

SELEÇÃO AUTOMÁTICA DE PARÂMETROS INICIAIS DO ALGORITMO K-SEGMENTOS COM TEACHING-LEARNING-BASED OPTIMIZATION

ANDRÉ DE AGUIAR BRAGA*, DANTON DIEGO FERREIRA†, BRUNO HENRIQUE GROENNER BARBOSA‡

**Núcleo de Eletricidade e Automação, DEG-UFLA
Lavras- MG*

*†Núcleo de Eletricidade e Automação, DEG-UFLA
Lavras- MG*

*‡Núcleo de Eletricidade e Automação, DEG-UFLA
Lavras- MG*

Emails: aabraga@posgrad.ufla.br, danton@deg.ufla.br, brunohb@deg.ufla.br

Abstract— Being a nonlinear generalization of principal component analysis, the principal curves technique is a robust tool for data analysis and classification. In pattern recognition one of the most popular algorithms to build Principal Curves is the k-segments. This algorithm presents good results and excellent applicability due to its guaranteed convergence and robustness. However, its use depends on user-defined parameters. This work presents an automatic selection technique of the quantity and length of segments of the k-segment algorithm using the TLBO (Teaching-Learning-Based Optimization) metaheuristic. A cost function is used that takes into account the length of the curve and the distance of the events to the segments where they are projected. Experimental tests made with synthetic databases are presented to demonstrate the efficiency of the proposed method in representation problems.

Keywords— Principal Curves, k-Segment, Teaching-learning-based optimization.

Resumo— Sendo uma generalização não linear de análise de componentes principais a técnica de curvas principais é um ferramenta robusta para análise e classificação de dados. Em reconhecimento de padrões um dos algoritmos mais populares para construção de Curvas Principais é o algoritmo k-segmentos. Esse algoritmo apresenta bons resultados e ótima aplicabilidade por sua convergência garantida e robustez. Contudo, sua utilização depende de parâmetros definidos pelo usuário. Este trabalho apresenta uma técnica de seleção automática da quantidade e comprimento dos segmentos do algoritmo de k-segmentos com a utilização da meta-heurística TLBO (Teaching-Learning-Based Optimization). É utilizada uma função-custo que leva em consideração o comprimento da curva e a distância dos eventos aos segmentos onde se projetam. Testes experimentais feitos com bases de dados sintéticos são apresentados para demonstrar a eficiência do método proposto em problemas de representação.

Palavras-chave— Curvas Principais, k-Segmentos, Otimização baseada em ensino-aprendizagem.

1 Introdução

O conceito de curvas principais (CP) foi apresentado por (Hastie and Stuetzle, 1989) como uma generalização não linear da técnica de análise de componentes principais. Definidas como curvas unidimensionais suaves que passam através do centro de um conjunto multidimensional gerando uma representação unidimensional e não linear dos dados, as CP podem ser vistas como uma ferramenta interessante de reconhecimento de padrões.

É importante destacar que há outras técnicas não lineares de redução de dimensão tal como a rede Self-Organizing Maps (SOM), que por sua vez tem sua capacidade de representação influenciada pela distribuição dos eventos, sendo necessário avaliar a dimensão da rede utilizada para representar certas bases de dados. A rede SOM de uma dimensão, técnica de redução de dimensionalidade não linear que mais se aproxima da técnica de CP, aloca neurônios ao longo dos dados para representá-los, enquanto o método de CP constrói curvas para tal objetivo, o que faz com que sua

capacidade de representação seja maior.

Computacionalmente, a extração das CP de um conjunto de dados se dá pela definição do primeiro componente principal do conjunto como ponto de partida para construção da curva, que é alterada sucessivamente até a convergência do algoritmo de obtenção das CP.

