

Explorando a correlação espaço-temporal no agrupamento de sensores de cidades inteligentes

Morgana Gabi Gomes, Pedro H. Barros, Heitor S. Ramos

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brasil

{morganagabigomes, pedro.barros, ramosh}@dcc.ufmg.br

Abstract. *We propose a similarity function called SMELL-TS, based on deep metric learning, for classification of time series in the context of Zero-shot Learning, i.e., we are able to classify objects of classes that have not yet been used in the training set. The data are pre-processed by the Short-Term Fourier Transform, and later, they are mapped into two new representation spaces, called latent space and S-Space. We tested our model on a real dataset of sensors distributed in an intelligent building, seeking to group sensors co-located in the same environment. Our method presented better results when compared to other techniques found in the literature, with a gain of 15 % in the Room Accuracy metric – percentage of correctly co-located sensors grouped by SMELL-TS.*

Resumo. *Neste trabalho, propusemos uma função de similaridade chamada de SMELL-TS, baseada em aprendizagem de métrica profunda, para classificação de séries temporais no contexto de Zero-shot Learning, i.e., nosso método é apto a classificar objetos que pertencem a classes que ainda não foram usadas no conjunto de treinamento. Os dados são pré-processados pela Transformada de Fourier de Curto Termo, e posteriormente, são mapeados em dois novos espaços de representação, chamados de espaço latente e Espaço-S. Testamos nosso modelo num conjunto de dados reais de sensores distribuídos em um edifício inteligente, buscando agrupar sensores co-localizados no mesmo ambiente. Nosso método apresentou melhores resultados quando comparado com outras técnicas encontradas na literatura, com um ganho de 15 % na métrica de Room Accuracy – porcentagem de sensores co-localizados corretamente agrupados pelo SMELL-TS.*

1. Introdução

Um grande desafio dentro do contexto de aprendizagem de máquina é o problema de *Few-Shot Learning* (FSL), paradigma definido quando um modelo precisa aprender através de um número muito limitado de amostras (de uma ou mais classes) supervisionadas [Wang and Yao 2019]. De modo mais extremo, quando o número de amostras de uma (ou mais) classe a ser analisada é nulo, o problema passa a ser chamado de *Zero-shot Learning* (ZSL), i.e., reconhecer um novo conceito tendo apenas o conhecimento das classes conhecidas [Romera-Paredes and Torr 2015]. Por exemplo, classificar itens de um determinado tipo que não apareceu no conjunto de treino. O paradigma de ZSL é muito interessante para a área de classificação automática, pois assim como os humanos

têm capacidade de identificar novas espécies e categorias, um modelo também precisaria saber lidar com este cenário [Romera-Paredes and Torr 2015].

Podemos dividir o ZSL em dois tipos: Convencional e o Generalizado. ZSL Convencional (ZSLC) consiste num modelo em que, no conjunto teste, todos os itens pertencem a classes não vistas no treinamento, enquanto no caso de ZSL Generalizado (ZSLG) o conjunto teste é composto tanto de classes vistas no treino, quanto de classes não vistas [Han et al. 2021]. É bom destacar que, no caso do ZSLG, o modelo pode interpretar de forma errônea os itens das classes não vistas no treino, devido ao problema de viés com as classes vistas. Entre as aplicações, ZSL é útil para situações em que há um grande desbalanceamento entre a quantidade de itens de cada classe diferente [Han et al. 2021]; bem como no contexto de classificação de mundo aberto [Schroff et al. 2015].

Funções de distância quantificam a similaridade entre dois elementos (e.g., distância Euclidiana). Dentre essas abordagens, algumas propostas medem a similaridade entre amostras usando uma função que aprende o contexto dos dados. Essa área de estudo é chamada de Aprendizagem de Métrica [Oh Song et al. 2016]. Entretanto, soluções tradicionais de Aprendizagem de Métrica possuem limitações, dentre elas, este paradigma tipicamente usa transformações lineares, uma suposição dificilmente encontrada no mundo real [Barros et al. 2020]. A Aprendizagem de Métrica Profunda foi proposta para resolver esta limitação usando modelos não lineares advindos de redes neurais [Cui et al. 2016, KAYA and BİLGE 2019]. Devido a esta propriedade, essas técnicas tendem a obter melhores resultados quando comparados às propostas convencionais da literatura.

Encontramos vários trabalhos usando aprendizado métrico com redes neurais [Sohn 2016, Wang et al. 2019]. Mesmo apresentando resultados promissores, essas propostas apresentam algumas limitações, como convergência lenta, o que torna a otimização do modelo uma tarefa desafiadora.

Em nosso trabalho anterior [Barros et al. 2020] hipotetizamos que o espaço de representação latente poder ser aperfeiçoado com um espaço auxiliar. Por exemplo, abordagens comuns de aprendizagem de métrica profunda baseadas em NNs extraem um espaço latente que codifica pontos semelhantes e diferentes, mas não a separabilidade entre eles. No entanto, esta representação única é limitada, uma vez que não captura informações de pares.

