

## **TRACTS: Um método para a classificação de trajetórias de objetos móveis usando séries temporais**

**Irineu Jr. Pinheiro dos Santos<sup>1</sup>, Luis Otavio Alvares<sup>1,2</sup>**

<sup>1</sup>Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

{ijpsantos,alvares}@inf.ufrgs.br

<sup>2</sup>Depto. de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)  
Caixa Postal 476 – 88.040-900 – Florianópolis – SC – Brasil

alvares@inf.ufsc.br

**Abstract.** *Trajectory classification is a relatively new research issue, but some methods have already been proposed. Most of them are developed for a specific application. A few are more general and can be applied to several domains. This paper presents a new trajectory classification method that transforms trajectories into time-series, in order to obtain more discriminative features for classification. The results of the processed experiments have been compared with the TraClass algorithm and have presented better results.*

**Resumo.** *A classificação de trajetórias é um assunto relativamente novo de pesquisa, mas alguns métodos já foram propostos. A maioria destes métodos foi desenvolvido para uma aplicação específica. Poucos propuseram um método mais geral, aplicável a vários domínios ou conjuntos de dados. Esse trabalho apresenta um novo método de classificação que transforma as trajetórias em séries temporais de forma a obter características mais discriminativas para a classificação. Os resultados dos experimentos realizados foram comparados com os obtidos pelo método TraClass e se mostraram, em geral, superiores.*

### **1. Introdução**

Nos anos recentes, observa-se o crescimento do uso de sistemas baseados em localização, tais como GPS, triangulação de rádio e de celulares, sistemas wireless e RFID. Esses sistemas possibilitaram e facilitaram o rastreamento de objetos móveis com uma resolução de apenas alguns metros no espaço, ou com maior precisão ainda, em alguns casos.

O aumento do uso de tais sistemas está levando à maior disponibilidade de dados espaço-temporais de pessoas, veículos e animais, por exemplo. Essa coleção de dados espaço-temporais representa um ambiente extremamente rico para análise, pois dentro dessa grande massa de dados estão diversas informações a respeito de comportamentos individuais e coletivos dos objetos móveis envolvidos.

Uma das formas de se extrair informações interessantes a partir de um conjunto de dados espaço-temporais é através da análise de trajetórias, que podem ser construídas a partir dos dados coletados. Uma trajetória é uma sequência de pontos  $(id, x, y, t)$ , em que  $x$  e  $y$  indicam a posição do objeto móvel identificado por  $id$  no tempo  $t$ . A

mineração e a descoberta de conhecimento a partir dessas trajetórias têm emergido como um tópico frequente em pesquisas acadêmicas e em aplicações práticas.

A análise de diversas trajetórias de objetos móveis permite, por exemplo, realizar a identificação de lugares importantes, ou POIs (*Points of Interest*), para esse determinado objeto móvel. Isso possibilita a armazenagem de históricos de sua localização em cada um desses POIs [Hariharan and Toyama 2004], descobrir a sequência de locais que geralmente são visitados pelo objeto móvel [Zhou et al. 2007], realizar a predição de destinos possíveis [Ashbrook and Starner 2003] e até mesmo permitir que um usuário seja avisado em tempo real que está pegando um ônibus errado [Liao et al. 2006]. Focando-se nas propriedades geométricas das trajetórias, existe a possibilidade de detecção de trajetórias similares, estabelecendo correlação entre objetos móveis [Tiakas et al. 2009], padrões de congruência [Gudmundsson et al. 2007] e predições de comportamento, como, por exemplo, no tempo de viagem de um veículo [Tiesyte and Jensen 2008]. Indo além da análise das propriedades geométricas das trajetórias e considerando-se também os objetos geográficos estáticos em suas proximidades, tais como hotéis, pontos turísticos de uma cidade e pedágios, é possível atribuir semântica a essas trajetórias, estabelecendo, por exemplo, rotas turísticas ou trechos de congestionamento no fluxo do trânsito [Palma et al. 2008].

Uma forma de análise dos dados espaço-temporais, que ainda foi pouco explorada na literatura atual, é a classificação de trajetórias. Essa abordagem utiliza o conceito de classificação tradicional da mineração de dados [Tan et al. 2006], onde é aprendida uma função alvo  $f$  que mapeia um conjunto de atributos  $x$  para um dos rótulos de classe predefinidos  $y$ . Na classificação de trajetórias, a função  $f$ , ou modelo de classificação, mapeia cada trajetória para um determinado rótulo de classe. Esse rótulo de classe  $y$  determina o objetivo do processo de classificação, que pode ser desde a identificação de diferentes tipos de objetos móveis [Lee et al. 2008], até a identificação de meios de transporte (carro, bicicleta, etc.) utilizado pelo objeto móvel [Zheng et al. 2008].