Devido à aplicabilidade das CP, é possível encontrar diferentes estratégias de implementação com diferentes paradigmas de programação aliados a outras técnicas de reconhecimento de padrões. Em (Wang et al., 2008) é proposto um algoritmo de aprendizado de CP baseado em redes neurais artificiais, que transforma aprendizado não supervisionado em aprendizado supervisionado com a projeção do conjunto de dados em linhas poligonais, com a utilização de algoritmos de CP já existentes. Um novo método de aprendizado de CP baseado em Análise de Componentes Principais (PCA) e *Back-propagation network* (BP) que também transforma aprendizado não supervisionado em aprendizado supervisionado é encontrado em (Wang et al., 2007). Em (Cleju et al., 2005) se propõe a utilização de uma es-

estrutura de árvores para obtenção das características principais dos dados de forma a aumentar o desempenho dos métodos baseados em CP. Em (Zhang et al., 2011) é proposta uma estratégia de partição não linear dos dados, feita em estrutura de árvore de CP que é implementada de maneira iterativa. A modelagem das curvas também pode ser realizada por meio da distribuição de probabilidade dos dados de uma CP como mistura de métodos não lineares, como feito em (Delicado and Smrekar, 2007), ou baseada em um esquema de subdivisões, de modo que as curvas são obtidas iterativamente, como proposto em (Zhang and Luo, 2005).

A robustez e a convergência garantida aliadas à sua baixa susceptibilidade a mínimos locais faz com que o algoritmo de extração de CP proposto por (Verbeek et al., 2002) seja amplamente utilizado. Fernandez (2005) utilizou o algoritmo k-segmentos para propor um método de classificação de navios por meio de dados de sonar passivo e imagens infravermelhas, assim como em (Fernandes et al., 2015) onde se propõe a utilização da técnica de CP para a determinação da classe dos navios, seguida pela classificação dos mesmos por meio de um índice de projeção. Em (Oliveira and Souza Filho, 2017) foi abordada a classificação hierárquica de navios por meio de CP com a utilização do algoritmo k-segmentos.

Ferreira et al. (2013) estudaram a utilização do método de CP para o monitoramento de qualidade de energia e em (Faier, 2006) tal método foi utilizado para identificar e classificar padrões em descargas parciais. Em (Ferreira et al., 2015) foi proposto um índice de desvio de qualidade de energia com o qual é possível identificar distúrbios na rede elétrica. Já em (Moraes and Ferreira, 2016) o algoritmo k-segmentos é utilizado na clusterização de dados apresentando bons resultados para clusters esféricos e alongados.

A extração das CP a partir do algoritmo k-segmentos é feita de maneira incremental, iniciando-se com apenas um segmento, sendo o número de segmentos aumentado progressivamente, com a reconstrução da curva a cada inserção. Para a utilização desse algoritmo, porém, é necessária a determinação de certos parâmetros, tais como: a quantidade de segmentos que compõem a curva, k , e o coeficiente que determina como é feita a ligação entre diferentes segmentos, λ . O comprimento do primeiro segmento, que tem ligação direta com a qualidade da representação dos dados é fixo, não sendo utilizado como parâmetro de entrada do algoritmo. É necessário observar que o comprimento do primeiro segmento tem influência direta no comprimento total da curva, uma vez que, os segmentos subsequentes são maiores à medida que se aumenta o comprimento do segmento inicial.

A determinação dos parâmetros depende de

fatores tais como o conhecimento prévio dos dados por parte dos usuários, o que, por si só, não é suficiente. Em (Wang and Lee, 2006) é apresentado um método de seleção automática de k e λ , analisado mais profundamente em (Lee and Wang, 2006), em que a etapa de estimação dos parâmetros é descrita em um problema de determinação do modelo estatístico, necessitando de uma grande quantidade de cálculos.

Neste trabalho é proposta uma técnica de seleção automática de k e do comprimento dos segmentos que compõem a linha poligonal com a utilização de uma meta-heurística baseada em ensino e aprendizagem em uma sala de aula, o *Teaching-learning-based optimization* (TLBO) (Rao et al., 2011). O método proposto dispensa a necessidade de se conhecer previamente os dados, sendo possível representá-los de maneira satisfatória com poucos ajustes iniciais por parte do usuário e sem a realização de cálculos complexos. Tal método também elimina a necessidade de realização de processos de busca exaustiva, de difícil execução, devido às inúmeras possibilidades de pares de parâmetros possíveis, destacando ainda, que o comprimento dos segmentos não é uma medida discreta, inviabilizando a determinação dos parâmetros ótimos de maneira manual.

