Classificação de Séries Temporais através da Probabilidade de Auto-Transição obtida do Grafo de Transição de Padrões Ordinais

Isadora Cardoso-Pereira[1], João B. Borges[3], Pedro H. Barros[1], Antonio F. Loureiro[1], Heitor S. Ramos[1]

[1]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

[2]Departamento de Computação e Tecnologia
Universidade Federal do Rio Grande do Norte

{isadoracardoso, loureiro, ramosh}@dcc.ufmg.br, joaoborges@dct.ufrn.br,

phbarros@ufmg.br

Abstract. *Regarding its interdisciplinary and broad scope of real-world applications, it is evident the need of extracting knowledge from time series data. Mining this type of data, however, faces several complexities due to its unique properties. In this work, we propose a new feature, retained from the Ordinal Pattern Transition Graph, called probability of self-transition. Our proposal was tested in a real Urban Computing problem, referred to the classification of transportation mode. We had better accuracy results than well-known Information Theory features, besides being less dependent on Ordinal Patterns parameters.*

Resumo. *Baseando-se em sua interdisciplinaridade e grande escopo em aplicações do mundo real, é clara a necessidade de extrair conhecimento de séries temporais. Porém, minerar dados de séries temporais é uma atividade complexa, devido às suas propriedades particulares. Neste trabalho, propomos a utilização de um novo atributo retirado do Grafo de Transição de Padrões Ordinais, chamado de probabilidade de auto-transição. Nossa proposta foi testada em um problema real de Computação Urbana, referente à classificação de modos de transporte. Obtivemos melhores valores de acurácia do que atributos bem conhecidos de Teoria de Informação , além de apresentar menor dependência em relação aos parâmetros de Padrões Ordinais.*

1. Introdução

A análise de trajetórias de GPS é um problema bastante estudado na área de Computação Urbana, sendo utilizada para rastrear pessoas [Zheng et al. 2008] ou veículos [Aquino et al. 2015]. Em especial, examinar pessoas e identificar o modo de transporte o qual utilizam é essencial para cidades que desejam reduzir o congestionamento e o tempo de deslocamento entre seus pontos, melhorando a vida do cidadão. Os dados da trajetória do objeto de estudo são representados como uma coleção discreta de pontos através do tempo – ou seja, uma série temporal. Há

uma ampla gama de áreas que estudam seus fenômenos usando observações temporais [Lines et al. 2018].

Baseando-se em sua interdisciplinaridade e grande escopo em aplicações no mundo real, é clara a necessidade de extrair conhecimento de séries temporais. Por consequência, estas tem sido objeto de estudo há décadas [Längkvist et al. 2014]. Porém, minerar dados de séries temporais é uma atividade complexa, sendo considerada por Yang e Wu [Yang and Wu 2006] um dos problemas mais desafiadores no contexto de mineração de dados, devido às suas propriedades particulares. Além da alta dimensionalidade e ruído, problemas bem conhecidos na era do *Big Data*, séries temporais possuem dados dependentes da ordenação. Logo, alterar a ordem desses dados pode alterar o significado do fenômeno observado. Isso contraria uma suposição comum dos algoritmos de aprendizado de máquina, como *Naive Bayes*, de dados independentes e identicamente distribuídos, resultando em algoritmos com baixo desempenho [Bagnall et al. 2017]. Uma diferente representação dos dados pode superar esses problemas.

Representação de dados, ou pré-processamento de dados, é um passo essencial na mineração de dados de séries temporais. Consiste em aplicar transformações diretamente na série temporal no mesmo domínio (por exemplo, resumindo os pontos dos dados originais em um formato mais compreensível [Wilson 2017]) ou em transformar os dados para diferentes domínios (por exemplo, frequência, *shapelets*, símbolos etc [Bagnall et al. 2017]). Uma representação adequada não apenas diminui a dimensionalidade e remove ruídos, mas também preserva os atributos globais e locais mais críticos dos dados originais [Wilson 2017]. Assim, atributos úteis, isto é, atributos que representam os dados originais, podem ser utilizados para diversas análises de série temporal, como classificação, possibilitando uma computação eficiente. Além disso, esses atributos devem ser robustos aos possíveis problemas dos dados, como falta de dados, *outliers*, espaço de tempo irregular, entre outros.

No presente trabalho, empregamos Padrões Ordinais (PO) [Bandt and Pompe 2002], em conjunto com sua transformação em grafos, conhecida como Grafo de Transição de Padrões Ordinais (GTPO) [Small 2013], para representar os dados de uma série temporal em um novo domínio e então classificá-la utilizando atributos retirados dessa transformação. Em especial, propomos a utilização de um novo atributo, chamado probabilidade de auto-transição. Com essas ferramentas, buscamos caracterizar séries temporais de acordo com seu comportamento. A validação de nossa proposta é feita em um problema real de Computação Urbana, referente à classificação de modos de transporte utilizado pelos usuários. A contribuição deste trabalho está baseada na colaboração desse novo atributo para uma apropriada caracterização e classificação de séries temporais.

As demais seções deste trabalho estão organizadas da seguinte maneira: Seção 2 discute brevemente os trabalhos relacionados; Seção 3 apresenta a metodologia utilizada no trabalho; Seção 4 discute os resultados obtidos e a Seção 5 encerra com as devidas conclusões.