Esse artigo propõe um modelo de Aprendizagem de Métrica Profunda, chamado SMELL-TS, que tem capacidade de classificar séries temporais em um contexto de ZSL. Para avaliar nosso modelo, o testamos em um conjunto de dados proposto por [Hong et al. 2013], composto por sensores em um edifício inteligente. Os sensores são responsáveis por prover dados que dão suporte à tomada de decisões automatizadas como abrir uma janela, ligar um ar-condicionado, etc. O objetivo é reagrupar os sensores que estão co-localizados no mesmo cômodo do edifício, e está no contexto de ZSLC devido ao fato dos cômodos vistos no treinamento não serem vistos no teste.

O SMELL-TS é composto de três partes: Inicialmente, usamos a transformada de Fourier de curto termo para extrair as características no domínio da frequência dos dados dos sensores. Em seguida, usamos uma rede neural para obter uma nova representação dos dados, chamado de espaço de representação latente. Nessa nova representação, induzimos

o modelo a aproximar os sensores localizados no mesmo ambiente. Por fim, agrupamos os sensores com a otimização de algoritmos genéticos usando uma heurística proposta no novo espaço de representação dos dados.

Medimos nosso resultado como a porcentagem dos sensores corretamente associados com cada sala, chamado de Edge Accuracy, e com a porcentagem de salas corretamente reagrupadas, chamada de Room Accuracy. Nosso modelo obteve o melhor resultado comparado com outras técnicas encontradas na literatura, com melhora de 8.34% e 15% no Edge Accuracy e Room Accuracy, respectivamente.

As principais contribuições deste trabalho são:

- Propusemos uma abordagem para quantificar a similaridade entre séries temporais baseada na correlação espaço-temporal.
- Através da nossa função de similaridade, podemos agrupar sensores co-localizados no mesmo espaço físico.
- Nossa abordagem é adequada para o paradigma de ZSL.

Organizamos o trabalho como segue: Seção 2 encontram-se os trabalhos relacionados; Seção 3 apresenta nossa proposta; Seção 4 descreve a metodologia usada para analisar os dados; Seção 5 apresentam os resultados principais; e Seção 6 conclui esse trabalho, e discute algumas perspectivas futuras.

2. Trabalhos Relacionados

Nas áreas de Aprendizagem de Métrica, temos trabalhos na literatura como [Xing et al. 2002], que dado exemplos de pares de pontos de mesma classe, aprende uma transformação linear (chamada de Distância de Mahalanobis) que respeite as relações de distância no espaço obtido pela transformação, baseado em um problema de otimização; Em [Globerson and Roweis 2005], os autores propuseram um algoritmo que aprende uma métrica quadrática Gaussiana, para classificação, com o objetivo de colapsar todos os exemplos de cada classe, de modo que, classes diferentes colapsem em locais diferentes.

Já na área de Aprendizagem de Métrica Profunda, temos [Bromley et al. 1994], que propõe um algoritmo para reconhecer assinaturas escritas manualmente utilizando uma rede neural dupla, chamada de rede siamesa, que tem como objetivo aproximar, em um novo espaço de representação, os vetores de mesmo rótulo e afastar os vetores de rótulos diferentes para formação de agrupamentos. O modelo tem algumas dificuldades de convergência, necessitando de uma maior etapa de refinamento da arquitetura da rede neural.

Baseado na rede siamesa, [Schroff et al. 2015] desenvolve um modelo em forma de tripletos – sendo 3 redes idênticas com o mesmo peso – para potencializar o sistema de reconhecimento de rostos. Cada tripla é formada por um elemento âncora, e os outros dois sendo um positivo (mesma classe que o elemento âncora) e um negativo (classe diferente). Dessa forma o modelo busca estimar uma nova representação dos dados que visa “empurrar” o negativo para longe do âncora e “puxar” o positivo para perto. Porém, devido ao fato de se utilizar triplas, o conjunto de dados escala com $O(n^3)$.

Por datasets com maior granularidade sofrerem de problemas como muitas categorias e poucos dados de treino, [Cui et al. 2016] desenvolveu um modelo de Aprendizagem de Métrica Profunda inspirado na rede de tripletos para facilitar o aprendizado

nesses datasets; [Wang et al. 2017] propõem um sistema de pesquisa de imagens com o intuito de capturar as abstrações utilizando um novo modelo de Aprendizagem de Métrica Profunda com uma função de custo baseado no ângulo entre os objetos.

Já para trabalhos envolvendo ZSL, vemos que este paradigma torna-se mais desafiador, necessitando de modelos geralmente adaptados especificamente para esse contexto. [Rabiner and Juang 1986] explicou como a técnica de Cadeias de Markov pode ser aplicada no contexto de processamento de fala; [Berndt and Clifford 1994] desenvolveu um algoritmo de detecção de padrões baseado em *Dynamic Time Warping* para uso na detecção de arquivos de áudios.

Pela dificuldade em se determinar distância ou similaridade entre séries temporais multivariadas, [Che et al. 2017] projetou um modelo que combina um novo método de alinhamento temporal (chamado de *expected alignment*) e uma métrica local dependente dos dados e treinada por redes neurais profundas. [Grabocka and Schmidt-Thieme 2018] utilizou redes neurais profundas para alinhar os índices de séries temporais num espaço de representação não-linear.

