EXPERIMENTOS COM UMAP E ELECTRE-TREE PARA ANÁLISE DE PAÍSES CONSIDERANDO INDICADORES SOCIOECONÔMICOS

Caio Mocny Coutinho, Rodrigo Celso de Lima Porto, Diogo Ferreira de Lima Silva, Valdecy Pereira, Marcos Costa Roboredo
Departamento de Engenharia de Produção (TEP), Universidade Federal Fluminense
caiomocny@id.uff.br, rodrigo_celso@id.uff.br, diogofls@id.uff.br, valdecypereira@id.uff.br, mcroboredo@id.uff.br

RESUMO

O recente método de Apoio Multicritério à Decisão ELECTRE-TREE foi proposto para problemas de classificação ordinal, unindo aspectos do tradicional ELECTRE-TRI-B com uma abordagem de aprendizado de máquina para uma elicitação de parâmetros de maneira indireta. Neste trabalho, usamos o ELECTRE-TREE em conjunto com uma técnica de redução de dimensionalidade (UMAP) para classificar países de acordo com suas performances em indicadores socioeconômicos. Dentre os experimentos realizados, considerou-se diferentes números de modelos para a construção do comitê de classificação dada uma projeção do UMAP. Em seguida, um experimento testou um número fixo de modelos (100), porém, com variações nas projeções do UMAP antes do aprendizado de parâmetros. Resultados consistentes foram observados com uma acurácia preditiva em torno de 80%. Detalhes da obtenção e tratamento de dados são discutidos ao longo do texto, assim como as principais contribuições dessa pesquisa.

PALAVRAS CHAVE. ELECTRE. Classificação Ordinal. UMAP.

Apoio à Decisão Multicritério, Inteligência Computacional

ABSTRACT

The recent ELECTRE-TREE method was proposed for ordinal classification problems, combining aspects of the traditional ELECTRE-TRI-B with a machine learning approach for an indirect parameter elicitation. In this work, we use the ELECTRE-TREE in conjunction with a dimensionality reduction technique (UMAP) to classify countries according to their performance in socioeconomic indicators. Among the experiments carried out, different numbers of models were considered for the construction of the classification committee given a UMAP projection. Then, an experiment tested a fixed number of models (100), however, with variations in the UMAP projections before the parameter learning step. Consistent results were observed with a predictive accuracy of around 80%. Details regarding data gathering and processing are discussed throughout the text, as well as the main contributions of this research.

KEYWORDS. ELECTRE. Sorting. UMAP.

Multicriteria Decision Support, Computational Intelligence

https://proceedings.science/p/157586?lang=pt-br

1. Introdução

Ao longo das últimas décadas, vários estudos analisaram as características de países/ sociedades que alcançaram êxito em termos de desenvolvimento econômico assim como daqueles que não conseguiram se desenvolver. Por exemplo, estudos recentes analisaram o problema sob o ponto de vista da complexidade econômica dos países [Hartmann et al., 2021; Hidalgo, 2021]. Segundo Hidalgo [2021], este índice mede a intensidade do conhecimento de um país considerando tanto a diversidade quanto a ubiquidade dos produtos que exporta, considerando não apenas o nível agregado de produção, como também quantos e quais tipos de produtos os países são capazes de produzir e exportar. Pesquisadores defendem que países com uma indústria capaz de desenvolver e exportar bens mais complexos estariam mais aptos a alcançar o desenvolvimento econômico.

O presente estudo explora o índice de complexidade econômica para a formulação de exemplos de classificação a serem utilizados para aprendizado de parâmetros em um método recente de classificação multicritério, o ELECTRE-TREE [Montenegro de Barros et al., 2021]. A área de Apoio Multicritério à Decisão (AMD) trata de problemas de decisão complexos caracterizados por situações onde há vários fatores (critérios, atributos) de influência sobre uma determinada decisão, sendo estes fatores frequentemente conflitantes [Basílio et al., 2022; Gomes e Gomes, 2019; Greco et al., 2016; de Almeida et al., 2015; Leoneti e Gomes, 2022]. Existem vários problemas reais e de diferentes áreas que podem ser tratados por métodos AMD, como apresentado nas revisões a seguir [Costa et al., 2022; de Almeida-Filho et al., 2021; Zanghelini et al., 2018; Chowdhury e Paul, 2020]. Cinelli et al. [2020] apresenta uma revisão compreensiva da taxonomia usada na área de AMD. Dentre os principais tipos de problemas está a classificação ordinal. Nestes problemas, deseja-se alocar alternativas em classes pertencentes a um conjunto pré-estabelecido e ordenado em termos de preferências.

Neste cenário, diversos métodos de classificação multicritério foram propostos ao longo das últimas décadas. Dentre eles, o método ELECTRE-TRI (também conhecido como ELECTRE-TRI-B), inicialmente proposto por [Roy e Bouyssou, 1993], destaca-se como uma abordagem não compensatória, pertencente a família de métodos ELECTRE, que realiza a alocação de alternativas explorando relações de sobreclassificação com perfis de fronteira. Vários procedimentos foram criados ao longo dos anos como variantes do método ELECTRE-TRI [Bouyssou et al., 2022; Montenegro de Barros et al., 2021; Kadziński e Martyn, 2021; Madhooshiarzanagh e Abi-Zeid, 2021; Fernández et al., 2020; Ramezanian, 2019]. Em algumas dessas variantes, destaca-se que os parâmetros são aprendidos por meio de exemplos de referência. Este é o caso do recente ELECTRE-TREE, que alia as características do método de sobreclassificação ELECTRE-TRI-B com técnicas de aprendizado de máquina (florestas randômicas) e um algoritmo genético para inferência indireta de parâmetros.