Entre os métodos existentes para classificação de trajetórias, a técnica utilizada para a identificação das características para a classificação pode variar bastante. Pode ser utilizando, por exemplo, técnicas de análise de similaridade [Panagiotakis et al. 2009] ou de análise de regiões do espaço [Lee et al. 2008].

Este trabalho apresenta um novo método para classificação de trajetórias, denominado *TRACTS* (*Trajectory Classification using Time Series*), baseado na transformação de cada trajetória em um conjunto de séries temporais para que destas sejam extraídas as características que serão submetidas ao processo de geração do modelo de classificação. Este método é, no nosso conhecimento, o primeiro método que utiliza séries temporais para a classificação de trajetórias.

O restante do trabalho está organizado da seguinte forma: na Seção 2 são apresentados os trabalhos relacionados e uma breve base conceitual; na Seção 3 é apresentado o *TRACTS*, o novo método para classificação de trajetórias; a Seção 4 descreve alguns experimentos realizados utilizando-se o método proposto e os resultados obtidos; por fim, a Seção 5 conclui o artigo e apresenta trabalhos futuros.

## 2. Trabalhos relacionados

Para a classificação de dados convencionais, como no exemplo clássico da planta Iris, o conjunto de dados contém atributos descritivos da planta como o comprimento e largura da pétala e da sépala. Cada registro corresponde a um exemplo, com os dados de uma planta específica: os valores de suas características e a sua classe. Já para realizar a classificação de trajetórias, os dados fornecidos se resumem a uma sequência  $(id, x, y, t)$ , onde cada identificador da trajetória  $id$  está associado a uma classe, o que dificulta bastante o processo.

Entre a base de dados de trajetórias e a construção do modelo de classificação final, algum tipo de processamento dos dados da trajetória deve ser realizado. O trabalho proposto por [Panagiotakis et al. 2009] introduz um método de classificação de trajetórias que une técnicas de análise de similaridade de trechos das trajetórias com uma função de votação definida pela distância entre esses trechos, possibilitando a construção de um modelo de classificação a partir dessa função de votação.

A segmentação e extração de características de trajetórias também tem sido uma estratégia bastante utilizada nos métodos existentes para classificação de trajetórias. Os segmentos, e suas características, podem definir, por exemplo, comportamentos específicos e, quando ordenados em sequência, podem estabelecer a classificação de uma trajetória. Essa abordagem permitiu, no trabalho de [Lee and Hoff 2007], a identificação de atividades humanas a partir de trajetórias, adquiridas por vídeo, de diferentes modalidades esportivas.

Da mesma forma, no trabalho desenvolvido por [Zheng et al. 2008], foi possível inferir o meio de transporte utilizado a partir das trajetórias de diversos objetos móveis. Uma trajetória podia ter sido realizada com mais de um meio de transporte e, neste caso, o meio de transporte utilizado em cada sub-trajetória era definido. O método usa conhecimento da aplicação, por exemplo, para a passagem de um meio de transporte para outro a velocidade deveria ficar por alguns instantes próxima de zero e usa como principais características a serem consideradas para a classificação a velocidade média e a velocidade máxima na subtrajetória. Com isso diferenciava entre as 4 classes consideradas: carro, ônibus, bicicleta e a pé.

No trabalho de [García et al. 2006] são utilizados filtros de Kalman para identificar modos de voo (subtrajetórias em trajetórias de aviões como curvas, acelerações, aproximação para pouso, etc.).

Jiawei Han e seu grupo propõem o framework *TraClass* para a classificação de trajetórias, com dois métodos para a extração de características em [Lee et al. 2008]. O primeiro método, denominado *Region Based Clustering* (RB) descobre regiões com trajetórias essencialmente de uma classe, independente do tipo de movimento realizado. O segundo método, denominado *Trajectory Based Clustering* (TB) descobre subtrajetórias que descrevem padrões de movimento comum para cada classe. Nenhuma característica geométrica, global ou pontual, da trajetória é extraída para análise. Somente propriedades de movimentação (de subtrajetórias) e de regiões das trajetórias são utilizadas para a sua classificação.

Como pode ser observado, a maioria dos trabalhos descreve métodos específicos para a aplicação sendo considerada. O trabalho de [Lee et al. 2008] é que apresenta um

método mais geral. O método *TRACTS* também é geral. Ele utiliza algumas técnicas que serão sucintamente descritas a seguir, como base conceitual para o restante do trabalho.

## 2.1 Séries Temporais

Séries temporais consistem de sequência de valores ou eventos obtidos sobre repetidas medidas de tempo [Han and Kamber 2006]. Os valores são geralmente medidos em intervalos regulares de tempo (a cada hora, diariamente, semanalmente, etc.). Aplicações envolvendo séries temporais têm sido muito utilizadas em áreas como bolsa de valores, análise orçamentária, projeções de produtividade, controle de qualidade, observações de fenômenos naturais (temperatura, ventos, terremotos, etc.) e tratamentos médicos.