Devido a análise contextual completa de dados produzidos por edifícios inteligentes ser uma tarefa custosa e demorada, [Hong et al. 2019] propôs um novo método estatístico que se baseia no fato dos sensores desses edifícios estarem sujeitos aos mesmos eventos, para buscar uma correlação espaço-temporal entre eles; [Li et al. 2020] usou uma rede neural convolucional em forma de tripletos para gerar uma nova representação de dados de sensores de um prédio inteligente com o objetivo de medir a correlação espaço-temporal entre eles. Os autores pré-processam os dados da série temporal utilizando a Transformada de Fourier de Curto Termo; [Barros et al. 2020] desenvolveram um framework que usa redes neurais siamesas para gerar duas novas representações distintas dos dados, sendo uma chamada de espaço latente e outra de espaço-S, usando as duas posteriormente para calcular a similaridade entre duas entradas.

Nossa proposta é inspirada em [Barros et al. 2020], mas adaptada para o contexto de séries temporais, diferentemente da proposta original, nosso modelo inicialmente processa as séries temporais pela Transformada de Fourier de Curto Termo e depois treina uma rede neural, com uso de um espaço latente e de um espaço de similaridade – chamado de espaço-S – para estimar similaridade entre as duas entradas. Avaliamos nosso modelo em um contexto de salas num edifício inteligente, e usamos Algoritmos Genéticos para reagrupar sensores co-localizados nesse ambiente. As salas (classes) vistas no treino não aparecem no teste, o que coloca nosso modelo no contexto de ZSL.

3. Encoder para Aprendizado de Métrica de Distância Supervisionado com Espaço de Similaridade Adaptado para Séries Temporais (SMELL-TS)

Uma série temporal \mathcal{T} é uma sequência de p observações de uma variável através do tempo, i.e., $\mathcal{T} = \{t_{n_1}, t_{n_2}, \dots, t_{n_p}\}$, onde $t_{n_i} \in \mathbb{R}$ é o valor da série temporal no tempo $n_i \in \mathbb{Z}$. No SMELL-TS, os dados (séries temporais) são pré processados pela transformada de Fourier de curto termo (STFT) [Daubechies 1990], que extrai características da série temporal no domínio da frequência (seção 3.1), i.e., o STFT transforma a série temporal em uma nova representação \mathcal{I} . Essa nova representação é utilizada como entrada de um encoder para estimar um novo espaço de representação, chamado Espaço de Representação Latente – seção 3.2. Depois disso, o modelo usa uma representação auxi-

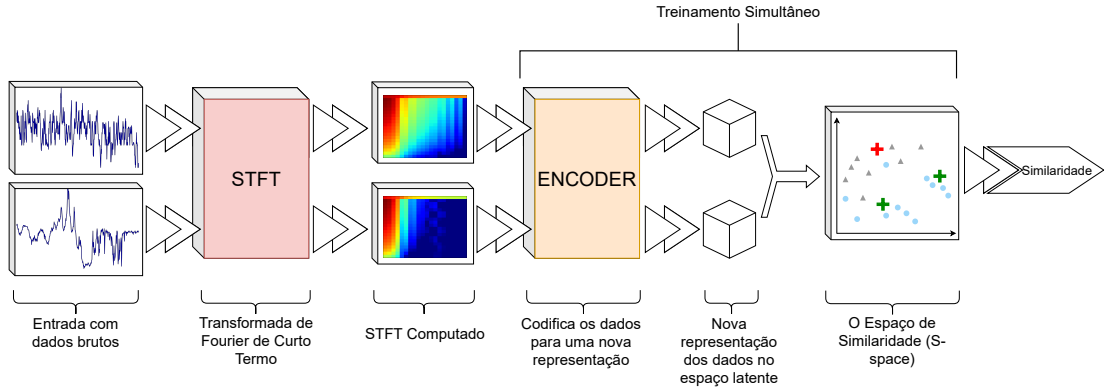


Figura 1. Esquemático do SMELL-TS para estimar similaridade entre séries temporais.

liar (espaço-S) para estimar a similaridade entre duas séries temporais – seção 3.3. Um esquemático do nosso método é apresentado na Figura 1, que retrata os passos mencionados anteriormente.

3.1. Transformada de Fourier de Curto Termo (STFT)

Dados de sensores apresentam uma variedade de dinâmicas diferentes e muitas irregularidades entre eles. Mesmo que eles seja correlatos ou localizados no mesmo ambiente, suas variações podem não acontecer simultaneamente, i.e., um sensor de luz e temperatura, que apesar de apresentarem correlação, apresentam variação diferente quando submetidos ao mesmo ambiente (e.g., luz do Sol num dia ensolarado).

Outro problema é a taxa de atualização dos dados dos sensores, que podem depender de diversos fatores (e.g., tipo do sensor e fabricante). No fim, a série temporal apresenta diferentes tamanhos e medidas em diferentes momentos, ou seja, dada duas séries temporais $\mathcal{T} = \{t_{n_1}, t_{n_2}, \dots, t_{n_p}\}$ e $\mathcal{S} = \{s_{m_1}, s_{m_2}, \dots, s_{m_p}\}$, não temos nenhuma garantia que $n_i = m_i$.

Para mitigar esse problema, usamos a STFT para extrair uma nova representação dos dados das séries temporais. É bom notar que as séries temporais presentes no conjunto de dados analisado são rotuladas, e que nós sabemos a localização dos sensores. Além disso, a transformação usada no STFT não altera os rótulos do sensor, que neste trabalho consiste na sala em que o sensor está localizado.