Neste estudo, um conjunto de 124 países é analisado em termos de um conjunto de indicadores relacionados à aptidão para alcançar um desenvolvimento econômico. Cada país é alocado pelo método para uma dentre três classes ordenadas. Em conjunto com o ELECTRE-TREE, destaca-se a utilização da técnica *Uniform Manifold Approximation and Projection for Dimension Reduction* (UMAP) para reduzir de dimensionalidade da matriz de decisão e facilitar o problema de inferência de parâmetros. Além disso, o estudo utiliza uma técnica de imputação de dados com base em vizinhos mais próximos das alternativas (KNN Imputer). Uma alta performance preditiva foi alcançada, porém, ao custo de uma perda de interpretabilidade do modelo de apoio à decisão.

O restante do artigo está organizado conforme a seguir. A segunda seção apresenta uma fundamentação teórica e prepara o leitor para o entendimento da aplicação, através da explanação dos passos do método ELECTRE-TREE. A terceira seção apresenta detalhes do experimento, in-

cluindo etapas de tratamento dos dados e análise dos resultados. Por fim, a Seção 4 apresenta conclusões para o trabalho e perspectivas para trabalhos futuros.

2. Fundamentação Teórica e Descrição do Método

2.1. Notação

Seja $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$ um conjunto de m alternativas analisadas de acordo com seus respectivos desempenhos em um conjunto de n critérios $G = \{g_1, g_2, ..., g_n\}$. Cada alternativa pode ser representada por um vetor $\mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]$, onde $x_{i,j}$ representa o desempenho da alternativa $\mathbf{x}_i$ no critério g_j. Desse modo, o conjunto de alternativas X, com seus respectivos desempenhos, também pode ser representado por uma matriz de decisão $\mathbf{X} = [x_{ij}]_{m \times n}$.

O presente estudo trata de um problema de classificação multicritério. Nesse cenário, cada alternativa deve ser alocada em alguma dentre três classes ordenadas e pré-estabelecidas. O conjunto de classes que pode ser representado por $C = \{C_1, ..., C_c\}$, onde $C_1 \succ C_2 \succ ... \succ C_c$. Com intuito de caracterizar a alocação das alternativas em classes por meio da relação de sobreclassificação, perfis de fronteira são definidos. O perfil $\mathbf{b}_k = [b_{k,1}, ..., b_{k,n}]$ define a fronteira entre duas classes consecutivas: C_k e C_{k+1}.

2.2. UMAP

O método UMAP (*Uniform Manifold Approximation and Projection*), publicado por McInnes et al. [2018], consiste em um método de redução da dimensionalidade dos dados, preferencialmente para duas dimensões, com o intuito de poder representá-los em um gráfico e facilitar a sua análise. Neste trabalho, a técnica é utilizada para reduzir a matriz de decisão, que recebe desempenhos dos países em relação aos indicadores socioeconômicos. Nesta técnica, amostras semelhantes tendem a se agrupar no resultado final, permitindo identificar similaridades e *outliers* no conjunto de dados original.

Considerando um conjunto de dados $\mathbf{X}_D = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$, onde D é a dimensão deste conjunto, o UMAP calcula as similaridade s_{ij} que cada dado $\mathbf{x}_i$ possui em relação com cada um dos outros dados $\mathbf{x}_j$, com o intuito de aglomerar os dados que possuírem mais similaridade entre si depois de reduzir a sua dimensionalidade. O algoritmo do UMAP é apresentado no Algoritmo 1 e mais detalhes podem ser encontrados em [McInnes et al., 2018].

2.3. ELECTRE-TREE

O método ELECTRE-TREE, proposto recentemente por Montenegro de Barros et al. [2021], consiste em uma variante do método clássico ELECTRE-TRI-B para a elicitação indireta de parâmetros. O ELECTRE-TREE combina características de Florestas Randômicas, algoritmo de Aprendizado de Máquina de alto desempenho, com o uso um algoritmo genético para inferência dos parâmetros respeitando a estrutura de classificação do método ELECTRE-TRI-B.

Dentre as principais características dos procedimentos da família ELECTRE, destacam-se a natureza não-compensatória, o uso de comparações paritárias, a noção de pesos como grau de importância relativa e o uso da relação de sobreclassificação S [Gomes e Gomes, 2019; Greco et al., 2016]. O método ELECTRE-TRI-B destaca-se como uma abordagem para a problemática de classificação. Neste caso, perfis de fronteira são utilizados para a caracterização das decisões de alocação de alternativas nas classes ordenadas. No caso do ELECTRE-TRI-B, as preferências do decisor são inseridas no modelo por meio de um conjunto de parâmetros:

- Perfis de fronteira ($\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{c-1}$): o perfil de fronteira $\mathbf{b}_k$ caracteriza o limite entre duas classes e é construído como um vetor de desempenhos referentes aos n critérios do modelo.
- Vetor de pesos ($\mathbf{w} = [w_1, ..., w_n]$): reflete o grau de importância relativa dos critérios. A entrada w_j indica o peso do critério g_j.