O restante do trabalho está organizado da seguinte maneira. A Seção 2 é dividida em duas partes, sendo que na Subseção 2.1 é descrito o funcionamento do algoritmo de otimização utilizado e na Subseção 2.2 o método proposto é descrito; os resultados alcançados são discutidos na Seção 3. Já na Seção 4 é apresentada a conclusão dessa etapa da pesquisa assim como os próximos passos a serem tomados.

2 Fundamentação Teórica

2.1 *Teaching-learning-based optimization* (TLBO)

Teaching-learning-based optimization (TLBO) é um método de otimização baseado em populações proposto em (Rao et al., 2011). Seu funcionamento se inspira nas interações sociais existentes dentro de uma sala de aula. É dividido em duas fases, como apresentadas a seguir e na Figura 1:

- **Fase Professor:** a cada iteração identifica-se na população o indivíduo que apresenta melhor aptidão para resolver o problema em questão. A este indivíduo é atribuído a função de professor. Os outros indivíduos “aprendem” com o professor se movendo em sua direção, por meio de fatores de aprendizado calculados de maneira aleatória.
- **Fase Aluno:** nesta fase os “alunos” interagem entre si de maneira a aumentar seu “conhecimento”. O fator de interação entre os

indivíduos é calculado de maneira aleatória. Esta interação faz com que os indivíduos da população se aproximem das soluções ótimas. Verifica-se, então, se os critérios de parada foram atingidos para a realização de mais um laço iterativo ou finalização do algoritmo.

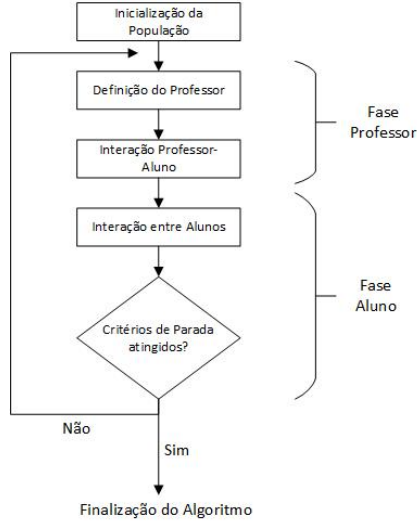


Figura 1: Fluxograma da Meta-heurística TLBO.

A escolha desse método de otimização foi feita devido à sua simplicidade computacional e bom desempenho para solução de problemas complexos, uma vez que apresenta fácil implementação e requer poucos parâmetros de ajuste pelo usuário, tais como o tamanho da população e número de iterações (Garcia et al., 2017). Estas características fazem com que esse método seja efetivo para a otimização da escolha dos parâmetros iniciais do algoritmo de k-segmentos proposta neste trabalho.

2.2 Método de seleção automática de parâmetros

A efetividade da classificação de dados pelo algoritmo de k-segmentos depende da escolha correta dos parâmetros do algoritmo, tais como o comprimento do primeiro segmento e a quantidade de segmentos que compõem a curva. É importante observar que a escolha do comprimento do primeiro segmento impacta de maneira significativa a construção da curva, uma vez que a construção dos segmentos subsequentes é feita de maneira incremental a partir dos segmentos existentes.