2. Trabalhos Relacionados

A caracterização e classificação de séries temporais é tema de estudo de diversas áreas e, sendo assim, é largamente explorada. Há diversas contribuições no campo da Aprendizagem de Máquina (ML, da nomeclatura em inglês), como pode ser visto em [Wang et al. 2010, Bagnall et al. 2017, Lines et al. 2018]. Esses estudos comparam a efetividade de mais de 20 técnicas de ML na classificação de séries temporais dos mais diversos domínios, incluindo Computação Urbana (por exemplo, contagem de pedestres e carros para entender a utilização de espaços públicos, predição de eventos como terremotos a partir de sensores espalhados pela cidade etc). Dentre as técnicas avaliadas, destaca-se Hierarchical Vote Collective of Transformation-based Ensembles (HIVE-COTE). Proposto por Lines et al. [Lines et al. 2018], HIVE-COTE é uma técnica *ensemble* composta de 35 classificadores, modularizados de acordo com o domínio no qual atuam. Apesar de apresentar boa acurácia na classificação em diversos domínios, é uma técnica de alto custo computacional, mesmo em comparação à Aprendizagem Profunda (como constatado em [Fawaz et al. 2019]), o que a torna inviável em séries temporais de alta dimensionalidade.

Técnicas advindas da Teoria da Informação também tem tido sucesso na caracterização de séries temporais na área de Computação Urbana. Tais métodos são capazes de distinguir séries temporais utilizando técnicas livres de modelo, computacionalmente baratas e de baixa dimensionalidade, como PO e Plano de Complexidade-Entropia [Rosso et al. 2007]. Por exemplo, Aquino et al. [Aquino et al. 2015] caracterizou o comportamento de veículos através de suas velocidades; Aquino et al. [Aquino et al. 2017] caracterizou o comportamento de cargas elétricas e Ribeiro et al. [Ribeiro et al. 2017] caracterizou o comportamento do preço de petróleo bruto.

Mais uma direção de pesquisa que também obteve sucesso na caracterização de séries temporais é baseada na transformação da série temporal em grafos. A partir destes, constrói-se redes que herdam as características da série temporal original (por exemplo, séries periódicas são transformadas em grafos regulares e séries aleatórias são transformadas em grafos aleatórios). Alguns exemplos são o grafo de visibilidade [Lacasa et al. 2008] e o grafo horizontal de visibilidade [Luque et al. 2009]. Porém, como cada amostra da série temporal é transformada em um vértice do grafo, há um impacto na escalabilidade dessas técnicas, as tornando inviáveis para séries temporais de alta dimensionalidade.

Recentemente, métodos que combinam mais de uma abordagem estão surgindo. Em [Small 2013], [Ravetti et al. 2014], [Zhang et al. 2017] e [Guo et al. 2018], vemos técnicas que obtêm grafos a partir das permutações dos padrões possíveis em PO, aproveitando as vantagens das duas abordagens.

Os estudos apresentados acima mostram que a classificação de séries temporais é factível. Este trabalho é inspirado pela utilização de técnicas de diferentes áreas, bem como pela sua combinação. Aqui, obtemos um novo atributo retirado do GTPO, baseado na probabilidade de auto-transição, e avaliamos seu impacto na classificação de séries temporais no contexto de Computação Urbana.

Tabela 1. Distância e duração dos modos de transporte

transporte	distância (km)	duração (h)
caminhada	10 123	5460
bicicleta	6495	2410
carro/táxi	32 866	2384
ônibus	20 281	1507

Tabela 2. Trajetórias obtidas da base de dados

transporte	trajetórias
caminhada	1653
bicicleta	840
ônibus	1017
carro/táxi	831
total	4341

3. Metodologia

3.1. Base de dados

Para validar nossa proposta, utilizamos a base de dados GeoLife [1], coletada por Zheng et al. [Zheng et al. 2008]. Essa base de dados apresenta a trajetória de GPS de 182 usuários em um período de cinco anos (de abril de 2007 a agosto de 2012), contendo informações da latitude, longitude e altitude. Dentre esses usuários, 73 possuem informações do modo de transporte, o que será classificado neste trabalho. Apenas os modos de transporte com uma duração superior a 1000 horas foram considerados, pois entendemos que, quanto menor a série temporal, mais difícil extrair informações relevantes, o que leva a geração de modelos de baixa qualidade. Os modos de transporte utilizados estão descritos na Tabela 1.

Neste trabalho, definimos trajetória como uma sequência ininterrupta de pontos de GPS (latitude e longitude) que pertencem ao mesmo modo de transporte. Para extrair as trajetórias, agrupamos os pontos do conjunto de dados em usuário, dia e modo de transporte. Trajetórias com menos de 10 pontos foram descartadas a fim de evitar a criação de trajetórias de baixa qualidade, o que pode afetar o modelo gerado. A Tabela 2 mostra o total de trajetórias obtidas para cada transporte.

Até o momento, nossa proposta é designada para séries temporais unidimensionais. Dessa forma, assumimos o modelo caricato no qual a latitude e longitude fornecem informações independentemente. Além disso, extraiu-se atributos que combinam essas duas medidas, como a distância euclidiana de um ponto a outro.

3.2. Transformação em Padrões Ordinais

Seja uma série temporal $\mathbf{X}(t) = \{x_1, x_2, ..., x_n\}$ de tamanho n e seja também uma dimensão $D \in \mathbb{N}$ e um atraso $\tau \in \mathbb{N}$. A cada instante $t = \{1, ..., n - (D-1)\tau\}$, gera-se uma janela deslizante $w_t \subseteq x$, tal que

[1] `https://www.microsoft.com/en-us/download/details.aspx?id=52367`

$$w_t = \{x_t, x_{t+\tau}, ..., x_{t+(D-2)\tau}, x_{t+(D-1)\tau}\},$$

isto é, cada elemento dentro da janela deslizante é obtido da série temporal no tempos $t, ..., t + (D-1)\tau$. Isso corresponde a uma amostra da série temporal em um intervalo regularmente espaçado.