Algoritmo 1 UMAP, Adaptado de McInnes et al. [2018]

Entrada: Conjunto de dados $\mathbf{X}_D$ de tamanho m; Dimensão $d < D$ a ser obtida; número n de dados que fazem parte da vizinhança de $\mathbf{x}_i$ incluindo ele próprio, que por padrão é $n = 15$.

Saída: Conjunto de dados $\mathbf{X}_d$ dimensionalmente reduzida.

Para $\mathbf{x}_i \in \mathbf{X}_D$ **faça**
  Para $\mathbf{x}_{j\neq i} \in \mathbf{X}_D \setminus \{\mathbf{x}_i\}$ **faça**
    Calcula a norma euclidiana entre $\mathbf{x}_i$ e $\mathbf{x}_j$;
  Fim Para
Fim Para
Para $\mathbf{x}_i \in \mathbf{X}_D$ **faça**
  $\mathbf{x}_i \leftarrow \vec{0}$
  Para $\mathbf{x}_{j\neq i} \in \mathbf{X}_D \setminus \{\mathbf{x}_i\}$ **faça**
    É atribuído valores às dimensões reduzidas de $\mathbf{x}_j$ de modo que $||\mathbf{x}_i - \mathbf{x}_j||_d = ||\mathbf{x}_i - \mathbf{x}_j||_D$;
    É calculado a similaridade s_{ij} entre $\mathbf{x}_i$ e $\mathbf{x}_j$ de modo que $\sum_{j=1}^{m} s_{ij} = \log_2 n$;
  Fim Para
Fim Para
/ Como não existe garantia de que $s_{ij} = s_{ji}$, então o seguinte comando é necessário para simetrizar as similaridades. */*
Para $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{X}_d$ **faça**
  $s_{ij} \leftarrow s_{ij} + s_{ji} - s_{ij} \times s_{ji}$;
  $s_{ji} \leftarrow s_{ij}$;
Fim Para
Os dados de $\mathbf{X}_d$ são inicializados utilizando Incorporação Espectral (*Spectral Embedding*);
Para $\mathbf{x}_i \in \mathbf{X}_d$ **faça**
  Seleciona aleatoriamente um ponto $\mathbf{x}_j$ que possua alta similaridade com $\mathbf{x}_i$;
  Seleciona aleatoriamente um $\mathbf{x}_k$ que possua baixa similaridade com $\mathbf{x}_i$;
  Os valores de s_{ij} e s_{ik} passam a serem descritas segundo a curva de *t-Student*, cuja média são os valores dimensionalmente reduzidas de $\mathbf{x}_j$ e $\mathbf{x}_k$, respectivamente;
  Os valores de $\mathbf{x}_i$ são reajustados de forma que $\mathbf{x}_i$ se aproxime mais de $\mathbf{x}_j$ do que de $\mathbf{x}_k$, buscando assim maximizar s_{ij} ao mesmo tempo que minimiza s_{ik};
Fim Para

Retorne $\mathbf{X}_d$

- Vetor de limiar de indiferença ($\mathbf{q} = [q_1, ..., q_n]$): para cada critério g_j, o limiar q_j indica valor da diferença de desempenho entre duas alternativas, $x_{ij} - b_{ij}$, para o qual um decisor deixaria de ser indiferente entre elas e passaria a ter uma preferência fraca por uma das alternativas neste critério.

- Limiar de preferência ($\mathbf{p} = [p_1, ..., p_n]$): para cada critério g_j, o limiar p_j indica valor da diferença de desempenho entre duas alternativas, $x_{ij} - b_{ij}$, para o qual um decisor deixaria de ter uma preferência fraca e passaria a apresentar uma preferência forte por uma das alternativas neste critério.

- Limiar de veto ($\mathbf{v} = [v_1, ..., v_n]$): a relação de sobreclassificação de uma alternativa $\mathbf{x}_i$ em relação a uma outra $\mathbf{x}_j$ pode ser vetada ou penalizada caso, em um determinado critério, a

alternativa $\mathbf{x}_j$ possua um desempenho muito superior ao de $\mathbf{x}_i$. Essa característica é modelada por meio dos limiares de veto.

A determinação dessa grande quantidade de parâmetros pode ser difícil para o decisor. Por exemplo, o decisor pode não disponibilizar o tempo necessário, não se sentir confortável ou mesmo não estar apto para essa determinação de parâmetros. Esse problema vem sendo tratado na literatura pela elaboração de métodos alternativos que elicitam parâmetros através de exemplos de decisão, em uma abordagem de decisão multicritério dirigida por dados [Doumpos et al., 2022]. O método ELECTRE-TREE, apresentado no presente estudo, se encaixa nessa abordagem. Os passos do método ELECTRE-TREE são apresentados brevemente a seguir, mais detalhes podem ser encontrados no trabalho de Montenegro de Barros et al. [2021].