A modelagem do problema de otimização no TLBO foi feita com a definição dos membros da população, sendo estes constituídos de duas variáveis (k, f) , onde $f\sigma$ é o comprimento do primeiro segmento, sendo σ^2 a variância dos dados ao longo do primeiro segmento, calculada pelo algoritmo, e k o número de segmentos. Tais variáveis são possíveis combinações de parâmetros iniciais para o algoritmo de k-segmentos. Definiu-se como 50 o tamanho da população, inicializada

aleatoriamente, de maneira a garantir uma alta variedade de pares (k, f) , que por sua vez passam pelas fases “professor” e “aluno” do TLBO por 30 iterações. Definiu-se como objetivo do algoritmo de otimização a minimização da equação (Verbeek et al., 2002):

$$(1-p)n \log l + p \sum_{i=1}^k \sum_{x \in V_i} d(s_i, x)^2 / (2\sigma^2), \quad (1)$$

em que n é o número de eventos, l é o comprimento total da curva, $d(s_i, x)$ a distância de cada evento à curva, σ^2 a variância dos dados ao longo do primeiro segmento, V_i as regiões de Voronoi à qual pertencem os eventos e p a ponderação a ser escolhida de acordo com o problema.

Segundo Verbeek et al. (2002) há uma variedade de critérios de parada para o algoritmo k-segmentos e, na prática, o mínimo global da Equação (1), com $p = 0,5$, é atingido quando há boa representatividade dos dados na curva gerada. Tal afirmação levou à escolha da Equação (1) como primeira função-custo a ser utilizada no presente trabalho. Apesar de Verbeek et al. (2002) ter utilizado $p = 0,5$ para todos os problemas, observou-se, durante a realização deste trabalho, a necessidade de ponderar o peso dos dois termos dessa equação de modo a representar melhor os dados, como é discutido na Seção 3.

3 Resultados e discussão

Os testes com o algoritmo de otimização foram realizados em duas etapas distintas, de modo a verificar sua potencialidade para a resolução do problema em questão. Na primeira etapa os indivíduos da população apresentam apenas uma variável, o número de segmentos k , sendo o comprimento do primeiro segmento fixado em $3\sigma/2$, que segundo (Verbeek et al., 2002) apresentou os melhores resultados experimentais em diversos problemas. Na segunda etapa, os indivíduos da população apresentam duas variáveis, o número de segmentos k e o parâmetro f , tal que $f\sigma$ é o comprimento do primeiro segmento.

3.1 Base de dados

De modo a verificar a capacidade de representação do método proposto foram utilizados dois diferentes padrões de dados. Tais conjuntos de dados artificiais formam uma espiral e um rabisco, que intercepta a si mesmo, ambos com ruído gaussiano.

Os mesmos conjuntos de dados foram utilizados em (Verbeek et al., 2002) e (Wang and Lee, 2006) para mostrar a efetividade dos respectivos algoritmos.

3.2 Comprimento do segmento inicial fixo

Com o valor de comprimento do segmento inicial fixado em $3\sigma/2$ verificou-se a capacidade do algoritmo de otimização de encontrar o número ideal de segmentos que compõem a CP para o conjunto de dados. Com a utilização da Equação (1), como função custo do TLBO, com ponderação $p = 0,5$, como proposto em (Verbeek et al., 2002), obtém-se os resultados mostrados nas Figuras 2(a) e 2(c). É possível observar que as curvas geradas não representam de maneira eficaz o contorno do conjunto de dados, sobretudo os dados visualizados em 2(c) onde se percebe de maneira clara a necessidade de se aumentar o número de segmentos.

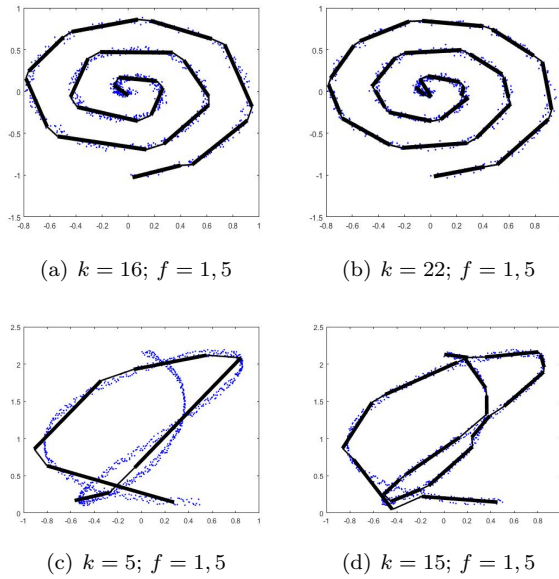


Figura 2: Curvas geradas com o valor do comprimento do primeiro segmento fixado em $3\sigma/2$, sendo (a) e (c) construídas a partir da minimização da Equação (1), com $p = 0,5$ e (b) e (d) construídas a partir da minimização da Equação (2), onde $p = 0,95$.