Devido ao grande tamanho e a alta dimensionalidade de muitas séries temporais, a redução de dados frequentemente serve como um primeiro passo na análise de séries temporais. A redução de dados leva não somente a um espaço de armazenamento muito menor, mas também a um processamento dos dados muito mais rápido.

Entre as estratégias para redução de dados, é possível encontrar a **redução dimensional**, que emprega técnicas de processamento de sinal para obter versões reduzidas do dado original. Essa estratégia de redução se mostra extremamente importante, pois, por exemplo, para computar a relação entre duas curvas de séries temporais, a redução da série temporal do comprimento (ou dimensão)  $n$  para  $k$  pode levar a redução de  $O(n)$  para  $O(k)$  em termos de complexidade computacional. Se  $k \ll n$ , a complexidade da computação será largamente reduzida.

A figura 1 mostra algumas técnicas de redução dimensional em séries temporais.

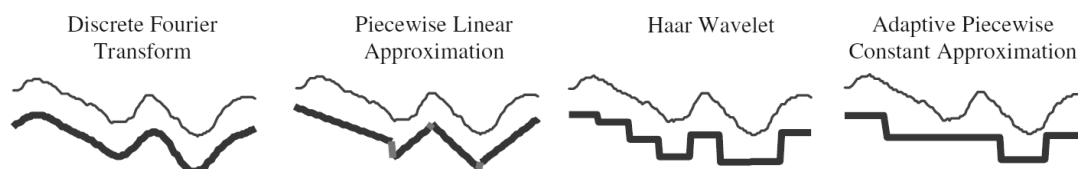


Figura 11: Redução dimensional de séries temporais

## 2.2 O método SAX

O método *Symbolic Aggregate Approximation* (SAX), proposto por [Lin et al. 2007], permite que uma série temporal de tamanho  $n$  seja reduzida para uma sequência de caracteres de tamanho  $w$ , onde  $w \leq n$ . O alfabeto utilizado para a sequência de caracteres é de tamanho  $a$ , onde  $a \geq 2$ . No método SAX, para cada um dos valores obtidos nas séries temporais do conjunto de dados, é realizada uma transformação que permite a sua representação simbólica.

A faixa de valores correspondentes a cada símbolo é determinada pelo valor dos elementos das diversas séries temporais consideradas e pela quantidade de faixas que são desejadas na representação. Nessa abordagem, todos os valores  $(v_1, v_2, v_3, \dots, v_n)$  de todas as séries temporais consideradas são primeiramente ordenados em ordem crescente de valor. Esse conjunto de valores é então dividido igualmente em  $f$  faixas  $(f_1, f_2, f_3, \dots, f_n)$ , sendo possível obter um número de  $\frac{n}{f}$  elementos em cada faixa. Dessa forma, é possível estabelecer que  $a = f$ . A aplicação do método SAX em duas séries temporais,  $C_1$  e  $C_2$ , é ilustrado na figura 2.

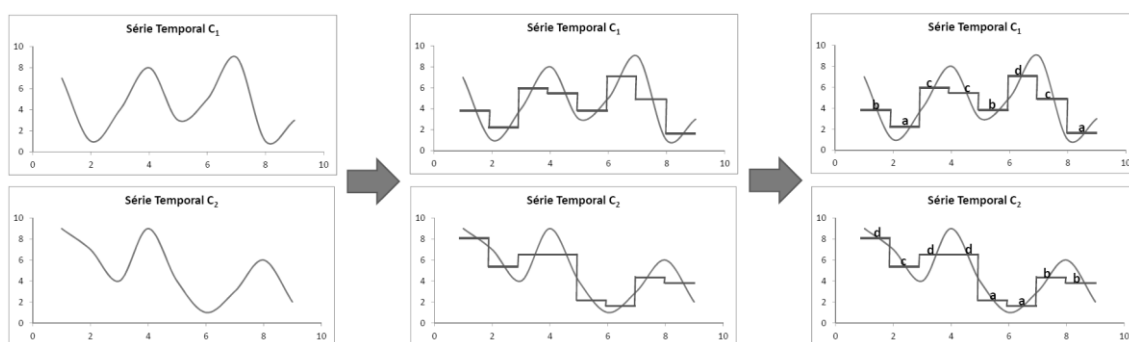


Figura 22: Aplicação do método SAX em duas séries temporais

### 2.3 O método Time-Series Bitmaps

Entre os métodos atualmente disponíveis para mineração de séries temporais que possuem capacidade de tratamento de sequências de caracteres, encontra-se o método que transforma séries temporais em mapas de bits, chamado *Time-Series Bitmaps* (TSB). Esse método foi proposto por [Kumar et al. 2005]. A ideia central do método é transformar uma sequência de caracteres, como, por exemplo, uma representação discretizada de uma série temporal, em uma matriz de mapa de bits, onde cada célula da matriz realizará a contagem das ocorrências de uma *substring* na *string* principal. Nesse método, a *substring* deve ter tamanho maior ou igual a um, definindo dessa forma a profundidade de análise desejada.