Algoritmo 2 ELECTREE-TREE, Adaptado de Montenegro de Barros et al. [2021]

Entrada: Definição dos parâmetros ELECTRE Tri-B que precisam ser inferidos; Tamanho da amostra S; Número de modelos N; Exemplos de Decisão ou *Clusters* Ordenados; e os parâmetros do Algoritmo Genético (GA). Os critérios precisam ser do tipo maximização.
Para $i = 0$ até N **faça**
Amostrar com reposição S observações da população.
Amostrar com reposição k critérios, onde $k \geq 2$.
Rodar modelo ELECTRE-TRI-B com as alternativas e critérios amostrados.
Se Exemplos de decisão são dados **então**
Usar exemplos como referências.
Se não
Usar clusters ordenados como referências.
Fim Se
Gerar parâmetros desconhecidos do ELECTRE-TRI-B usando o algoritmo genético. A função perda é dada pela acurácia entre o ELECTRE-Tri-B e as referências.
Incluir o modelo com os parâmetros convergentes na lista de soluções.
Fim Para
Para cada alternativa j **faça**
Para $i = 0$ até n **faça** Classificar as alternativas usando o modelo i. Contar os votos das classes para a alternativa j.
Fim Para
Fim Para

Retorne votos e parâmetros elicitados.

3. Estudo de Caso

O estudo de caso aqui apresentado ilustra a aplicação de uma metodologia multicritério com o objetivo de classificar um conjunto de 124 países em termos de indicadores socioeconômicos. Neste caso, os países representam o conjunto de alternativas do modelo multicritério, os indicadores compõem os critérios, e cada um dos países será classificado em uma dentre três classes, cujas fronteiras são delimitados por perfis gerados pelo método ELECTRE-TREE.

3.1. Coleta e Tratamento de Dados

A primeira etapa desta aplicação consistiu na busca por indicadores socio-econômicos para compor o conjunto de critérios do modelo. Três fontes foram utilizadas com esse intuito: o

Atlas de Complexidade Econômica [The Growth Lab at Harvard University, 2022], o Índice de Liberdade Econômica [Heritage Foundation, 2022] e o banco de dados do Banco Mundial [The World Bank, 2022].

No total, 38 indicadores foram coletados dessas três bases para utilização neste trabalho, conforme indicam as duas primeiras colunas da Tabela 1.

Observou-se uma grande quantidade de dados faltantes nos anos mais recentes, principalmente para os países menos desenvolvidos. Isso reduziria drasticamente o número de países a serem considerados no trabalho. Dessa forma, visando suavizar a ausência de dados em certos critérios e de possíveis efeitos econômicos passageiros, foi decidido utilizar indicadores médios na matriz de decisão. Neste cenário, para os indicadores provenientes do Índice de Liberdade Econômica, cujos dados estão mais disponíveis para os países em anos recentes, esta pesquisa utiliza a média entre os anos de 2018 e 2022. Já para os indicadores provenientes do banco mundial, com mais dados faltantes para um número considerável de países, optou-se por trabalhar com médias de 10 anos, entre 2011 e 2020. Em relação ao indicador de Complexidade Econômica, este é utilizado na pesquisa como atributo alvo para a construção dos exemplos de decisão, atribuições de alternativas dentre as três classes ordenadas. Neste caso, o ano de 2019 (mais recente disponível) foi utilizado.

Mesmo com a utilização valores médios os dados presentes dentro de intervalos de anos recentes, alguns países continuaram com dados faltantes para a alguns critérios. Dois procedimentos foram utilizados. Primeiramente, arbitrou-se um número, 3 dados faltantes, acima do qual os objetos do conjunto de dados seriam desconsiderados. Para as alternativas que passaram por esse filtro, decidiu-se aplicar um procedimento de imputação de dados com base nos dois vizinhos mais próximos (duas alternativas cujas distâncias Euclideanas considerando os demais critérios são as menores dentro do conjunto). Este procedimento foi aplicado com o método KNN Imputer presente no pacote Scikit-Learn no Python.

Ainda durante a etapa de tratamento dos dados, foram analisadas as correlações entre os critérios dentro do conjunto de dados. Altas correlações são consideradas prejudiciais para caráter multidimensional do estudo. Procura-se um conjunto de critérios exaustivo mas não repetitivo. Alguns dos achados são listados a seguir. Além disso, A Figura 1 apresenta uma parte da matriz de correlações obtida na análise, evidenciando a alta correlação entre os indicadores do Índice de Liberdade Econômica.

- Uma correlação clara foi visto nos indicadores obtidos da base do Economic Freedom Index, por isso, todos os indicadores obtidos desta base, exceto o Overall Score, foram retirados;
- Como o Crescimento Anual do PIB era altamente correlacionado ao Crescimento do PIB per Capita, escolheu-se manter apenas o Crescimento do PIB per Capita;
- Como havia 4 colunas relacionadas a Despesas Públicas com Educação (Total, Primária, Secundária e Terciária) que eram altamente correlacionadas entre si, decidiu-se manter apenas a primeira delas;
- Como o Índice GINI possuía uma correlação altíssima com o índice de Distribuição entre os 20% mais Ricos, que também estava correlacionada com a Distribuição entre os 20% mais Pobres, então optou-se por manter apenas o Índice GINI;
- Como Despesas Públicas estava altamente correlacionado com Receitas Públicas, decidiu-se criar uma coluna única com Receitas/Despesas.