De modo a se obter melhor representatividade dos dados é necessária uma análise mais detalhada da função custo do algoritmo de otimização. O primeiro termo dessa função é relativo ao comprimento total da curva, enquanto o segundo é proporcional ao somatório quadrático das distâncias de cada evento aos segmentos da curva. Dessa maneira, a função objetivo utilizada pondera de forma igual o aumento do comprimento da curva e o distanciamento dos pontos aos segmentos. Essa ponderação mostra-se ineficaz para representação de dados com alto número de eventos e esse tipo de conjunto de dados geralmente requer um alto número de segmentos para que seja representado de maneira eficaz pela CP. A partir dessa observação verificou-se a necessidade de utilizar uma ponderação diferente da utilizada por (Verbeek et al., 2002) para que o impacto do aumento de

número de segmentos na ponderação da função objetivo seja menor.

Com a realização de testes heurísticos observou-se que o uso de valor de ponderação $p = 0,95$ retorna bons resultados, fazendo com que as curvas representem de maneira melhor o contorno dos dados, resultando na seguinte função objetivo:

$$0,05 n \log l + 0,95 \sum_{i=1}^k \sum_{x \in V_i} d(s_i, x)^2 / (2\sigma^2) \quad (2)$$

As Figuras 2(b) e 2(d) apresentam os resultados obtidos com a utilização da função objetivo apresentada na Equação (2). O aumento do número de segmentos em ambos os conjuntos de dados faz com que a curva siga de maneira mais suave o contorno dos mesmos, evitando o subdimensionamento visualizado quando se utiliza a função objetivo proposta em (Verbeek et al., 2002). É importante destacar que os testes heurísticos realizados para a determinação do valor da ponderação p tem como objetivo a definição da melhor função custo a ser utilizada pelo método de seleção automática de parâmetros iniciais proposto.

3.3 Comprimento do segmento inicial como parâmetro otimizado

Com o intuito de verificar a viabilidade da função objetivo proposta por (Verbeek et al., 2002) quando o valor de f é um dos parâmetros de entrada do algoritmo a serem otimizados, utilizou-se a Equação (1), com valor de $p = 0,5$.

Observa-se nas Figuras 3(a) e 3(c) os resultados obtidos. Para o conjunto de dados de 3(a) os resultados encontrados são satisfatórios, apesar de ser perceptível uma baixa suavidade de contorno da curva. Já em 3(c) verifica-se o subdimensionamento da curva, sendo a representação pior do que a obtida na Figura 2(c). Para ambos os casos a alteração do valor de p da função objetivo é capaz de solucionar tal deficiência. É importante destacar o fato que para ambos os conjuntos de dados o comprimento do segmento que minimiza a função custo é pouco maior que σ , divergindo do valor $3\sigma/2$, recomendado por Verbeek et al. (2002).

Com a utilização da Equação (2) os problemas citados de representação são minimizados. Para o conjunto de dados em espiral a suavidade da curva gerada aumentou consideravelmente com o aumento do número de segmentos, gerado pela utilização desta função objetivo, como visto na Figura 3(b). Observa-se que o valor de f encontrado é menor que $3/2$.