3.2. Espaço de Representação Latente

No contexto de Aprendizado de Máquina, um encoder é uma rede neural (NN) que busca transformar os dados de entrada numa nova representação. Seja $\mathcal{I} = \{i_1, i_2, \dots, i_l\}$ o conjunto dos dados das séries temporais após o pré-processamento no STFT, a função de encoder $f_{\Theta} : \mathcal{I} \rightarrow \mathcal{L}$ mapeia o conjunto \mathcal{I} num novo espaço de representação latente \mathcal{L} com pesos Θ .

Usamos o espaço de representação latente para o problema de agrupamento de sensores co-localizados e, inspirados por [Li et al. 2020], normalizamos todos os vetores $\ell_j = f_{\Theta}(i_j)$, de modo que $\|\ell_j\|_2 = 1$, para todo $i_j \in \mathcal{I}$.

Para calcular a similaridade entre os vetores, usamos a equação $\rho(\ell_i, \ell_j) = 1 - \frac{\|\ell_i - \ell_j\|_2^2}{\|\ell_i\|_2^2 + \|\ell_j\|_2^2}$, pois devido à normalização em \mathcal{L} , essa equação é equivalente ao coeficiente de correlação de Pearson [Li et al. 2020]. Portanto, quanto menor a distância Euclidiana entre dois vetores que se encontram exclusivamente no espaço de representação latente, maior é a correlação de Pearson.

3.3. Espaço de Similaridade (espaço-S)

Diferentemente da literatura, usamos uma abordagem para quantificar similaridade utilizando um espaço auxiliar chamado *Espaço-S*, como podemos ver na Figura 1 e iremos detalhar agora. Definimos a função $f^S : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{S}$, que constrói o espaço-S (\mathcal{S}) como

$$\begin{aligned} \mathbf{s}_{mn} &= f^S(\mathbf{i}_m, \mathbf{i}_n) \\ &= |f_\Theta(\mathbf{i}_m) - f_\Theta(\mathbf{i}_n)| \\ &= |\ell_m - \ell_n| \\ &= (|\ell_m^1 - \ell_n^1|, |\ell_m^2 - \ell_n^2|, \dots, |\ell_m^n - \ell_n^n|) \end{aligned} \quad (1)$$

onde ℓ_m^i é a i -ésima característica da m -ésima amostra dos dados no espaço de representação latente \mathcal{L} .

Com isso, o vetor \mathbf{s}_{mn} , pertencente ao espaço-S, representa um par de séries temporais no espaço de representação latente $(\mathbf{i}_m, \mathbf{i}_n)$. Definimos que um dado vetor \mathbf{s}_{mn} é chamado de vetor de similaridade se, e somente se, \mathbf{i}_m tem a mesma classe que \mathbf{i}_n . Da mesma forma, definimos que \mathbf{s}_{mn} é um vetor de dissimilaridade se \mathbf{i}_m possui classe diferente de \mathbf{i}_n .

Para quantificar a similaridade, o SMELL-TS estima um conjunto de marcadores \mathcal{M} . Os marcadores quantificam a semântica do vetor localizado no espaço-S, i.e., nós definimos um marcador como (dis)similar representando-o pelos conjuntos (\mathcal{M}^-) \mathcal{M}^+ , respectivamente. Com isso, quanto mais próximo um vetor \mathbf{s}_{mn} está do marcador de similaridade, maior é a probabilidade do vetor representar séries temporais similares. Dado o conjunto dos marcadores de similaridade $\mu_i^+ \in \mathcal{M}^+$ e dissimilaridade $\mu_i^- \in \mathcal{M}^-$ temos que

$$\mathcal{M} = \mathcal{M}^+ \cup \mathcal{M}^- = \{\mu_i^+\}_{i=1}^k \cup \{\mu_j^-\}_{j=k+1}^w, \quad (2)$$

onde $\mathcal{M}^- \cap \mathcal{M}^+ = \emptyset$, k é o número de marcadores positivos, $w - k$ é o número de marcadores negativos e w é o total de marcadores.

Seja $\Sigma = \{\Theta, \mathcal{M}\}$ o conjunto de parâmetros do SMELL-TS, que correspondem, respectivamente, aos pesos do encoder e os marcadores. O parâmetro Σ é estimado durante o treinamento do SMELL-TS. Com isso, nossa proposta estima simultaneamente o espaço de representação latente e o espaço-S. Enquanto o espaço-S tende a encontrar uma similaridade entre pares de séries temporais, o espaço latente extrai uma nova representação que permite calcular a correlação espaço-temporal entre duas séries temporais distintas. Portanto, o treinamento simultâneo de ambos os espaços tende a aumentar a performance do SMELL-TS.