Tabela 1: Indicadores Coletados

Indicador	Fonte
Índice de Complexidade Econômica	Atlas de Complexidade Econômica
Score Global de Liberdade Econômica	Índice de Liberdade Econômica
Direito de propriedade	Índice de Liberdade Econômica
Integridade Pública	Índice de Liberdade Econômica
Eficácia Judicial	Índice de Liberdade Econômica
Carga Tributária	Índice de Liberdade Econômica
Gastos Públicos	Índice de Liberdade Econômica
Saúde Fiscal	Índice de Liberdade Econômica
Liberdade de negócios	Índice de Liberdade Econômica
Liberdade de Trabalho	Índice de Liberdade Econômica
Liberdade monetária	Índice de Liberdade Econômica
Liberdade Comercial	Índice de Liberdade Econômica
Liberdade de investimento	Índice de Liberdade Econômica
Liberdade financeira	Índice de Liberdade Econômica
PIB (Bilhões, PPP)	Índice de Liberdade Econômica
1/Desemprego	Índice de Liberdade Econômica
Taxa de Imposto de Renda (%)	Índice de Liberdade Econômica
1/Inflação	Banco Mundial
Crescimento anual do PIB (%)	Banco Mundial
Crescimento do PIB per Capita (%)	Banco Mundial
Reservas Totais	Banco Mundial
Receitas do Governo (% do PIB)	Banco Mundial
Despesas do Governo (% do PIB)	Banco Mundial
População Ativa (%) (15 - 65 anos)	Banco Mundial
Despesas do Governo em Educação (% do PIB)	Banco Mundial
Despesas do Governo em Educação Primária (% do PIB)	Banco Mundial
Despesas do Governo em Educação Secundária (% do PIB)	Banco Mundial
Despesas do Governo em Educação Terciária (% do PIB)	Banco Mundial
Taxa de alfabetização	Banco Mundial
Valor agregado de fabricação de média e alta tecnologia	Banco Mundial
Despesas em P&D (% do PIB)	Banco Mundial
Exportações de Alta Tecno (US$)	Banco Mundial
Importação de Bens e Serviços (% do PIB)	Banco Mundial
Índice GINI	Banco Mundial
1/População que vive em favelas	Banco Mundial
PL 20% mais ricos	Banco Mundial
PL 20 % mais pobres	Banco Mundial

	Score Global de Liberdade Econômica	Direito de Propriedade	Integridade do Governo	Eficácia Judicial	Carga Tributária	Gastos Públicos	Saúde Fiscal	Liberdade de Negócios	Liberdade de Trabalho	Liberdade Monetária	Liberdade Comercial	Investiment Freedom	Liberdade Financeira
Score Global de Liberdade Econômica	1,00	0,88	0,81	0,82	-0,08	-0,19	0,50	0,83	0,63	0,66	0,80	0,80	0,81
Direito de Propriedade	0,88	1,00	0,92	0,92	-0,33	-0,42	0,28	0,86	0,51	0,63	0,76	0,74	0,74
Integridade do Governo	0,81	0,92	1,00	0,93	-0,43	-0,50	0,24	0,81	0,45	0,55	0,68	0,64	0,67
Eficácia Judicial	0,82	0,92	0,93	1,00	-0,38	-0,41	0,18	0,83	0,54	0,53	0,66	0,63	0,65
Carga Tributária	-0,08	-0,33	-0,43	-0,38	1,00	0,46	-0,01	-0,20	0,05	-0,17	-0,08	-0,26	-0,15
Gastos Públicos	-0,19	-0,42	-0,50	-0,41	0,46	1,00	0,02	-0,33	0,01	-0,17	-0,39	-0,19	-0,24
Saúde Fiscal	0,50	0,28	0,24	0,18	-0,01	0,02	1,00	0,26	0,16	0,25	0,34	0,32	0,26
Liberdade de Negócios	0,83	0,86	0,81	0,83	-0,20	-0,33	0,26	1,00	0,54	0,53	0,70	0,62	0,67
Liberdade de Trabalho	0,63	0,51	0,45	0,54	0,05	0,01	0,16	0,54	1,00	0,35	0,41	0,39	0,42
Liberdade Monetária	0,66	0,63	0,55	0,53	-0,17	-0,17	0,25	0,53	0,35	1,00	0,56	0,65	0,62
Liberdade Comercial	0,80	0,76	0,68	0,66	-0,08	-0,39	0,34	0,70	0,41	0,56	1,00	0,61	0,68
Investiment Freedom	0,80	0,74	0,64	0,63	-0,26	-0,19	0,32	0,62	0,39	0,65	0,61	1,00	0,80
Liberdade Financeira	0,81	0,74	0,67	0,65	-0,15	-0,24	0,26	0,67	0,42	0,62	0,68	0,80	1,00

Figura 1: Correlações entre Indicadores

Os indicadores destacados em vermelho na Tabela 1 foram retirados no filtro de correlação. Além disso, como os indicadores Receitas do Governo e Despesas do Governo foram transformados em um único critério, restaram para a análise um conjunto de 13 critérios além do indicador de Índice de Complexidade Econômica (usado como atributo alvo neste experimento).