O comprimento do segmento tem conexão direta com a quantidade de segmentos da curva. Dessa forma, a medida em que a quantidade de segmentos aumenta o comprimento ideal deles

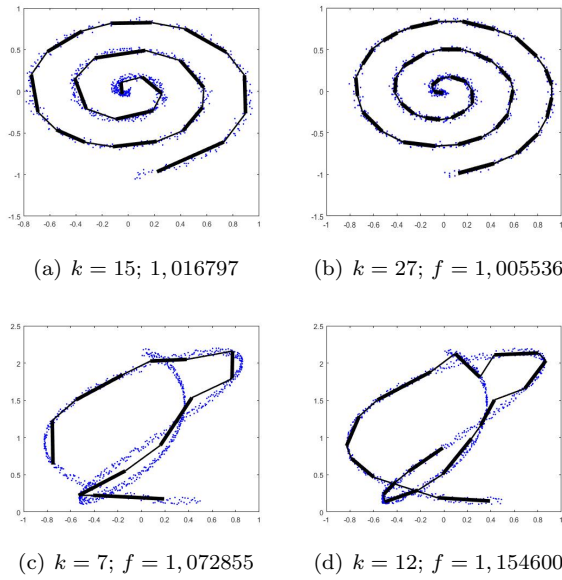


Figura 3: Curvas geradas com o valor do comprimento do primeiro segmento variável, sendo (a) e (c) construídas a partir da minimização da Equação (1), com $p = 0,5$ e (b) e (d) construídas a partir da minimização da Equação (2).

tende a diminuir, de modo a garantir a minimização da distância dos eventos à curva. A manutenção do comprimento em $3\sigma/2$ mostra-se incoerente à medida que se aumenta o valor de k , uma vez que a suavidade da CP gerada é proporcional à ponderação entre comprimento e quantidade de segmentos, sendo penalizada quando não há equilíbrio entre tais valores.

É importante notar que o número de segmentos gerados é proporcional ao comprimento total da curva. Por esse motivo a variação do primeiro termo da função objetivo utilizada impacta o número de segmentos que compõem a CP.

Para o segundo conjunto de dados, visualizados na Figura 3(d), a utilização da função objetivo com $p = 0,95$ fez com que os dados fossem representados de maneira mais eficiente, uma vez que ocorre um aumento considerável do número de segmentos. Por outro lado, o aumento de segmentos não é suficiente para que a representação seja, de fato, satisfatória.

Seguindo a premissa que, para a função objetivo, o aumento da ponderação dada à distância dos eventos à curva gera melhores resultados, realizou-se testes heurísticos de modo a definir uma equação cujos parâmetros que a minimizam sejam capazes de representar satisfatoriamente o contorno de ambos os conjuntos de dados. Tal resultado foi obtido com a utilização da seguinte equação ($p = 0,99$):

$$0,01 n \log l + 0,99 \sum_{i=1}^k \sum_{x \in V_i} d(s_i, x)^2 / (2\sigma^2). \quad (3)$$

A curva de convergência do método proposto para os dados em espiral é apresentada na Figura 4. É possível observar que no caso apresentado a função-custo atinge seu mínimo antes das 30 iterações utilizadas, o que mostra uma rápida convergência do algoritmo para encontrar uma solução satisfatória. A qualidade dos resultados encontrados com a utilização da Equação (3) pode ser visualizada nas Figuras 5(a) e 5(b).

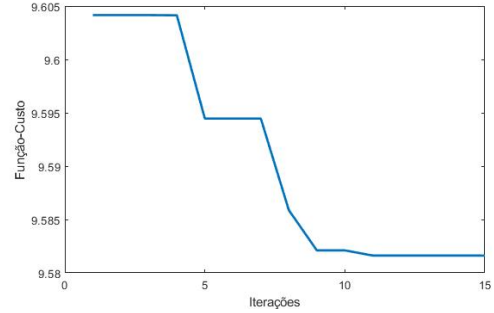


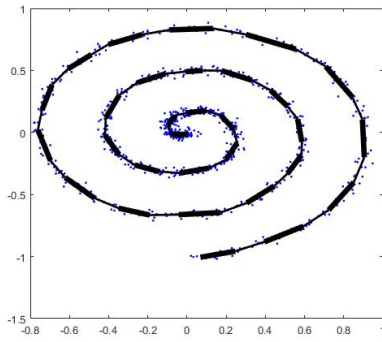
Figura 4: Convergência do algoritmo TLBO nas primeiras 15 iterações.