A relação ordinal para cada instante t consiste da permutação $\pi = \{r_0, r_1, ..., r_{D-1}\}$ de $\{0, 1, ..., D-1\}$, de forma que

$$x_{t-r_{D-1}} \leq x_{t-r_{D-2}} \leq \cdots \leq x_{t-r_1} \leq x_{t-r_0}.$$

Em outras palavras, π representa a permutação dos elementos da janela w_t, organizados em ordem crescente. A fim de se obter resultados únicos, define-se que, se a série temporal tiver elementos de tal forma que $x_{t-r_i} = x_{t-r_{i-1}}$, considera-se que $r_i < r_{i-1}$. Assim, a série temporal é convertida em um conjunto de padrões ordinais, $\Pi = \{\pi_1, ..., \pi_m\}$, onde $m = n - (D-1)\tau$, e cada π_m representa uma permutação do conjunto das possíveis permutações de $D!$ [Aquino et al. 2017].

A escolha do valor de D depende do tamanho da série temporal e deve satisfazer a condição $n \gg D!$ – quanto mais se aumenta D, mais deve-se aumentar a série temporal para que as estatísticas extraídas dos dados sejam confiáveis [Rosso et al. 2007]. Explicações mais aprofundadas são fornecidas em [Staniek and Lehnertz 2007]. Para propósitos práticos, Bandt e Pompe [Bandt and Pompe 2002] recomendam valores tais que $3 \leq D \leq 7$, os quais são adotados neste trabalho.

Para todas as $D!$ possíveis permutações π de D, a frequência relativa pode ser computada pelas vezes que certa sequência apareceu na série temporal, dividida pelo total de número total de sequências, obtendo-se o histograma da distribuição de probabilidade $P \equiv \{p(\pi)\}$, que é definido por:

$$p(\pi) = \frac{\mid s_\pi \mid}{n - (D-1)\tau},$$

onde $\mid s_\pi \mid \in \{0, ..., m\}$ é o número de padrões observados do tipo π.

Dessa nova representação é possível extrair atributos, tais como quantificadores de Teoria da Informação, que podem ser utilizados para caracterizar as dinâmicas da série temporal [Rosso et al. 2007]. Neste trabalho, extraímos dois quantificadores, a entropia de permutação e complexidade estatística.

3.2.1. Entropia de Permutação

A Entropia de Permutação é uma medida de incerteza associada aos processos descritos por p_π e é definida por:

$$H[p_\pi] = -\sum p(\pi) \log_2 p(\pi),$$

onde $0 \leq H[p_\pi] \leq logD!$. Essa medida é equivalente a entropia de Shannon [Aquino et al. 2017]. Baixos valores de $H[p_\pi]$ representam uma sequência de aumento ou diminuição de valores na distribuição de permutação, indicando que a série temporal original é determinística, enquanto altos valores indicam um sistema completamente aleatório [Bandt and Pompe 2002].

O máximo valor para $H[p_\pi]$ ocorre quando todas as possíveis permutações de $D!$ tem a mesma probabilidade de ocorrer, o que é o caso para a distribuição uniforme p_u das permutações. Assim, $H_{max} = H[p_u] = \log D!$ [Zunino et al. 2012]. Podemos definir a Entropia normalizada de Shannon, para o caso da entropia de permutação, como:

$$H_S[p_\pi] = \frac{H[p_\pi]}{H_{max}}, \tag{1}$$

onde $0 \leq H_S[p_\pi] \leq 1$ [Rosso et al. 2007].

3.2.2. Complexidade Estatística

A Complexidade Estatística é baseada na divergência de Jensen-Shannon (JS) entre a distribuição de probabilidade associada p_π e a distribuição uniforme p_u (o caso trivial para o conhecimento mínimo do processo) e é definida por:

$$C_{JS}[p_\pi] = Q_{JS}[p_\pi, p_u] H_S[p_\pi],$$

onde $p_\pi = \{p(\pi)\}$ é a distribuição de probabilidade de padrões ordinais, p_u é a distribuição uniforme e H_S é a entropia normalizada de Shannon, como definido na Equação 1. O desequilíbrio $Q_{JS}[p_\pi, p_u]$ é dado por:

$$Q_{JS}[p_\pi, p_u] = Q_0 JS[p_\pi, p_u] = Q_0 \left(S\left[\frac{p_\pi + p_u}{2}\right] - \frac{S[p_\pi] + S[p_u]}{2} \right),$$

onde S é a entropia de Shannon e Q_0 é definida por:

$$Q_0 = -2 \left[\left(\frac{D! + 1}{D!} \right) ln(D! + 1) - 2ln(2D!) + ln(D!) \right]^{-1},$$

o que descreve a constante de normalização, a qual é igual ao inverso do máximo valor de $JS[p_\pi, p_u]$ e $0 \leq Q_{JS} \leq 1$ [Aquino et al. 2017, Rosso et al. 2007].

3.3. Grafo de Transição de Padrões Ordinais

Dada a sequência de padrões ordinais Π, o grafo de transição de padrões ordinais representa a relação entre os consecutivos padrões ordinais e é definido como um grafo dirigido ponderado $G_\pi = (V, A)$, com vértices $v_{\pi_i} \in V = \{v_{\pi_i} : i = 1, ..., D!\}$,

que corresponde a uma das possíveis permutações de $D!$ para a dimensão D, e arestas $A = \{(v_{\pi_i}, v_{\pi_j}) : v_{\pi_i}, v_{\pi_j} \in V\}$.