### 3. TRACTS: o método proposto para classificação de trajetórias

O método para a classificação de trajetórias *TRACTS* contém três etapas principais. A Figura 3 apresenta uma visão geral do método, que será detalhado nas seções seguintes.

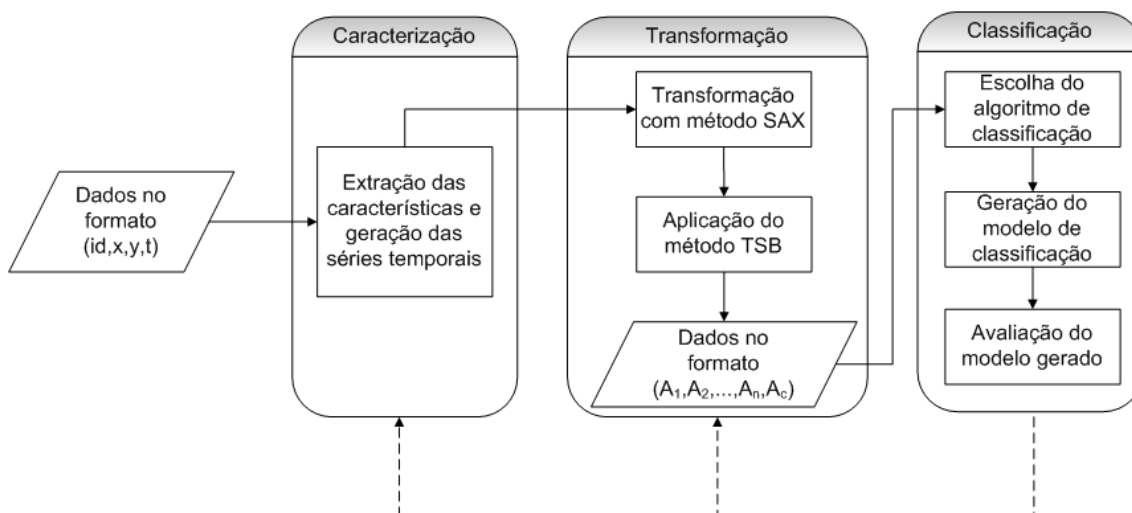


Figura 33: Visão geral do método TRACTS

Na figura não está representada a etapa inicial, presente em qualquer método de mineração de trajetórias, que é a parte de limpeza e eventualmente reconstrução das trajetórias, de acordo com a necessidade das operações que serão realizadas nas etapas seguintes. Um exemplo de caso onde ocorre esse tipo de situação é na segmentação de trajetórias em sub-trajetórias menores.

### 3.1 Caracterização

Na etapa de caracterização dos dados é realizada a identificação e extração das características importantes a partir do conjunto de dados. Essa é uma etapa fundamental para o sucesso do método *TRACTS*, pois ela fornece a semântica necessária para a análise no processo de classificação.

Uma tarefa não trivial nessa etapa é a identificação dos atributos que são relevantes para o domínio considerado. Entretanto, como estamos trabalhando com trajetórias de objetos em movimento, certas características serão sempre universais para todas as trajetórias, tais como velocidade, aceleração e direção vetorial absolutas, assim como a variação de cada uma dessas características em determinado intervalo. Essas são as características utilizadas por *default*. Cada trajetória é então representada como um conjunto de séries temporais, uma para cada característica considerada. Esse é um diferencial do método *TRACTS*, pois é o primeiro a considerar não apenas valores globais da trajetória como média e desvio padrão, mas o valor das características em cada ponto da trajetória.

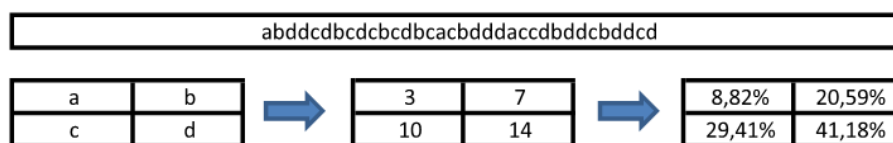
### 3.2 Transformação

Nesta etapa, os dados contínuos das séries temporais geradas na etapa anterior são processados de tal forma que um conjunto de séries temporais seja resumido a um único registro com as características relevantes dessas séries, em formato adequado para os algoritmos tradicionais de classificação.

A primeira técnica utilizada é o método de agregação simbólica SAX, introduzido na subseção 2.2. Na aplicação deste método, o conjunto das séries temporais de cada característica é considerado, de tal forma que cada letra do alfabeto utilizado tenha o mesmo número de ocorrências no conjunto das séries temporais daquela característica.