Ao se observar a Equação (3) é possível notar a discrepância entre os pesos dados a cada um dos termos, gerando questionamentos a respeito da importância do primeiro deles, referente ao comprimento da curva. A não utilização desse termo, por outro lado, pode acarretar no sobreajuste, *overfitting*, do modelo que representa o conjunto de dados. Isso ocorre, uma vez que a não penalização do aumento excessivo do número de segmentos pode gerar uma CP em que cada par de eventos gere um segmento. Nesse caso a distância de cada ponto à curva seria nulo fazendo com que o mínimo global da função objetivo também o fosse.

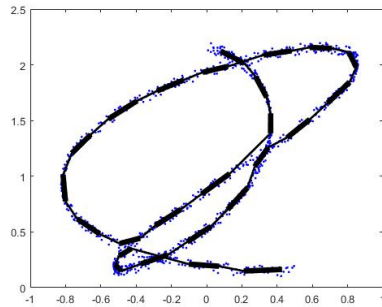
Por esse motivo é importante a manutenção dos dois termos da equação, mesmo que a ponderação de ambos seja desigual e, mesmo com diferentes ponderações desta, caso se mostre necessário em alguma aplicação específica.

Em uma observação mais detalhada da Figura 5(b) verifica-se a ligação equivocada de alguns segmentos nas proximidades dos pontos onde há cruzamento dos dados. A correta construção das linhas que ligam os segmentos, segundo (Wang and Lee, 2006), é determinada pelo parâmetro λ e o efeito de sua variação é visível, principalmente, em conjuntos de dados onde há interseções, como o da figura. Para o método aqui proposto o valor de λ foi fixado em 1 para todos os conjuntos de dados. Na Figura 6 é mostrado que com a variação do valor de λ é possível a obtenção de um contorno mais adequado, para os valores de f e k obtidos pelo método de seleção automática proposto.

Mostra-se necessário observar que a variação de λ não influencia de maneira significativa a variação dos valores da função objetivo. Assim, para



(a) $k = 30$; $f = 1,025893$



(b) $k = 25$; $f = 1,006802$

Figura 5: Curvas geradas a partir da minimização da Equação (3).

que λ seja considerada uma variável a ser escolhida automaticamente pelo algoritmo, é preciso redefinir a função objetivo.

De maneira geral, o algoritmo proposto, com a utilização da Equação (3) apresenta resultados satisfatórios para a escolha dos parâmetros iniciais, (f, k) , de grande relevância para a correta construção de uma curva principal pelo algoritmo de k -segmentos.

4 Conclusões

O método de seleção automática de parâmetros iniciais f e k do algoritmo de k -segmentos com a utilização de TLBO proposto por este trabalho apresentou resultados satisfatórios, como demonstrado pelos exemplos utilizados. É possível observar que a boa representação dos dados passa pela escolha adequada da função objetivo.

De modo a aumentar a potencialidade do algoritmo para que o mesmo seja capaz de determinar automaticamente a melhor ligação entre os segmentos, é preciso definir uma função objetivo mais ampla. A definição de tal função juntamente com a utilização de algoritmos evolucionários multi-objetivo, capazes de otimizar a ponderação dos pesos relativos à distância dos eventos e comprimento da curva, são os próximos passos desta pesquisa, juntamente com avaliações da técnica proposta em bases de dados reais.

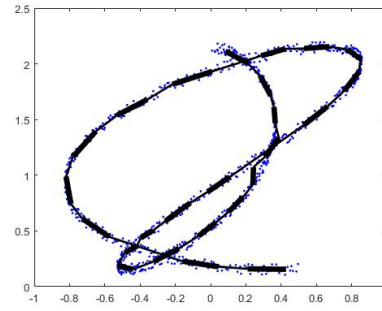


Figura 6: Curva gerada a partir da minimização da Equação (3) com $k = 25$; $f = 1,006802$ e $\lambda = 5$.