Uma aresta dirigida conecta dois PO no grafo se esses aparecem sequencialmente na série temporal original, representando a transição entre os padrões. Os pesos $w : E \to \mathbb{R}$ das arestas representam a probabilidade da existência de uma específica transição em Π e é dado por:

$$w(v_{\pi_i}, v_{\pi_j}) = \frac{\mid \Pi_{\pi_i,\pi_j} \mid}{m-1},$$

onde $\mid \Pi_{\pi_i,\pi_j} \mid \in \{0, ..., m-1\}$ é o número de transições entre as permutações π_i e π_j e $\sum_{v_{\pi_i},v_{\pi_j}} w(v_{\pi_i}, v_{\pi_j}) = 1$.

Uma vez que o grafo é construído a partir do conjunto de PO, algumas das propriedades dessa transformação são herdadas. As mais notáveis são:

- **simplicidade e agilidade**: a construção do grafo depende apenas do número m de padrões ordinais, precisando somente contabilizar o número de transições em $m-1$ passos. Por sua vez, a transformação da série temporal em PO depende do tamanho n da série e da dimensão D. A complexidade dessa transformação é limitada por $O(nD^2)$, assumindo que as permutações são obtidas ordenando as janelas deslizantes por um algoritmo simples de ordenação, como *Selection Sort*, em $O(D^2)$ e $\tau = 1$, no pior caso. Por razões práticas, como D é recomendado estar no intervalo entre 3 e 7, a ordenação dessa estratégia tem no máximo 7 elementos, então a complexidade dessa estratégia tem maior dependência do tamanho n da série temporal;
- **escalabilidade**: As abordagens que utilizam grafo de visibilidade [Lacasa et al. 2008], por exemplo, transformam cada amostra da série temporal em um vértice no grafo – uma abordagem inviável para séries temporais de alta dimensionalidade devido ao espaço necessário para armazenamento. Por outro lado, o número de vértices do grafo de transição é dado pela dimensão D, não dependendo do tamanho da série e sendo limitado por $D!$;
- **robustez**: PO são robustos a presença de ruídos e invariantes com respeito a transformações não-lineares monótonas [Aquino et al. 2017, Rosso et al. 2007].

Na Figura 1 vemos uma ilustração do processo descrito acima: (a) Tem-se a série temporal original; (b) calcula-se janelas deslizantes, com valores de D e τ ($D = 3$ e $\tau = 2$, na figura); (c) retira-se os padrões ordinais, repetindo o processo para toda a série temporal; e (d) constrói-se o GTPO com um nó para cada PO encontrado na série temporal e com arestas que descrevem a sucessão temporal dos padrões [McCullough et al. 2017].

3.4. Probabilidade de auto-transição

As auto-transições do grafo de transição são as arestas a partir de um vértice direcionadas ao próprio vértice, também conhecidas como laço ou *loop*. Sua presença em um grafo representa a ocorrência de padrões ordinais consecutivos.

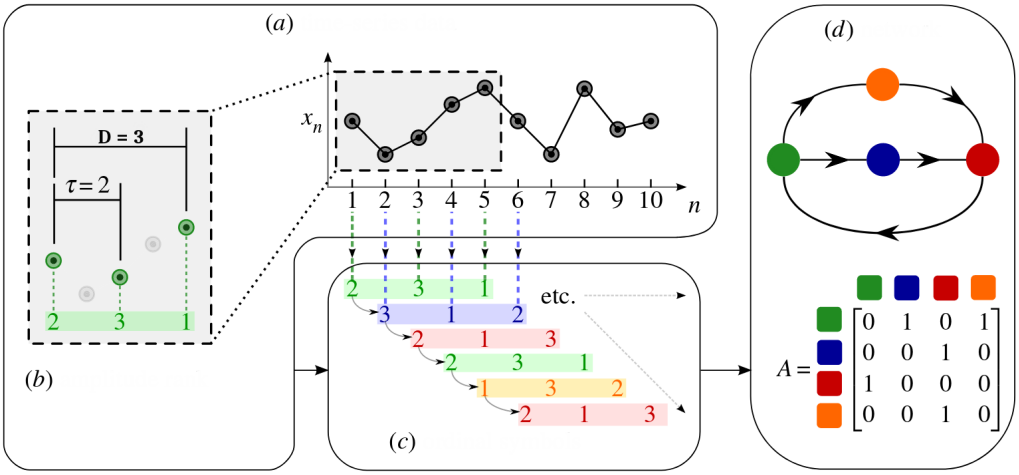

Figura 1. Ilustração da transformação das séries temporais para GTPO (Adaptada de [McCullough et al. 2017])

Zhan et al. [Zhang et al. 2017] propuseram a análise da entropia computada através dos pesos das arestas do grafo de transição após a remoção das arestas de auto-transição. Porém, essas transições estão diretamente relacionadas a correlação temporal da série temporal original e são uma valiosa indicação das dinâmicas ocultas e, sendo assim, não deveriam ser descartadas. A forma como essas arestas estão colocadas é um elemento importante para a posterior análise do grafo.

A probabilidade de auto-transição é definida como a probabilidade de ocorrência de uma sequência de padrões iguais dentro do conjunto de PO e pode ser expressada como:

$$p_{st} = p(\pi_i, \pi_i) = \sum_{i \in 1,\ldots,D!} w(v_{\pi_i}, v_{\pi_i})$$

A normalização dos pesos do grafo adotada neste trabalho é semelhante a adotada por Zhan et al. [Zhang et al. 2017], onde os autores normalizam os pesos do grafo de forma que todos somem 1. Porém, em nosso caso, aceitamos a presença de auto-transição.