Com as séries temporais transformadas em sequências de caracteres, é necessário utilizar um método de tratamento de *strings* para possibilitar o processo de classificação dessas trajetórias. Para isso é utilizado o método TSB, explicado na subseção 2.3, que realiza a transformação das sequências de caracteres produzidas pelo método SAX em matrizes. É construída uma matriz TSB para cada característica de cada trajetória. Conforme explicado anteriormente, elas mapearão o número de *substrings* da *string* principal formada, em cada uma das células da matriz. Entretanto, como as trajetórias dos domínios podem ter quantidades diferentes de pontos, utiliza-se o percentual de cada uma delas. A Figura 4 apresenta um exemplo de geração da matriz TSB com profundidade um (considera como *substring* somente um símbolo por vez) para uma série temporal com quatro faixas (*a*, *b*, *c*, *d*). Quando a profundidade é maior, é obtida uma análise mais detalhada dos eventos sequenciais da série temporal, porém a complexidade computacional cresce exponencialmente para maiores níveis de profundidade.

Cada atributo de classificação para cada característica considerada é mapeado para uma das células da matriz TSB. Por exemplo, se na Figura 4 estivéssemos representando os valores de velocidade de uma trajetória, teríamos quatro atributos (com seus respectivos valores entre parênteses): *velocidade\_a* (8,82%), *velocidade\_b* (20,59%), *velocidade\_c* (29,41%) e *velocidade\_d* (41,18%).



**Figura 4: Geração de uma matriz TSB**

Portanto, após a aplicação do método TSB, cada célula da matriz, juntamente com seu valor, vira um atributo, possibilitando a construção do conjunto de dados no formato  $(A_1, A_2, A_3, \dots, A_n, A_c)$ , onde  $A_1, A_2, A_3, \dots, A_n$  correspondem aos atributos de característica e o atributo  $A_c$  ao atributo classe, que servirá como entrada para a próxima etapa, onde será realizada a classificação propriamente dita.

### 3.3 Classificação

Finalmente, o conjunto de atributos obtidos na etapa anterior é submetido a um algoritmo de classificação tradicional, tal como C4.5 ou Naive Bayes. O desempenho do algoritmo de classificação deve ser medido, permitindo estabelecer se o modelo de classificação gerado foi satisfatório ou não.

Assim como o processo tradicional de descoberta de conhecimento estabelece [Fayyad et al. 1996], caso seja necessário, no método *TRACTS* também é possível retornar a alguma das etapas anteriores, possibilitando o ajuste de parâmetros para geração de um novo modelo de classificação.

## 4. Experimentos Realizados

O método *TRACTS* foi testado e validado utilizando os mesmos dados do trabalho realizado pelo grupo de Jiawei Han [Lee et al. 2008], que introduziu o método *TraClass*, também para classificação de trajetórias. Isso permitiu uma base comparativa para os resultados obtidos.

Três bases de dados foram utilizadas nos experimentos. A primeira é o conjunto de trajetórias de três tipos de animais, a segunda é uma base de dados com trajetórias de navegação de dois barcos, e por fim, uma base de dados contendo trajetórias de rastreamento de furacões. Para cada base de dados, foi realizado um experimento em separado.

Para os três experimentos foi utilizado o mesmo conjunto de características: velocidade ( $v_n$ ), aceleração ( $a_n$ ), direção ( $d_n$ ) e variação da direção ( $vd_n$ ), calculadas da seguinte forma (onde  $\Delta D$  e  $\Delta t$  são a distância e o tempo entre dois pontos consecutivos da trajetória, respectivamente,  $\theta(p_{n-1}, p_n)$  retorna o ângulo entre a reta formada pelos pontos  $p_{n-1}$  e  $p_n$  e o norte geográfico e  $diff$  retorna a diferença entre dois ângulos):

$$v_n = \frac{\Delta D_{p_{n-1}}^{p_n}}{\Delta t_{p_{n-1}}^{p_n}} \quad a_n = \frac{v_{n+1} - v_n}{t_{n+1} - t_n} \quad d_n = \theta(p_{n-1}, p_n) \quad vd_n = diff(d_n, d_{n+1})$$

Com isso obtiveram-se quatro séries temporais para cada trajetória.

Na etapa de transformação, cada um dos valores das características obtidos na etapa anterior foi mapeado para um símbolo, através do método SAX (como as trajetórias eram relativamente pequenas, com uma média de 433 pontos em 2033 trajetórias, não houve necessidade de redução de dimensionalidade) e, logo após,

inserido em uma matriz de mapa de bits de série temporal. Para realizar a agregação simbólica com o método SAX, todos os  $n$  valores, para cada uma das quatro características obtidas de cada domínio, são considerados juntamente. Para cada característica, é formada uma lista ordenada do menor para o maior valor, que é dividida em um número equiprovável de elementos, de acordo com o número de faixas estabelecido. Conforme explicado na seção 3, essas faixas irão formar os símbolos SAX que serão mapeados para cada uma das trajetórias do domínio.