Diferentemente de [Barros et al. 2020], realizamos uma investigação na literatura acerca de funções que poderiam quantificar a similaridade entre vetores no espaço-S, e

escolhemos a função mais adequada devido a suas propriedades, bem como desempenho. Dessa forma, para cálculo da similaridade, usamos uma função baseada em Wavelet [Li Zhang et al. 2004]. Dado o vetor s_{mn} , o marcador $\mu_k \in \mathcal{M}$ e o kernel k_w , o SMELL-TS quantifica a similaridade como

$$q_{mn}^r = \frac{k_w(s_{mn}, \mu_r)}{\sum_{\mu_{r'} \in \mathcal{M}} k_w(s_{mn}, \mu_{r'})}, \quad (3)$$

com

$$k_w(x, y) = \prod_{i=1}^N \left[\cos(1.75(x^i - y^i)) \exp\left(-\frac{(x^i - y^i)^2}{2}\right) \right], \quad (4)$$

onde x^i e y^i são as i -ésimas características das amostras x e y no espaço-S.

Por fim, consideramos que q_{mn}^+ é a probabilidade do par (ℓ_m, ℓ_n) estar localizado no mesmo ambiente (similar). Então, temos $q_{mn}^+ = \sum_p q_{mn}^p$ para todo $\mu_p \in \mathcal{M}^+$. Consequentemente, q_{mn}^- é a probabilidade do par (ℓ_m, ℓ_n) estar localizado num ambiente diferente (dissimilar), e portanto, $q_{mn}^- = \sum_t q_{mn}^t$ para todo $\mu_t \in \mathcal{M}^-$. Devido à restrição de $\mathcal{M}^- \cap \mathcal{M}^+ = \emptyset$, temos que $q_{mn}^+ + q_{mn}^- = 1$.

Para treinar o modelo, usamos a função de custo entropia cruzada binária. Para a função de custo do SMELL-TS, $\mathcal{L}(\{\mathcal{I} \times \mathcal{I}\})$, a função de custo entropia cruzada J é responsável por guiar o processo de otimização para encontrar os parâmetros otimizados. Seja q_{mn}^+ o resultado do SMELL-TS, u_{mn} o valor associado a cada rótulo da similaridade dos pares (i_m, i_n) de modo que $u_{mn} = 1$ se s_{mn} representa um vetor de similaridade no espaço-S, e $u_{mn} = 0$ caso contrário. Logo, a função de custo é definida como

$$\mathcal{L}(\mathcal{I} \times \mathcal{I}) = \mathbb{E}_{i_m, i_n} [u_{mn} \log(q_{mn}^+) + (1 - u_{mn}) \log(q_{mn}^-)]. \quad (5)$$

4. Metodologia

4.1. Conjunto de dados

Para treinar e avaliar o SMELL-TS, usamos o conjunto de dados KETI, proposto por [Hong et al. 2013], que consiste em 9 dias de dados reais observados num escritório nos EUA, com 50 salas espalhadas por 4 andares. Cada sala contém 4 sensores diferentes – temperatura, umidade, CO₂ e luminosidade, resultando em um conjunto de dados composto de 200 séries temporais diferentes. Como usamos a taxa de k-fold com $k = 5$, i.e., a taxa de treino/teste é igual a 40/10, logo temos 160 séries temporais para treino, correspondendo a 40 classes diferentes, e 40 séries temporais para teste, correspondendo a 10 classes diferentes. Note que os sensores têm frequência de atualização diferente, e além disso, as classes utilizadas no teste não foram utilizadas no treinamento.

4.2. Algoritmos Genéticos

Como nossa proposta estima a similaridade entre duas séries temporais, precisamos de um método para agrupá-las. Neste trabalho, usamos a otimização de Algoritmos Genéticos (GA) para agrupar os sensores localizados na mesma sala. O GA busca a melhor combinação possível através da competição entre indivíduos numa população. Uma nova população é gerada para cada rodada do GA, contendo indivíduos da rodada anterior e novos indivíduos (gerados de pares de indivíduos da mesma população, tendo sua

operação chamada de crossover). Além disso, cada indivíduo tem uma probabilidade de mutação.

Neste trabalho, modelamos um indivíduo como uma solução possível de agrupamento, i.e., cada indivíduo tem precisamente 10 conjuntos de 4 séries temporais (sensores), totalizando então as 10 salas usadas no teste do GA. Por fim, uma heurística específica foi usada para medir a pontuação de cada indivíduo (chamada fitness) e selecionar o indivíduo com o melhor fitness - nossa solução de agrupamento. Então, para esse problema, essa proposta é eficiente, em contraste com o excessivo custo para olhar cada combinação possível.

A heurística foi adaptada da Savage-loss [Masnadi-Shirazi and Vasconcelos 2008] e funciona calculando a média de todos os pares possíveis dentro de cada sala, e então calculando a média da heurística de todas as salas. Devido ao fato de termos 10 salas no teste e 4 sensores por sala, cada sala contém 6 pares possíveis para o cálculo da correlação de Pearson representada como $\rho(m, n)$. Definimos um indivíduo $\mathcal{P} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{10}\}$, onde $\mathcal{S}_i = \{f_{i_1}, f_{i_2}, f_{i_3}, f_{i_4}\}$ corresponde à i -ésima sala de \mathcal{P} e f_{ij} corresponde ao j -ésimo sensor de \mathcal{S}_i . Portanto, o GA utiliza a função de fitness

$$H(\mathcal{P}) = \frac{1}{10} \sum_{i=1}^{10} \left(\frac{1}{6} \sum_{m=1}^3 \sum_{n=m+1}^4 \sqrt{\exp(\rho(\ell_{im}, \ell_{in}))} \right), \quad (6)$$

onde $\ell_{im} = f_{\Theta}(f_{im})$ corresponde ao mapeamento encontrado pela função de encoder do sensor f_{im} no espaço de representação latente estimado pelo SMELL-TS.