4. Resultados e Discussões

Testamos a qualidade da nossa proposta através da classificação de modos de transporte, utilizando a base de dados descrita anteriormente na seção 3.1. Nesse caso, assumimos que uma boa qualidade se refere a bons resultados de acurácia, precisão e sensibilidade na classificação – apresentando bons resultados nessas métricas, pode-se inferir que a representação dos dados utilizada neste trabalho são satisfatórias, isto é, o atributo proposto retém informações da série temporal original, tornando capaz a sua classificação.

O processo utilizado para a extração de atributos e posterior classificação está mostrado na Figura 2. Dependendo da natureza dos dados, é possível utilizá-los sem nenhuma extração de informação prévia, assim como é possível extrair

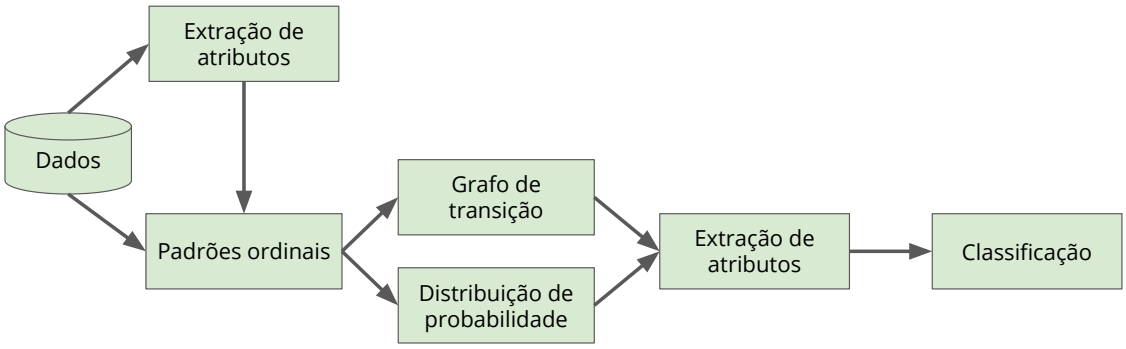

Figura 2. Metodologia utilizada neste trabalho

atributos que possam expressar melhor o conhecimento oculto dos dados, os destacando na transformação. Nesse trabalho utilizamos as duas abordagens para comparação: trabalhamos com a latitude e longitude proporcionada pelos dados, e também transformamos esses dois atributos em um terceiro, a distância euclidiana entre dois pontos consecutivos. Após, transformamos esses três atributos em PO, de onde extraímos o GTPO e a distribuição de probabilidade dos PO. Do GTPO extraímos a probabilidade de auto-transição e da distribuição de probabilidade extraímos a Entropia de Permutação e Complexidade Estatística. Com esses atributos, realizamos a classificação.

Como a ideia deste trabalho é ressaltar as particularidades da transformação dos dados, pouco esforço foi dedicado ao ajuste dos algoritmos de classificação. É possível obter melhores resultados das métricas de avaliação ajustando os parâmetros dos classificadores, porém, nosso objetivo não é apenas apresentar bons resultados de tais métricas, mas sim saber se nossa proposta é adequada para caracterização e classificação de séries temporais. Nossa classificação foi feita utilizando algoritmos simples, que são: k-Vizinhos Mais Próximos (k-NN), com $k = 2$; Máquina de Vetores de Suporte, com *kernel* linear (SVM-L) e radial (SVM-R) e Árvore de Decisão (DT).

Para avaliar a capacidade de generalização de nossa proposta e validar os resultados, utilizamos validação cruzada, com 5-*folds*. É importante destacar que essa validação cruzada é realizada nos atributos extraídos, imediatamente antes da classificação. Esses atributos, após passarem pelas transformações descritas neste trabalho, podem ser interpretados como independente e identicamente distribuídos, permitindo a utilização dessa técnica de validação.

A implementação dos métodos foi realizada em R (versão 3.4.4), numa máquina com a seguinte configuração: SO Linux, 8 GB de RAM e Intel® Core™ i5-2410M CPU @ 2.30GHz.

Primeiramente, vamos avaliar a influência da dimensão D na classificação. Na Figura 3, tem-se a acurácia obtida ao classificar latitude, longitude e distância, utilizando diferentes valores de D (entre 3 e 7, inclusive, como recomendado em [Bandt and Pompe 2002]). Ao classificar cada atributo isoladamente, vemos que p_{st} atinge o melhor valor de acurácia, com cerca de 70 %, enquanto $H[p_\pi]$ e $C_{JS}[p_\pi]$ alcançam cerca de 65 % e 63 %, respectivamente. Ou seja, utilizando apenas um único atributo, p_{st}, tem-se um ganho de cerca de 5 % de acurácia. Nessa mesma