As séries temporais geradas pelo método SAX foram então submetidas ao método TSB para a geração das matrizes e, conseqüentemente, dos atributos utilizados para a classificação.

Com o arquivo de classificação contendo os atributos e seus respectivos valores, foi utilizada a ferramenta *Weka* [Frank et al. 2005], que possibilitou a utilização de diversos algoritmos de classificação, bem como a análise comparativa da acurácia de cada um dos modelos gerados. Em todas as execuções foi utilizada validação cruzada com 10 subconjuntos.

#### 4.1 Experimento com a base de dados de animais

A base de dados com a trajetória de animais está disponível em [Pacific Northwest Research Station 2005]. Ela contém a trajetória de três diferentes tipos de animais: alce, gado e veado, que determinam a classe das trajetórias. Todos foram rastreados a partir de coleiras RFID.

Nesse conjunto de dados, houve necessidade de eliminação de ruídos, já que foi observado que a velocidade média entre alguns pontos era maior que a velocidade possível para o respectivo animal que originava essa trajetória. Também foi possível observar certa irregularidade no intervalo de obtenção das coordenadas geográficas, o que eventualmente ocasionava grande tempo decorrido entre a aquisição de dois pontos consecutivos. Dessa forma uma segmentação das trajetórias em sub-trajetórias menores foi realizada.

A tabela 1 mostra os resultados comparativos, para a acurácia de classificação, entre o método TRACTS e o método *TraClass*.

**Tabela 11: Resultado comparativo para as trajetórias dos animais**

Método	Acurácia	Algoritmo de Classificação Utilizado
TRACTS.7.2	95,26%	<i>MultiLayerPerceptron</i>
TRACTS.3.4	93,28%	<i>SVM (SMO)</i>
TRACTS.5.3	93,28%	<i>SVM (SMO)</i>
TRACTS.3.5	93,28%	<i>SVM (SMO)</i>
TRACTS.3.3	91,70%	<i>Adaboost+NaiveBayes</i>
<i>TraClass RB-</i>	83,30%	<i>SVM</i>
<i>TraClass TB-</i>	50,00%	<i>SVM</i>

Cada uma das linhas da tabela onde o método *TRACTS* é citado representa um parâmetro de entrada para o método, onde o primeiro algarismo representa o tamanho do alfabeto SAX e o algarismo seguinte representa a profundidade de análise no método TSB. Foram executados experimentos para alfabetos SAX de tamanhos 3, 4, 5 e 7, com profundidade da matriz TSB variando de 1 a 5 e diversos algoritmos disponíveis na ferramenta *Weka*. Nas tabelas de resultados comparativos mostrados nesse trabalho, são exibidos os cinco melhores resultados obtidos com o método *TRACTS*. *TraClass RB-TB*



corresponde ao método *TraClass* aplicado com as duas opções de geração de características, *region-based* e *trajectory-based*. *TraClass TB-only* corresponde a execução do método *TraClass* apenas considerando a característica baseada na trajetória. Não implementamos o método *TraClass*. Nas tabelas com os resultados dos experimentos, no que se refere ao método *TraClass*, reproduzimos os valores constantes no artigo que descreve o método [Lee et al. 2008]. Como pode ser observado na Tabela 1, o método *TRACTS* obteve resultados sensivelmente melhores que o *TraClass* em várias configurações de parâmetros.

#### 4.2 Experimento com a base de dados de barcos

Para o segundo experimento foi utilizada a base disponível em [Monterey Bay Aquarium Research Institute 2001], contendo a trajetória de dois barcos. Como as trajetórias dos barcos eram apenas duas trajetórias extremamente longas, para que a análise fosse possível, cada uma das trajetórias foi dividida em 200 sub-trajetórias menores, com distribuição equivalente de pontos. A classe correspondia ao nome do barco.

Da mesma forma que na base de dados dos animais, as mesmas características foram extraídas, as strings SAX foram transformadas em matrizes de bitmaps e a ferramenta *Weka* foi utilizada para a etapa de classificação. O resultado desse experimento é apresentado na Tabela 2. Como pode ser observado, o resultado foi melhor que o do método *TraClass* em várias configurações de parâmetros.

