locais de acelerações bruscas, desacelerações bruscas e mudanças bruscas de direção. Com este algoritmo foram realizados testes em bases de dados distintas, com o objetivo de definir o comportamento dos objetos e validar este trabalho. Os experimentos foram iniciados com dados de trajetórias de Santa Catarina, onde o algoritmo encontrou as regiões onde houve mau comportamento. Também foi verificado se os comportamentos anormais se repetiam em mais de uma trajetória na mesma região. Com isso foi possível verificar se o objeto móvel tinha um comportamento anormal em virtude da maneira como dirige ou se a razão era a rodovia.

Em relação à trabalhos futuros, pretende-se continuar com o estudo para que o usuário que utilizar o sistema possa apenas selecionar algumas trajetórias de um único condutor, e o algoritmo trazer como retorno uma afirmação do comportamento do objeto móvel, sem que o usuário tenha de olhar e decidir se a trajetória é normal ou não. Além disso, pretende-se definir novas medidas que garantam com mais certeza se um indivíduo tem ou não comportamento anormal durante sua trajetória.

6 Referências

- [1] Spaccapietra, S., Parent, C., Damiani, M.L., Macedo, J.A.F., Porto, F., Vangenot., C. A conceptual view on trajectories. Data and Knowledge Engineering 2007.
- [2] Alvares, L.O., Bogorny, V., Kuijpers, B., Macedo, J.A.F., Moelans, B. A model for enriching trajectories with semantic geographical information. In GIS'07: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, New York, NY, USA, 2007. ACM Press.
- [3] Palma, A.T., Bogorny, V. and Alvares, L.O., 2008, A Clustering-based approach for discovering interesting places in trajectories. In Proceedings of the ACMSAC, Fortaleza, Brazil (New York, NY, USA: ACM Press).
- [4] Manso, J. A.; Times, V. C.; Oliveira, G.; Alvares, L.O.; Bogorny, V. DB-SMoT: a Direction-based spatio-temporal clustering method. Fifth IEEE International Conference on Intelligent Systems (IEEE IS 2010), 2010 (accepted for publication).
- [5] CHIECHELSKI, G. O. ; BOGORNY, V. Uma Extensão do PostGIS para a Geração Automática de Trajetórias Semânticas. 2008.
- [6] Avancini, H. M. Análise da Redução de Dados em Trajetórias de Objetos Móveis. 2010
- [7] Aplicativo Weka. Disponível em URL: <http://www.cs.waikato.ac.nz/~ml/weka/>
- [8] Reportagem revista Época. Disponível em URL: <http://epocanegocios.globo.com/Revista/Common/0,,EMI85936-16366,00-O+NOVO+GARIMPO+DA+ERA+DIGITAL.html>
- [9] Conceitos de Reality Mining. Disponível em URL: <http://reality.media.mit.edu>
- [10] Bogorny, V.; Heuser, C.A; Alvares, L.O. A conceptual data model for trajectory data mining. Sixth International Conference on Geographic Information Science (GIScience 2010), 2010 (accepted for publication).
- [11] – Postgis - <http://postgis.refractorions.net/>
- [12] – QuantumGis - Documentação sobre o aplicativo Quantum Gis, versão 1.4.0. Disponível em URL: <http://www.qgis.org/pt/documentation/manuals.html>
- [13] – Weka- STPM - Alvares, Luis O. ; PALMA, Andrey ; Oliveira, G. ; BOGORNY, V. Weka-STPM: from trajectory samples to semantic trajectories. In: Workshop de Software Livre, 2010, Porto Alegre. WSL, 2010.
- [14] – Knowledge Discovery and Data Mining Laboratory – URL: <http://www-kdd.isti.cnr.it/>
- [15] – Renso, Chiara ; Giannotti, Fosca ; Nanni, Mirco ; Pedreschi, Dino ; Trasarti, Roberto. Mining Mobility Behavior from Trajectory Data.
- [16] – Wikipedia – conceito de azimuth. Disponível em URL: <http://en.wikipedia.org/wiki/Azimuth>