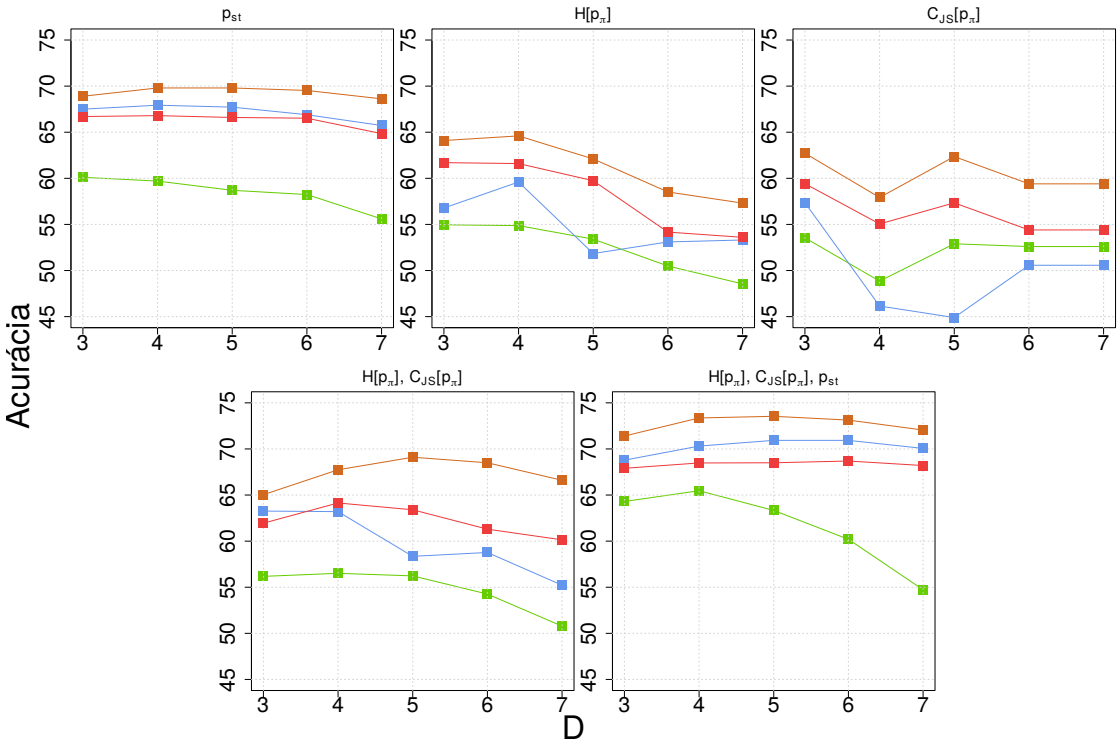

Figura 3. Acurácia das classificações com SVM-R (■), SVM-L (■), DT (■) e k-NN (■), utilizando diferentes valores de D.

figura vemos a classificação utilizando conjuntos de atributos: **(1)** $\{H[p_\pi], C_JS[p_\pi]\}$ e **(2)** $\{H[p_\pi], C_JS[p_\pi], p_{st}\}$. Para o conjunto (1), p_{st} ainda apresenta resultados sensivelmente superiores, o que sugere que o ganho de informação para p_{st} é superior ao ganho dos outros dois atributos em conjunto. No conjunto (2) vemos que a classificação apresenta os melhores resultados dentre os apresentados, ou seja, a classificação utilizando p_{st} pode ser melhorada se combinada com outros atributos. Além disso, vemos que o classificador que apresenta os melhores resultados é SVM-R e $D = 5$ é o melhor valor para D nesse caso.

Agora iremos avaliar a influência de τ na classificação. O valor máximo para τ depende do tamanho da série temporal n e da dimensão D, sendo limitado por:

$$\tau < \frac{n}{D-1}.$$

Quanto maior o valor de τ, maior deve ser o número das amostras da série temporal. Por exemplo, para $D = 5$ e $\tau = 1$, a série temporal deve ser, no mínimo, maior que 4; para $D = 5$ e $\tau = 2$, $n > 8$; para $D = 5$ e $\tau = 3$, $n > 12$; $D = 5$ e $\tau = 5$, $n > 20$; $D = 5$ e $\tau = 10$, $n > 40$; $D = 5$ e $\tau = 15$, $n > 60$ e assim segue. Em outras palavras, a medida que o τ aumenta, maior deve ser o tamanho das trajetórias. Com isso, trajetórias menores, e consequentemente dados, são descartados. Na Tabela 3 vemos como o valor de τ impacta no tamanho n da série.