3.2. Redução de Dimensionalidade com UMAP e Aplicação do ELECTRE-TREE

Mesmo com a retirada dos critérios com alta correlação na etapa de tratamento dos dados, a matriz de decisão $\mathbf{X}$ ficou com um total de 13 atributos. Esse grande número de dimensões pode ser um complicador para algoritmos de aprendizado. Neste cenário, algumas técnicas computacionais buscam reduzir a dimensionalidade de um conjunto de dados, porém, mantendo as suas características. Dentre as opções, destacam-se técnicas como *Principal Component Analysis* (PCA) e o recente método *Uniform Manifold Approximation and Projection* (UMAP) [McInnes et al., 2018]. De todo modo, não é comum a aplicação de técnicas de redução de dimensionalidade em métodos de elicitação indireta de preferências para AMD. Neste experimento, exploramos essa combinação com a utilização do UMAP para reduzir a dimensionalidade do problema de 13 para 2 dimensões antes da aplicação do método ELECTRE-TREE. A Figura 2 apresenta o exemplo de uma projeção dos dados após uma aplicação do UMAP seguida por uma normalização por padronização.

As cores dos pontos no gráfico de dispersão da Figura 2 indicam os países inicialmente alocados nas classes C_1 (vermelho), C_2 (verde) e C_3 (azul) de acordo uma discretização do atributo alvo (Complexidade Econômica). Lembrando que neste experimento de classificação ordinal, temos a ordem de preferências estabelecida como $C_1 \succ C_2 \succ C_3$. Pode-se visualizar que o UMAP distribuiu os objetos de uma maneira que faz sentido em termos das expectativas dos rótulos atribuídos.

Em seguida, aplica-se o ELECTRE-TREE considerando o problema reduzido à nova dimensionalidade. A menor dimensionalidade da matriz facilita a busca por parâmetros capazes de alocar os exemplos de decisão nas classes corretas. O ELECTRE-TREE funciona conforme apresentado no Algoritmo 2 e seu código está disponível para uso acadêmico em `https://github.com/Valdecy/ELECTRE-Tree`. Inicialmente, um experimento foi realizado variando o número de modelos gerados pelo ELECTRE-TREE para compor o comitê classificador do método para uma mesma projeção do UMAP, entre 100 e 150. Um segundo teste foi realizado mantendo o número de modelos fixo em 100, porém, variando projeções do UMAP antes do aprendizado de parâmetros do ELECTRE-TREE. Os resultados da aplicação estão resumidos na Figura

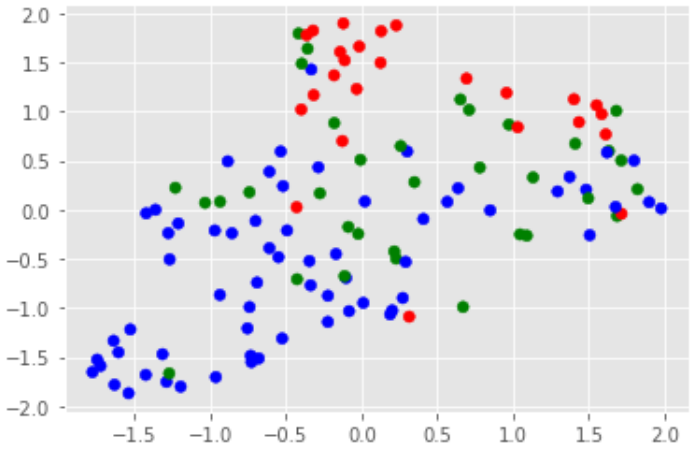


Figura 2: Projeção dos Dados com o UMAP em Duas Dimensões

3 e os hiperparâmetros escolhidos são detalhados a seguir:

- Inferimos limiares de preferência, indiferença, pesos dos critérios e perfis de fronteira;
- λ-corte inferido entre 0,5 e 1;
- Limiares de veto não foram considerados;
- Número de classes: $c = 3$;
- Número de modelos: no primeiro experimento, N variou entre 100 e 150. No segundo experimento, N foi fixado em 100;
- Exemplos de decisão foram considerados;
- Algoritmo Genético: $elite = 1, \eta = 1, \mu = 1, size = 25, rate = 0.10, generations = 50, samples = 0.15$.