4.3. Otimização dos Hiperparâmetros

A arquitetura do encoder utilizada neste trabalho corresponde a um conjunto de várias camadas convolucionais, conforme proposto em [Li et al. 2020]. No SMELL-TS, usamos a busca em grade para calibrar os hiper-parâmetros do modelo. Utilizamos 5 marcadores, três deles sendo de similaridade e dois de dissimilaridade, e treinamos nosso modelo com 10.000 rodadas de treinamento com mini-batch (40 pares de séries temporais por rodada). Além disso, para a rede neural, usamos a taxa de aprendizado de 0,01, e usamos o otimizador de Gradiente Estocástico Descendente (SGD).

Treinamos o Algoritmo Genético com 2.000 iterações, com 100 indivíduos (população) em cada rodada e uma quantidade de sobreviventes por rodada igual a 20 indivíduos, taxa de crossing-over igual a 0,1, e taxa de mutação igual a 0,001.

4.4. Métricas de Avaliação

Para avaliar o modelo, usamos duas métricas diferentes, em concordância com [Li et al. 2020]:

- Edge Accuracy: corresponde a porcentagem de sensores corretamente atribuídos às suas respectivas salas.
- Room Accuracy: corresponde a porcentagem de salas corretamente agrupadas.

Usamos k-fold com $k = 5$; consequentemente, a proporção treino/teste corresponde a 40/10 salas para cada rodada do k-fold para a relação de inferência entre as séries temporais, de acordo com [Li et al. 2020].

Tabela 1. Resultados mostrados como intervalo de 95% de confiança estatística para cada experimento. Os melhores resultados estão em negrito e * representa os valores com uma diferença estatística significativa.

Proposta	Edge Accuracy	Room Accuracy
HMM [Rabiner and Juang 1986]	12.67	4.0
MEMO [Hong et al. 2019]	19.0	10.00
DTW [Berndt and Clifford 1994]	37.27 ± 2.29	14.40 ± 2.80
DECADE [Che et al. 2017]	12.47 ± 1.53	2.00 ± 5.72
WN [Grabocka and Schmidt-Thieme 2018]	17.47 ± 2.07	8.00 ± 9.35
TN [Schroff et al. 2015]	25.73 ± 1.85	6.00 ± 8.77
SSN [Li et al. 2020]	67.93 ± 8.25	50.20 ± 13.65
STN [Li et al. 2020]	88.61 ± 1.98	*80.00 ± 3.61
SMELL-TS	*96.00 ± 4.67	*92.00 ± 9.34

5. Resultados e Discussão

Para avaliar a eficiência do nosso modelo, o comparamos com oito modelos encontrados na literatura. Note que por separarmos as salas através da técnica de k-fold, as salas (classes) usadas no treinamento não aparecem no teste. Com isso, essa tarefa é definida como problema de ZSL, requerendo técnicas adequadas para esse propósito.

A Tabela 1 mostra a média de Edge Accuracy e Room Accuracy junto com os desvios padrões correspondentes. Com exceção do HMM [Rabiner and Juang 1986] e MEMO [Hong et al. 2019], todos os outros modelos são supervisionados. Como descrito na seção 4.4, os resultados correspondem à média das 5 rodadas.

Analisando a métrica de Edge Accuracy, o SMELL-TS obteve a maior média, 96,00% intervalo de confiança de (91,33%, 100%). Para o STN [Li et al. 2020] e SSN [Li et al. 2020], Edge Accuracy corresponde a 88,61% – com intervalo de confiança de (86,63%, 90,59%) – e 67,93% com intervalo de confiança de (59,68%, 76,18%), sendo assim o segundo e terceiro melhores resultados, respectivamente. Para verificar a diferença estatística dos nossos resultados, realizamos um teste t de student com 95% de confiança entre os dois melhores resultados (SMELL-TS e STN). Obtivemos um p-valor de 0,0134, fazendo com que rejeitássemos a hipótese de que ambos modelos possuem resultados estatisticamente semelhantes.

Considerando a métrica de Room Accuracy, observamos que o segundo e terceiro melhores resultados são obtidos por STN [Li et al. 2020] e SSN [Li et al. 2020], respectivamente. STN [Li et al. 2020] obteve 80% de média no Room Accuracy com intervalo de confiança de (76,39%, 83,61%). STN [Li et al. 2020] obteve 50.20% de média no Room Accuracy com intervalo de confiança de (36,55%, 83,85%). Como no experimento anterior, verificamos a diferença estatística entre os dois melhores resultados, e encontramos um p-valor igual a 0,051, indicando que o resultado do SMELL-TS não possui diferença estatística quando comparado com o STN.