Na Figura 4 vemos a classificação para $D = 5$ (o valor que apresentou me-

Tabela 3. Valores de n a medida que τ aumenta

τ	1	2	3	5	10	15
n	4341	4329	4290	4201	4024	3871

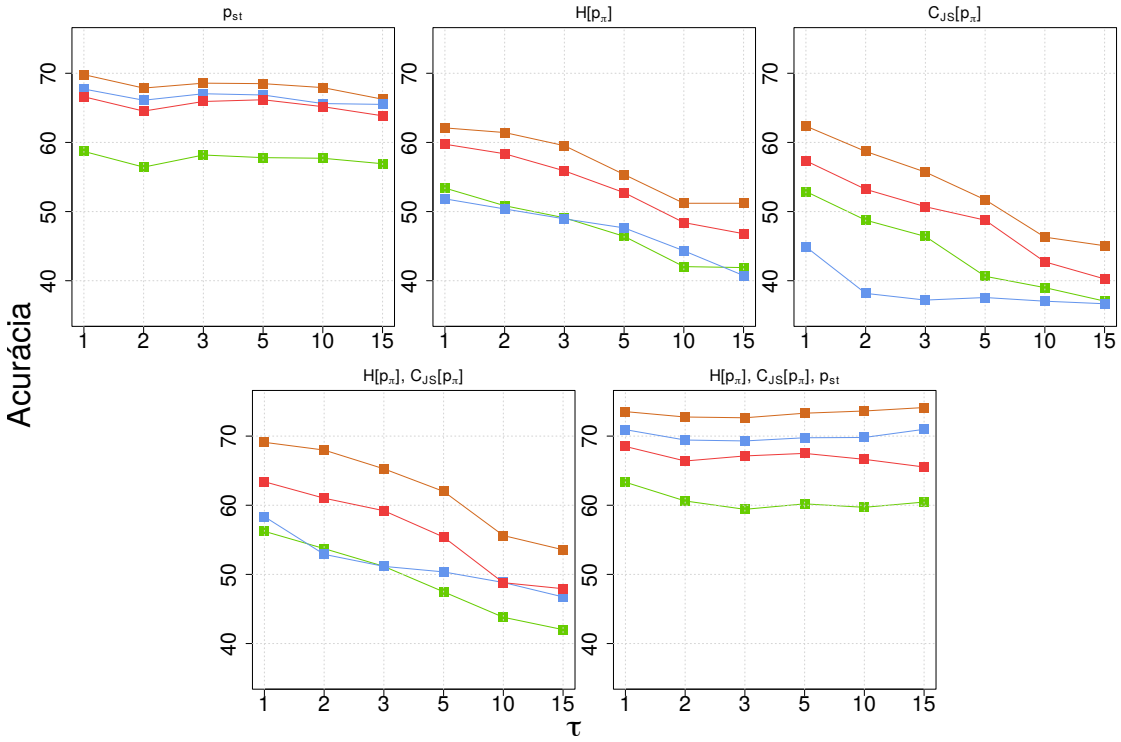

Figura 4. Acurácia das classificações com SVM-R (■), SVM-L (■), DT (■) e k-NN (■), utilizando diferentes valores de τ.

lhores resultados, no geral, entre os valores possíveis de D) e diferentes valores de $\tau \in \{1, 2, 3, 5, 10, 15\}$. É possível notar que, dentre os atributos classificados isoladamente, p_{st} apresenta um comportamento mais estável, sofrendo menor variação de acurácia a medida que o τ varia. Por outro lado, $H[p_\pi]$ e $C_{JS}[p_\pi]$ apresentam resultados de acurácia que vão decrescendo a medida que τ aumenta, tanto ao serem classificados isoladamente, quanto em conjunto. É possível notar também que p_{st} é capaz de manter sua robustez em relação ao valor de τ mesmo em conjunto com os outros atributos. Desse resultado, compreende-se que, mesmo com séries temporais menores, não há um comprometimento brusco do rendimento de p_{st}. Em outras palavras, p_{st} possui menor dependência dos valores dos parâmetros D e τ. Ademais, nessa figura também nota-se que $\tau = 1$, no geral, apresenta melhores resultados em todos os conjuntos classificados. Isso faz sentido, pois a medida que se aumenta o valor de τ, perde-se mais informações dentro da trajetória, devido a distância entre os pontos na janela deslizante. Além disso, novamente se tem SVM-R como o classificador com os melhores resultados de acurácia.

Com essa análise, iremos adotar os valores de $D = 5$ e $\tau = 1$ e o classificador

Tabela 4. Métricas de avaliação das classificações dos conjuntos dos dados

classes	Pre	Sen	F1	Acurácia
caminhada	93,98% ($\pm 1,42$)	80,66% ($\pm 1,05$)	86,77% ($\pm 0,93$)	81,05% ($\pm 1,52$)
bicicleta	55,83% ($\pm 3,57$)	82,46% ($\pm 4,86$)	66,45% ($\pm 3,85$)	
caminhada	96,47% ($\pm 0,66$)	91,20% ($\pm 0,77$)	93,75% ($\pm 0,65$)	92,03% ($\pm 0,90$)
ônibus	84,75% ($\pm 1,67$)	93,69% ($\pm 1,32$)	88,96% ($\pm 1,46$)	
caminhada	97,46% ($\pm 0,81$)	88,62% ($\pm 1,22$)	92,82% ($\pm 0,50$)	90,00% ($\pm 0,71$)
carro/táxi	75,20% ($\pm 2,75$)	93,70% ($\pm 1,75$)	83,40% ($\pm 1,41$)	
bicicleta	93,40% ($\pm 1,43$)	87,40% ($\pm 2,37$)	90,26% ($\pm 1,64$)	90,86% ($\pm 1,21$)
ônibus	88,81% ($\pm 1,80$)	94,10% ($\pm 0,80$)	91,37% ($\pm 1,00$)	
bicicleta	92,40% ($\pm 2,00$)	86,86% ($\pm 1,20$)	89,53% ($\pm 1,15$)	89,13% ($\pm 1,01$)
carro/táxi	85,85% ($\pm 1,24$)	91,70% ($\pm 1,94$)	88,67% ($\pm 0,85$)	
ônibus	81,05% ($\pm 3,82$)	85,12% ($\pm 0,95$)	83,00% ($\pm 2,04$)	81,74% ($\pm 1,60$)
carro/táxi	82,60% ($\pm 1,90$)	78,11% ($\pm 2,85$)	80,24% ($\pm 1,33$)	
caminhada	90,53% ($\pm 2,03$)	68,25% ($\pm 1,82$)	77,74% ($\pm 1,10$)	73,54% ($\pm 0,70$)
bicicleta	51,16% ($\pm 2,53$)	78,22% ($\pm 4,93$)	61,76% ($\pm 3,05$)	
ônibus	70,75% ($\pm 2,03$)	81,06% ($\pm 1,82$)	75,40% ($\pm 1,10$)	
carro/táxi	65,93% ($\pm 4,15$)	77,13% ($\pm 2,32$)	70,95% ($\pm 2,32$)	