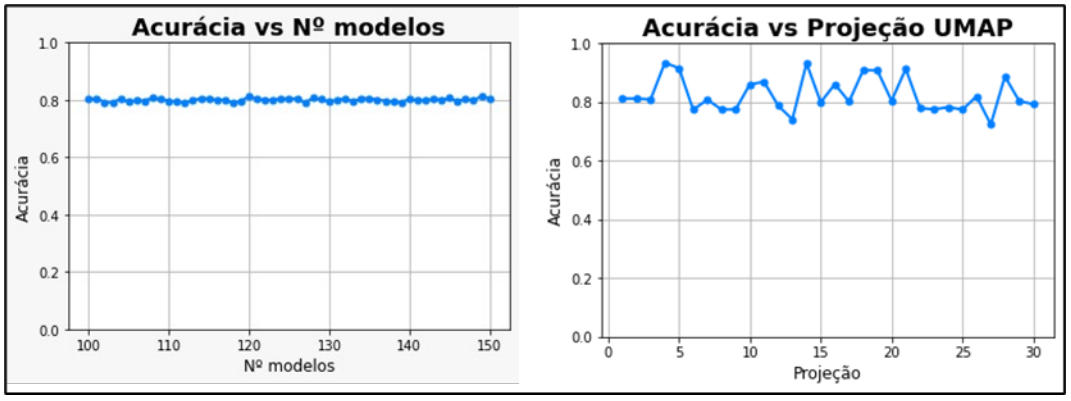


Figura 3: Acurácia preditiva

O gráfico da esquerda na Figura 3 ilustra a acurácia preditiva do primeiro experimento, onde percebe-se um valor que varia em torno de 80% florestas randômica geradas a partir de um número entre 100 a 150 árvores do ELECTRE-TREE. Por sua vez, a gráfica da direita na Figura 3 ilustra os resultados de um experimento com 30 projeções do UMAP geradas seguidas da aplicação do ELECTRE-TREE com os parâmetros listados acima e o número de modelos fixo em 100. Neste caso, a acurácia variou em um intervalo entre 72,4% e 93,4%, com uma acurácia média de 82,5%.

4. Conclusões

Este artigo apresenta uma nova aplicação do recente método de aprendizado de preferências ELECTRE-TREE [Montenegro de Barros et al., 2021]. O método alia características do tradicional ELECTRE-TRI-B com aspectos de modelos de aprendizado de máquina, mais especificamente das árvores randômicas.

Adicionalmente, esta pesquisa usa o procedimento de redução de dimensionalidade UMAP para facilitar o processo de aprendizado de parâmetros do ELECTRE-TREE. Neste tipo de aprendizado, um decisor não precisa explicitar diretamente a grande quantidade de parâmetros necessárias para uma aplicação do ELECTRE-TRI. Ao invés disso, exemplos de aplicação podem ser utilizados para um processo de inferência desses parâmetros, tais como: pesos, limiares de indiferença, limiares de preferência e perfis de fronteira. É possível ainda usar o método com a determinação dos exemplos de classificação realizada por um método não supervisionado como o k-Means. Com o uso do UMAP em uma etapa prévia ao aprendizado do ELECTRE-TREE, a matriz de dados é reduzida de 13 para 2 dimensões. Apesar da contribuição ao processo de aprendizado sob o aspecto de um modelo preditivo, um lado negativo consiste na perda de interpretabilidade dos resultados, característica importante dos métodos AMD.

Em suma, dois experimentos foram realizados com resultados ilustrados na Figura 3. Primeiramente, o número de modelos (correspondente ao número de árvores em uma floresta randômica) variou entre 100 e 150 para uma mesma projeção do UMAP. A acurácia teve pouca variação, com resultados entre torno de 77% e 83%. Em um segundo momento, o número de modelos foi fixado em 100 e 30 projeções do UMAP foram testadas. A acurácia média foi de 82,5%, porém, para alguns pontos, superou-se 90%.

De maneira geral, os resultados foram consistentes com a expectativa dos autores. A pesquisa contribui com a literatura aplicando um método AMD recente em um problema novo e testando um procedimento recente de redução de disponibilidade junto ao método multicritério. Ao longo do texto, detalha-se também o processo de coleta e tratamento de dados. Trabalhos futuros podem incluir novos experimentos considerando a inclusão e exclusão de indicadores, além da análise variações das projeções do UMAP em conjunto com variações de outros hiperparâmetros do método de aprendizado ELECTRE-TREE.

Referências

Basílio, M. P., Pereira, V., Costa, H. G., Santos, M., e Ghosh, A. (2022). A Systematic Review of the Applications of Multi-Criteria Decision Aid Methods (1977–2022). *Electronics*, 11(11): 1720. ISSN 2079-9292. URL https://www.mdpi.com/2079-9292/11/11/1720.

Bouyssou, D., Marchant, T., e Pirlot, M. (2022). A theoretical look at Electre Tri-nB and related sorting models. *4OR*. ISSN 1619-4500. URL https://link.springer.com/10.1007/s10288-022-00501-9.

Chowdhury, P. e Paul, S. K. (2020). Applications of MCDM methods in research on corporate sustainability. *Management of Environmental Quality: An International Journal*, 31(2):385–405. ISSN 1477-7835. URL https://www.emerald.com/insight/content/doi/10.1108/MEQ-12-2019-0284/full/html.

Cinelli, M., Kadziński, M., Gonzalez, M., e Słowiński, R. (2020). How to support the application of multiple criteria decision analysis? Let us start with a comprehensive taxonomy. *Omega*, 96: 102261. ISSN 03050483. URL https://linkinghub.elsevier.com/retrieve/pii/S0305048319310710.