Nossa proposta estima um novo espaço de representação (espaço-S) para estimar a similaridade entre pares de séries temporais. Cada par de séries temporais é representado

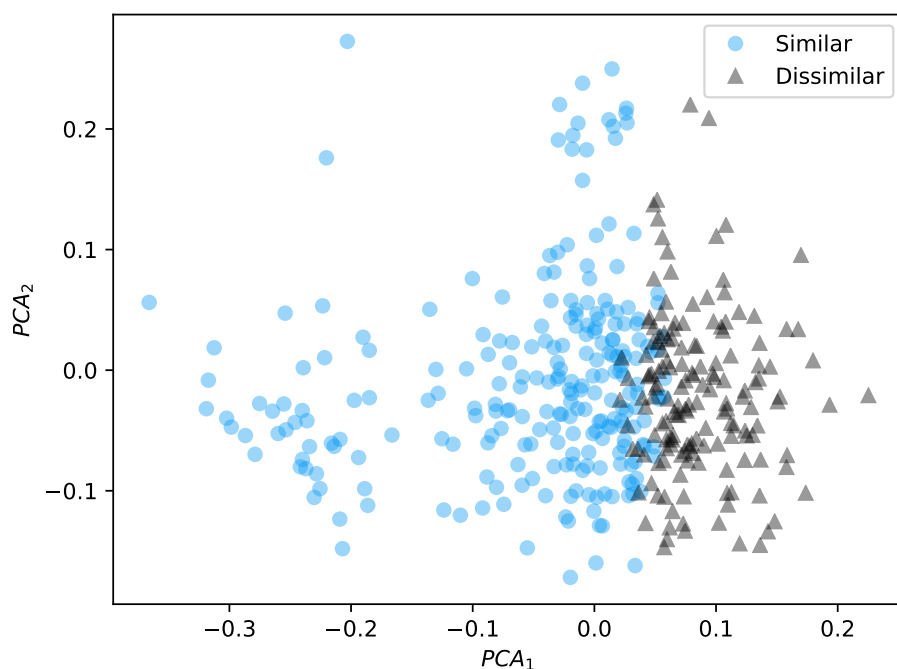


Figura 2. Espaço-S estimado pelo SMELL-TS. Nesse espaço, nossa proposta separa os vetores de similaridade/dissimilaridade em regiões diferentes.

como um vetor com semânticas (similaridade ou dissimilaridade). A Figura 2 mostra uma representação do espaço-S para 400 pares aleatórios de séries temporais, 200 dos quais são similares e 200 são dissimilares. Para melhor visualização, realizamos uma redução de dimensionalidade usando PCA. Observamos que nossa proposta tende a separar os vetores de similaridade e dissimilaridade em regiões diferentes no espaço-S.

Como visto na seção 3.2, a normalização do espaço de representação latente gera uma correspondência entre a correlação de Pearson e a distância Euclidiana. O SMELL-TS encontra uma representação para o espaço de representação latente que maximiza a correlação de Pearson entre as séries temporais.

Por fim, com a matriz de correlação obtida pela nossa proposta, usamos a heurística proposta na seção 4.2 para realizar o agrupamento das séries temporais. O treinamento do Algoritmo Genético para a primeira rodada do k-fold pode ser visto na Figura 3. Essa figura mostra o melhor fitness individual na população em cada rodada e o fitness do ground-truth, o indivíduo que tem todas as séries temporais corretamente agrupadas. Observamos que, a cada rodada de treinamento, a melhor solução encontrada pelo GA tende ao fitness do ground-truth.

6. Conclusão e trabalhos futuros

Esse trabalho propõe o SMELL-TS para calcular a similaridade entre séries temporais baseado numa correlação espaço-temporal entre sensores em um edifício inteligente. Com a heurística extraída do espaço de representação latente, usamos GA para realizar o agrupamento dos sensores co-localizados, i.e., localizados na mesma sala.

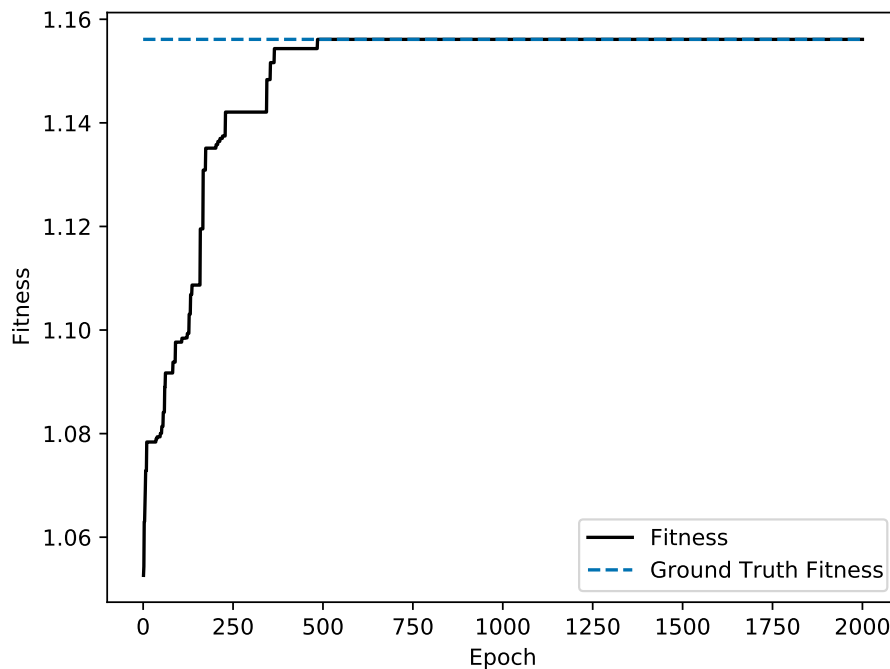


Figura 3. Evolução do melhor fitness individual entre rodadas (linha preta), e a linha pontilhada azul representa o fitness do grupo correto.