SVM-R para análise das classes da série temporal do conjunto de dados utilizado. A classificação foi feita utilizando os três atributos, p_{st}, $H[p_\pi]$ e $C_{JS}[p_\pi]$, pois, como vimos anteriormente, consegue melhor valor de acurácia com os mesmos em conjunto. Na Tabela 4 vemos a classificação para os conjuntos das classes presentes na série temporal, juntamente com o intervalo de confiança com 95% de confiança. Além de acurácia, utilizamos sensibilidade (sen), precisão (pre) e F1-score como métricas de avaliação, definidas como:

- A sensibilidade (sen) explica o quão efetivamente o classificador identifica as predições positivas. Ou seja, a habilidade do nosso modelo de identificar corretamente quais indivíduos pertencem a uma classe;
- A precisão (pre) expressa a proporção de pontos no dado os quais o modelo diz que são relevantes e que de fato são realmente relevantes;
- F1-score é a média harmônica entre precisão e sensibilidade.

É possível notar que há maior dificuldade em distinguir entre transportes que, intuitivamente, percorrem distâncias semelhantes entre dois pontos consecutivos, como os conjuntos {caminhada, bicicleta} e {ônibus, carro/táxi}. Para transportes mais distintos, apresenta-se melhores resultados.

5. Conclusão

Neste trabalho, utilizamos Grafo de Transição de Padrões Ordinais para classificar modos de transporte de trajetórias de GPS. Transformamos os dados da série temporal original em Padrões Ordinais e em seguida transformamos estes padrões ordinais para o Grafo de Transição e a distribuição de probabilidade de frequência

dos padrões. Dessas duas transformações, extraiu-se os atributos Entropia de Permutação e Complexidade Estatística. Além desses, extraímos um novo atributo chamado probabilidade de auto-transição, o qual é proposto neste trabalho. Esse atributo é extraído das auto-transições do grafo e está diretamente relacionado com a correlação temporal da série temporal original.

O atributo proposto apresenta melhores resultados de classificação do que os atributos Entropia de Permutação e Complexidade Estatística, até mesmo quando estes estão combinados. Assim, pode-se afirmar que o atributo possui uma caracterização satisfatória da série temporal. Além disso, nosso atributo possui menor dependência dos valores de D e τ.

Para trabalhos futuros, pretende-se estudar mais sobre o atributo aqui proposto, especialmente sobre sua combinação com outros atributos, para que consigamos uma melhor caracterização, e consequentemente uma melhor classificação das série temporais. Além disso, pretendemos testar nossa abordagem em diferentes conjuntos de dados a fim de analisar a robustez de nossa técnica.

Agradecimentos

Este trabalho foi parcialmente financiado por CNPq e CAPES.

Referências

Aquino, A., Cavalcante, T., Almeida, E., Frery, A., and Rosso, O. (2015). Characterization of vehicle behavior with information theory. *The European Physical Journal B*, 88(10):257.

Aquino, A., Ramos, H., Frery, A., Viana, L., Cavalcante, T., and Rosso, O. (2017). Characterization of electric load with information theory quantifiers. *Physica A: Statistical Mechanics and its Applications*, 465:277 – 284.

Bagnall, A., Lines, J., Bostrom, A., Large, J., and Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660.

Bandt, C. and Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102.

Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*.

Guo, H., Zhang, J.-Y., Zou, Y., and Guan, S.-G. (2018). Cross and joint ordinal partition transition networks for multivariate time series analysis. *Frontiers of Physics*, 13(5):130508.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuño, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13):4972–4975.

Lines, J., Taylor, S., and Bagnall, A. (2018). Time series classification with hive-cote: The hierarchical vote collective of transformation-based ensembles. *ACM Trans. Knowl. Discov. Data*, 12(5):52:1–52:35.

Luque, B., Lacasa, L., Ballesteros, F., and Luque, J. (2009). Horizontal visibility graphs: Exact results for random time series. *Phys. Rev. E*, 80:046103.

Längkvist, M., Karlsson, L., and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11 – 24.

McCullough, M., Small, M., Iu, H., and Stemler, T. (2017). Multiscale ordinal network analysis of human cardiac dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2096):20160292.

Ravetti, M. G., Carpi, L. C., Gonçalves, B. A., Frery, A., and Rosso, O. A. (2014). Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph. *PLOS ONE*, 9(9):1–15.

Ribeiro, H. V., Jauregui, M., Zunino, L., and Lenzi, E. K. (2017). Characterizing time series via complexity-entropy curves. *Phys. Rev. E*, 95:062106.

Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., and Fuentes, M. A. (2007). Distinguishing noise from chaos. *Phys. Rev. Lett.*, 99:154102.

Small, M. (2013). Complex networks from time series: Capturing dynamics. In *2013 IEEE International Symposium on Circuits and Systems*, pages 2509–2512. IEEE.

Staniek, M. and Lehnertz, K. (2007). Parameter selection for permutation entropy measurements. *International Journal of Bifurcation and Chaos*, 17(10):3729–3733.

Wang, X., Ding, H., Trajcevski, G., Scheuermann, P., and Keogh, E. J. (2010). Experimental comparison of representation methods and distance measures for time series data. *CoRR*, abs/1012.2789.

Wilson, S. J. (2017). Data representation for time series data mining: time domain approaches. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(1):e1392.

Yang, Q. and Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, 5(04):597–604.

Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., and Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. *Scientific reports*, 7(1):7795.

Zheng, Y., Li, Q., Chen, Y., Xie, X., and Ma, W.-Y. (2008). Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM.

Zunino, L., Soriano, M. C., and Rosso, O. A. (2012). Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Phys. Rev. E*, 86:046210.