Costa, I. P. d. A., Costa, A. P. d. A., Sanseverino, A. M., Gomes, C. F. S., e dos Santos, M. (2022). Bibliometric studies on Multi-Criteria Decision Analysis (MCDA) methods applied in military problems. *Pesquisa Operacional*, 42. ISSN 1678-5142. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-74382022000100209&tlng=en.

de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., e Garcez, T. V. (2015). *Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis*, volume 231 of *International Series in Operations Research & Management Science*. Springer International Publishing, Cham. ISBN 978-3-319-17968-1. URL http://link.springer.com/10.1007/978-3-319-17969-8.

de Almeida-Filho, A. T., de Lima Silva, D. F., e Ferreira, L. (2021). Financial modelling with multiple criteria decision making: A systematic literature review. *Journal of the Operational Research Society*, 72(10):2161–2179. ISSN 0160-5682. URL https://www.tandfonline.com/doi/full/10.1080/01605682.2020.1772021.

Doumpos, M., Grigoroudis, E., Matsatsinis, N. F., e Zopounidis, C. (2022). Preference Disaggregation Analysis: An Overview of Methodological Advances and Applications. In Greco, S., Mousseau, V., Stefanowski, J., e Zopounidis, C., editors, *Intelligent Decision Support Systems: Combining Operations Research and Artificial Intelligence - Essays in Honor of Roman Słowiński*, p. 73–100. Springer Cham, 1 edition. ISBN 978-3-030-96318-7. URL https://link.springer.com/10.1007/978-3-030-96318-7_5.

Fernández, E., Figueira, J. R., e Navarro, J. (2020). Interval-based extensions of two outranking methods for multi-criteria ordinal classification. *Omega*, 95:102065. ISSN 03050483. URL https://linkinghub.elsevier.com/retrieve/pii/S0305048319300726.

Gomes, L. F. A. M. e Gomes, C. F. S. (2019). *Princípios e métodos para tomada de decisão: enfoque multicritério*. Atlas, São Paulo, 6 edition. ISBN 978-85-97-02158-5.

Greco, S., Ehrgott, M., e Figueira, J. R., editors (2016). *Multiple Criteria Decision Analysis: State of Art Surveys*, volume 233 of *International Series in Operations Research & Management Science*. Springer New York, New York, NY, 2 edition. ISBN 978-1-4939-3093-7. URL http://link.springer.com/10.1007/978-1-4939-3094-4.

Hartmann, D., Zagato, L., Gala, P., e Pinheiro, F. L. (2021). Why did some countries catch-up, while others got stuck in the middle? Stages of productive sophistication and smart industrial policies. *Structural Change and Economic Dynamics*, 58:1–13. ISSN 0954349X. URL https://doi.org/10.1016/j.strueco.2021.04.007.

Heritage Foundation (2022). Index of Economic Freedom. URL https://www.heritage.org/index/.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3 (2):92–113. ISSN 25225820.

Kadziński, M. e Martyn, M. (2021). Enriched preference modeling and robustness analysis for the ELECTRE Tri-B method. *Annals of Operations Research*, 306(1-2):173–207. ISSN 0254-5330. URL https://link.springer.com/10.1007/s10479-020-03833-z.

Leoneti, A. B. e Gomes, L. F. A. M. (2022). A typology for MCDM methods based on the rationality of their pairwise comparison procedures. *Pesquisa Operacional*, 42. ISSN 1678-5142. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-74382022000100215&tlng=en.

Madhooshiarzanagh, P. e Abi-Zeid, I. (2021). A disaggregation approach for indirect preference elicitation in Electre TRI-nC: Application and validation. *Journal of Multi-Criteria Decision Analysis*, 28(3-4):144–159. ISSN 10991360. URL https://onlinelibrary.wiley.com/doi/10.1002/mcda.1730.

McInnes, L., Healy, J., Saul, N., e Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29):861. ISSN 2475-9066. URL http://joss.theoj.org/papers/10.21105/joss.00861.

Montenegro de Barros, G. M., Pereira, V., e Roboredo, M. C. (2021). ELECTRE tree: a machine learning approach to infer ELECTRE Tri-B parameters. *Data Technologies and Applications*, 55(4):586–608. ISSN 2514-9288. URL https://www.emerald.com/insight/content/doi/10.1108/DTA-10-2020-0256/full/html.

Ramezanian, R. (2019). Estimation of the profiles in posteriori ELECTRE TRI: A mathematical programming model. *Computers & Industrial Engineering*, 128(November 2018):47–59. ISSN 03608352. URL https://doi.org/10.1016/j.cie.2018.12.034https://linkinghub.elsevier.com/retrieve/pii/S0360835218306375.

Roy, B. e Bouyssou, D. (1993). *Aide multicritère à la décision: méthodes et cas*. Economica Paris.

The Growth Lab at Harvard University (2022). The Atlas of Economic Complexity. URL https://atlas.cid.harvard.edu/.

The World Bank (2022). World Bank Open Data. URL https://data.worldbank.org/.

Zanghelini, G. M., Cherubini, E., e Soares, S. R. (2018). How Multi-Criteria Decision Analysis (MCDA) is aiding Life Cycle Assessment (LCA) in results interpretation. *Journal of Cleaner Production*, 172:609–622. URL https://linkinghub.elsevier.com/retrieve/pii/S0959652617325350.

https://proceedings.science/p/157586?lang=pt-br