Agradecimentos

Nossos agradecimentos à Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Referências

- Cleju, I., Franti, P. and Wu, X. (2005). Clustering by principal curve with tree structure, *International Symposium on Signals, Circuits and Systems, 2005. ISSCS 2005.*, Vol. 2, pp. 617–620 Vol. 2.
- Delicado, P. and Smrekar, M. (2007). Mixture of nonlinear models: a bayesian fit for principal curves, *2007 International Joint Conference on Neural Networks*, pp. 195–200.
- Faier, J. M. (2006). *Curvas principais na identificação de descargas parciais em equipamentos de potência*, Master's thesis, Universidade Federal do Rio de Janeiro.
- Fernandes, T. S. P., de Oliveira e Souza Filho, J. B. and dos Santos Alves, E. (2015). Identificação de navios em sinais de sonar passivo por curvas principais, in C. J. A. Bastos Filho, A. R. Pozo and H. S. Lopes (eds), *Anais do 12 Congresso Brasileiro de Inteligência Computacional*, ABRICOM, Curitiba, PR, pp. 1–6.
- Fernandez, H. L. (2005). *Classificação de navios baseada em curvas principais*, Master's thesis, Universidade Federal do Rio de Janeiro.
- Ferreira, D. D., de Seixas, J. M., Cerqueira, A. S. and Duque, C. A. (2013). Exploiting principal curves for power quality monitoring, *Electric Power Systems Research* **100**: 1 – 6.
- Ferreira, D. D., de Seixas, J. M., Cerqueira, A. S., Duque, C. A., Bollen, M. H. and Ribeiro,

- P. F. (2015). A new power quality deviation index based on principal curves, *Electric Power Systems Research* **125**: 8 – 14.
- Garcia, C. M., Catalano, M. D., Soares, E. A. and Barbosa, B. H. (2017). Teaching-learning-based optimization no treinamento de redes neurais artificiais para problemas de classificação.
- Hastie, T. and Stuetzle, W. (1989). Principal curves, *Journal of the American Statistical Association* **84**(406): 502–516.
- Lee, T. C. M. and Wang, H. (2006). On a k-segments algorithm for computing principal curves, *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 183–187.
- Moraes, E. C. C. and Ferreira, D. D. (2016). A principal curve-based method for data clustering, *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 3966–3971.
- Oliveira, T. C. C. and Souza Filho, J. B. d. O. (2017). Otimização de classificadores de sinais de sonar passivo baseado em curvas principais hierárquicas, *Anais do XIII Simpósio Brasileiro de Automação Inteligente*, Porto Alegre, RS, pp. 569–574.
- Rao, R., Savsani, V. and Vakharia, D. (2011). Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems, *Computer-Aided Design* **43**(3): 303 – 315.
- Verbeek, J., Vlassis, N. and Krose, B. (2002). A k-segments algorithm for finding principal curves, *Pattern Recognition Letters* **23**(8): 1009 – 1017.
- Wang, H. and Lee, T. C. (2006). Automatic parameter selection for a k-segments algorithm for computing principal curves, *Pattern Recognition Letters* **27**(10): 1142 – 1150.
- Wang, Y. H., Guo, Y., Fu, Y. C. and Shen, Z. Y. (2007). An algorithm for learning principal curves with principal component analysis and back-propagation network, *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pp. 447–453.
- Wang, Y., Sun, Y., Liu, C. and Guo, Y. (2008). An algorithm for learning principal curves with explicit formulation based on neural network, *2008 Congress on Image and Signal Processing*, Vol. 1, pp. 399–404.
- Zhang, J., Wang, X., Kruger, U. and Wang, F. Y. (2011). Principal curve algorithms for partitioning high-dimensional data spaces, *IEEE Transactions on Neural Networks* **22**(3): 367–380.
- Zhang, L. and Luo, Z. (2005). Adaptive smooth principal curves design, *International Conference on Computer Graphics, Imaging and Visualization (CGIV'05)*, pp. 241–244.