**Tabela 22: Resultado comparativo para as trajetórias de barcos**

Método	Acurácia	Algoritmo de Classificação
<i>TRACTS.5.2</i>	100,00%	<i>Adaboost+J48</i>
<i>TRACTS.7.2</i>	100,00%	<i>Adaboost+SimpleCart</i>
<i>TRACTS.3.3</i>	99,75%	<i>NaiveBayes</i>
<i>TRACTS.3.4</i>	99,75%	<i>BayesNet</i>
<i>TRACTS.3.5</i>	99,75%	<i>BayesNet</i>
<i>TraClass RB-TB</i>	98,20%	<i>SVM</i>
<i>TraClass TB-Only</i>	84,40%	<i>SVM</i>

#### 4.3 Experimento com a base de dados de furacões

As trajetórias de furacões, disponível em [Unisys 2009], rastreou um conjunto de furacões que se originaram próximos ao golfo do México. Nessa base de dados, cada ponto continha a informação de força do furacão, que ia de F0 (mais fraco) até F5 (mais forte). Para que fosse possível classificar a trajetória como um todo, a trajetória inteira de cada furacão recebeu como atributo classe a força máxima que ele chegou. Nesse experimento, foram considerados somente os furacões de classe F2 e F3, como nos experimentos relatados no artigo que descreve o método *TraClass*, para possibilitar a comparação.

Como nos experimentos anteriores, foram executadas as etapas de caracterização, transformação e classificação, obtendo os resultados exibidos na Tabela 3.

Como pode ser observado pelos resultados obtidos, a acurácia foi bem menor do que nos experimentos anteriores, porque a trajetória dos furacões não está fortemente correlacionada com a sua intensidade. O método que obteve os melhores resultados foi o *TraClass RB-TB*, pois analisa a região de ocorrência das trajetórias (*Region Based*), já

que furacões mais fortes tendem a atingir locais mais afastados do que furacões mais fracos. No método *TRACTS*, os melhores resultados do método foram a partir de análises mais profundas do alfabeto SAX (com três ou quatro *substrings* sendo avaliadas por vez).

**Tabela 33: Resultado comparativo para as trajetórias de furacões**

Método	Acurácia	Algoritmo de Classificação
<i>TraClass RB-TB</i>	73,10%	SVM
<i>TRACTS.4.3</i>	69,45%	<i>Adaboost+J48</i>
<i>TRACTS.7.3</i>	69,45%	SVM (SMO)
<i>TRACTS.3.4</i>	68,41%	<i>Adaboost+RandomForest</i>
<i>TRACTS.5.3</i>	66,32%	<i>BayesNet</i>
<i>TraClass TB-Only</i>	65,40%	SVM
<i>TRACTS.3.2</i>	64,23%	<i>BayesNet</i>

#### 4.5 Resultado geral

Considerando os melhores resultados de cada uma das execuções dos experimentos, é possível compilar um resultado geral para os experimentos, através da média de acurácia dos resultados, conforme pode ser visualizado na Tabela 4.

**Tabela 44: Resultado comparativo da aplicação do método *TRACTS* com o método *TraClass***

Método	Acurácia			
	Animais	Barcos	Furacões	Média
<i>TRACTS</i>	95,26%	100,00%	69,45%	<b>88,24%</b>
<i>TraClass RB-TB</i>	83,30%	98,20%	73,10%	<b>84,87%</b>
<i>TraClass TB-Only</i>	50,00%	84,40%	65,40%	<b>66,60%</b>

Mesmo com a dificuldade de classificação no domínio de dados dos furacões, o método *TRACTS* apresentou resultado superior quando é considerada a média da acurácia dos experimentos realizados.

## 5. Conclusão e trabalhos futuros

Nesse trabalho foi apresentado o *TRACTS*, um novo método para classificação de trajetórias, utilizando séries temporais e técnicas de mineração de dados para esse tipo de representação, possibilitando a criação de modelos de classificação com boa acurácia.

O método *TRACTS* demonstrou-se eficaz na tarefa de criar um modelo de classificação com boa acurácia, no geral superior a um método proposto para o mesmo fim.

Fica possível observar que o método, apesar de possuir uma grande independência de domínio, possui certa dificuldade para realizar a classificação de domínios caóticos, tal como o domínio dos furacões. Em parte, isso se deve as características extraídas de cada uma das trajetórias desse domínio, já que furacões não se movem mais rápida ou vagorosamente de acordo com algum critério conhecido. Outra possível causa da dificuldade de classificação nesse domínio foi estabelecer o atributo classe como a força máxima atingida pelo furacão, já que a força máxima pode ter sido atingida somente em um breve período de tempo, enquanto o método considera a trajetória inteira. Seria possível segmentar a trajetória dos furacões em subtrajetórias para cada classe, o que melhoraria de forma significativa a acurácia, mas como o

método usado na comparação não tratou o conjunto de dados dessa forma, essa formatação não foi considerada.

Como trabalho futuro será realizando um estudo mais aprofundado, considerando outros domínios de aplicação, sobre a influência do tamanho do alfabeto utilizado no método SAX e do tamanho da *substring* avaliada no método TSB na acurácia de classificação. Também será investigado o uso de outros métodos de tratamento de *strings*, além do TSB.