Neste trabalho, avaliamos nossa proposta no contexto de ZSL, e realizamos testes estatísticos para indicar a relevância dos nossos resultados.

O método proposto apresenta resultados satisfatórios quando comparado com alguns modelos na literatura. Além do mais, apresentamos evidências de que nosso método está agrupando sensores localizados no mesmo ambiente.

Como trabalhos futuros, avaliaremos o SMELL-TS em mais conjunto de dados, bem como diferentes cenários de ZSL e investigaremos o comportamento do espaço de representação latente formado por Redes Generativas Adversariais (GANs). E também, temos a intenção de estender nossos experimentos para o contexto de Aprendizado Federado.

Agradecimentos

Esse trabalho foi parcialmente financiado pela CNPq (grant 311750/2018-4), CAPES, FAPEMIG, e processo nº 2020/05121-4, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- [Barros et al. 2020] Barros, P. H., Queiroz, F., Figueredo, F., dos Santos, J. A., and Ramos, H. S. (2020). A new similarity space tailored for supervised deep metric learning.
- [Berndt and Clifford 1994] Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Confe-*

- rence on *Knowledge Discovery and Data Mining*, AAAIWS'94, page 359–370. AAAI Press.
- [Bromley et al. 1994] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1994). Signature verification using a "siamese" time delay neural network. *Advances in neural information processing systems*, pages 737–737.
- [Che et al. 2017] Che, Z., He, X., Xu, K., and Liu, Y. (2017). Decade: a deep metric learning model for multivariate time series. In *KDD workshop on mining and learning from time series*. sn.
- [Cui et al. 2016] Cui, Y., Zhou, F., Lin, Y., and Belongie, S. (2016). Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1153–1162.
- [Daubechies 1990] Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005.
- [Globerson and Roweis 2005] Globerson, A. and Roweis, S. (2005). Metric learning by collapsing classes. *Advances in neural information processing systems*, 18:451–458.
- [Grabocka and Schmidt-Thieme 2018] Grabocka, J. and Schmidt-Thieme, L. (2018). Neuralwarp: Time-series similarity with warping networks. *CoRR*, abs/1812.08306.
- [Han et al. 2021] Han, Z., Fu, Z., Chen, S., and Yang, J. (2021). Contrastive embedding for generalized zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2371–2381.
- [Hong et al. 2019] Hong, D., Cai, R., Wang, H., and Whitehouse, K. (2019). Learning from correlated events for equipment relation inference in buildings. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '19, page 203–212, New York, NY, USA. Association for Computing Machinery.
- [Hong et al. 2013] Hong, D., Ortiz, J., Whitehouse, K., and Culler, D. (2013). Towards automatic spatial verification of sensor placement in buildings. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys'13, page 1–8, New York, NY, USA. Association for Computing Machinery.
- [KAYA and BİLGE 2019] KAYA, M. and BİLGE, H. S. (2019). Deep metric learning: A survey. *Symmetry*, 11(9).
- [Li et al. 2020] Li, S., Hong, D., and Wang, H. (2020). Relation inference among sensor time series in smart buildings with metric learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):4683–4690.
- [Li Zhang et al. 2004] Li Zhang, Weida Zhou, and Licheng Jiao (2004). Wavelet support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):34–39.
- [Masnadi-Shirazi and Vasconcelos 2008] Masnadi-Shirazi, H. and Vasconcelos, N. (2008). On the design of loss functions for classification: theory, robustness to outliers, and savageboost. *sign $[f^*(x)]$* , 10:2.
- [Oh Song et al. 2016] Oh Song, H., Xiang, Y., Jegelka, S., and Savarese, S. (2016). Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4004–4012.
- [Rabiner and Juang 1986] Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16.

- [Romera-Paredes and Torr 2015] Romera-Paredes, B. and Torr, P. (2015). An embarrassingly simple approach to zero-shot learning. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2152–2161, Lille, France. PMLR.
- [Schroff et al. 2015] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823.
- [Sohn 2016] Sohn, K. (2016). Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems 29*, pages 1857–1865. Curran Associates, Inc.
- [Wang et al. 2017] Wang, J., Zhou, F., Wen, S., Liu, X., and Lin, Y. (2017). Deep metric learning with angular loss. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2593–2601.
- [Wang et al. 2019] Wang, X., Han, X., Huang, W., Dong, D., and Scott, M. R. (2019). Multi-similarity loss with general pair weighting for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5022–5030.
- [Wang and Yao 2019] Wang, Y. and Yao, Q. (2019). Few-shot learning: A survey. *CoRR*, abs/1904.05046.
- [Xing et al. 2002] Xing, E. P., Ng, A. Y., Jordan, M. I., and Russell, S. (2002). Distance metric learning, with application to clustering with side-information. In *Proceedings of the 15th International Conference on Neural Information Processing Systems*, NIPS’02, page 521–528, Cambridge, MA, USA. MIT Press.