## 6. Referências

- Ashbrook, D., & Starner, T. (2003). "Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users". *Personal and Ubiquitous Computing Journal*, 275 - 286.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases". *AI Magazine*, 37-54.
- Frank, E., Hall, M., Holmes, G., Kirkby, R., & Pfahringer, B. (2005). "WEKA - A Machine Learning Workbench for Data Mining". In: *Collection of The Data Mining and Knowledge Discovery Handbook* (pp. 1305-1314). Springer.
- García, J., Concha, O. P., Molina, J. M., & Miguel, G. (2006). "Trajectory classification based on machine-learning techniques over tracking data". *Proceedings of 9th International Conference on Information Fusion*, (pp. 1-8). Florence, Italia.
- Gudmundsson, J., Kreveld, M. v., & Speckmann, B. (2007). "Efficient Detection of Patterns in 2D Trajectories". *GeoInformatica*, 195-215.
- Han, J., & Kamber, M. (2006). "Mining Stream, Time-Series and Sequence Data". In: *Data Mining: Concepts and Techniques*. (pp. 467-489). San Francisco, CA, EUA: Morgan Kaufmann.
- Hariharan, R., & Toyama, K. (2004). "Project Lachesis: Parsing and Modeling Location Histories". *Proceedings of the 3rd International Conference on Geographic Information Science* (pp. 106-124). Adelphi, EUA: Springer.
- Kumar, N., Lolla, V. N., Keogh, E., Lonardi, S., Ratanamahatana, C. A., & Wei, L. (2005). "Time-series Bitmaps: A Practical Visualization Tool for working with Large Time Series Databases". *Proceedings of 5th SIAM International Conference on Data Mining - SDM'05, Proceedings* (pp. 531-535). Newport Beach, CA, EUA: SIAM.
- Lee, J. Y., & Hoff, W. (2007). "Activity Identification Utilizing Data Mining Techniques". *Motion and Video Computing, IEEE Workshop on* (p. 12). Austin, Texas, EUA: IEEE Computer Society.
- Lee, J.-G., Han, J., Gonzalez, H., & Li, X. (2008). "TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering". *VLDB Endowment* (pp. 1081-1094). Auckland, Nova Zelândia: VLDB Endowment.
- Liao, L., Patterson, D., Fox, D., & Kautz, H. (2006). "Building Personal Maps from GPS Data". *Annals of the New York Academy of Sciences*, 249 - 265.
- Lin, J., Keogh, E., Wei, L., & Lonardi, S. (1 de Outubro de 2007). "Experiencing SAX: a novel symbolic representation of time series". *Data Mining and Knowledge Discovery*, pp. 107-144.

- Monterey Bay Aquarium Research Institute. (01 de Setembro de 2001). Acesso em 16 de Julho de 2010, disponível em MUSE Project: <http://www.mbari.org/MUSE/platforms/ships.htm>
- Pacific Northwest Research Station. (2005). *US Forest Service*. Acesso em 16 de Julho de 2010, disponível em The Starkey Project: <http://www.fs.fed.us/pnw/starkey/data/tables/>
- Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). “A Clustering-based Approach for Discovering Interesting Places in Trajectories”. *Proceedings of 23rd Annual Symposium on Applied Computing*, (pp. 863-868). Fortaleza, Ceara, Brasil.
- Panagiotakis, C., Pelekis, N., & Kopanakis, I. (2009). “Trajectory Voting and Classification based on Spatiotemporal Similarity in Moving Object Databases”. *Proceedings of the 8th International Symposium on Intelligent Data Analysis: Advances in Intelligent Data Analysis VIII* (pp. 131-142). Lyon, France: Springer.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Addison-Wesley.
- Tiakas, E., Papadopoulos, A. N., Nanopoulos, A., Manolopoulos, Y. P., Stojanovic, D., & Kajan, S. D. (2009). “Searching for similar trajectories in spatial networks”. *Journal of Systems and Software*, 772-788.
- Tiesyte, D., & Jensen, C. S. (2008). “Similarity-based prediction of travel times for vehicles traveling on known routes”. *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems* (pp. 105-114). Irvine, CA, EUA: ACM.
- Unisys. (2009). *Atlantic Tropical Storm Tracking by Year*. Acesso em 16 de Julho de 2010, disponível em Unisys Weather: <http://weather.unisys.com/hurricane/atlantic/>
- Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). “Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web”. *Proceeding of the 17th international conference on World Wide Web* (pp. 247-256). Pequim, China: ACM.
- Zhou, C., Bhatnagar, N., Shekhar, S., & Terveen, L. (2007). “Mining Personally Important Places from GPS Tracks”. *IEEE 23rd International Conference on Data Engineering Workshop* (pp. 517-526). Istambul, Turquia: IEEE Computer Society.