57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

PREPPY: UMA ABORDAGEM PARA INDUÇÃO DE REGRAS NO CONTEXTO DRSA

Hugo de Moraes Ramos Borba
UFPE - Universidade Federal de Pernambuco
Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, 50670-901
hugo.mrborba@ufpe.br

Caroline Maria de Miranda Mota
UFPE - Universidade Federal de Pernambuco
Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, 50670-901
caroline.mota@ufpe.br

Duan Vilela Ferreira
UFPE - Universidade Federal de Pernambuco
Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, 50670-901
duanvilela@gmail.com

RESUMO

Ambientes decisórios complexos pedem métodos multicritério alinhados às preferências do decisor. Neste artigo, é apresentado o PREPPY, algoritmo de indução de regras baseado na DRSA que, de forma incremental e top-down, especializa conjuntos de condições guiado por ganho de informação, precisão e cobertura, explorando permutações de critérios promissores. O método integra atributos quantitativos e qualitativos com ordenações flexíveis, evitando discretizações artificiais e atenuando a explosão combinatória por critérios de parada. Após a indução, regras redundantes são podadas e conflitos resolvidos priorizando maior precisão e, em seguida, cobertura, preservando concisão e interpretabilidade. Testes preliminares sugerem modelos estáveis e com boa capacidade de generalização, embora o custo computacional cresça em bases extensas, o que motiva otimizações, validações adicionais e avaliação sob consistência variável.

PALAVRAS CHAVE. Dominance-based Rough Set Approach (DRSA). Método de Decisão Multicritério (MDCM). Algoritmo de Aprendizado de Máquina.

Desenvolvimento e Aplicação de Método Multicritério de Apoio à Decisão (MCDA)

ABSTRACT

Complex decision-making environments demand multicriteria methods aligned with the decision maker's preferences. In this article, we present PREPPY, a DRSA rule induction algorithm which, in an incremental and top-down manner, specializes condition sets guided by information gain, precision, and recall, while exploring permutations of promising criteria. The method integrates quantitative and qualitative attributes with flexible orderings, avoiding artificial discretizations and mitigating combinatorial explosion through stopping criteria. After induction, redundant rules are pruned and conflicts are resolved by prioritizing higher precision and, subsequently, coverage, thus preserving conciseness and interpretability. Preliminary tests suggest stable models with good generalization ability, although computational cost increases on large datasets, motivating optimizations, further validations, and evaluation under variable consistency.

KEYWORDS. Dominance-based Rough Set Approach (DRSA). Multicriteria Decision Method (MCDM). Machine Learning Algorithm.

Development and Application of MCDA

1. Introdução

A crescente complexidade dos processos decisórios em ambientes com dados heterogêneos impulsiona o desenvolvimento de métodos que integrem múltiplos critérios diversos para extrair conhecimento alinhado às preferências de decisores [Blaszczynski et al., 2022]. Neste contexto, a construção de modelos de decisão multicritério normalmente requer o estabelecimento de parâmetros preferenciais como limiares e funções de preferência, pesos ou constantes de escala que normalmente demandam alto esforço cognitivo para os decisores, dada sua complexidade [Greco, Matarazzo, Slowinski e 2001]. No entanto, em cenários de problemáticas de classificação ordenada, o método *Dominance-based Rough Set Approach (*DRSA) [Greco, Matarazzo e Slowinski, 2001] apresenta formas de superar barreiras da definição parâmetros preferenciais, dado que a aplicação de modelos baseados no DRSA não requer tais informações [Błaszczyński, Greco e Słowiński, 2006].

O método DRSA é derivado da *Rough Set Theory* [Pawlak, 1982] e incorpora o conceito de dominância entre objetos (alternativas de decisão) para construção de subconjuntos dominantes e dominados, aproximações (superiores e inferiores) e uniões de classe, sendo estas a principal referência para indução de regras de decisão que permitem a extrapolação de decisões passadas em cenários futuros, principal produto deste tipo de modelo [Greco, Matarazzo e Slowinski, 2001]. No entanto, sabe-se que a geração de regras de decisão baseadas em dados rotulados é amplamente difundida na literatura e pode ser realizada através de variadas ferramentas de aprendizado de máquina [Grzymala-Busse, 1992]. Ainda assim, a maioria destas ferramentas está associada à geração de regras de decisão para classes nominais que usualmente não possuem relação preferencial entre si. Neste sentido, o algoritmo DomLEM [Greco et al., 2000] representa um avanço no desenvolvimento de abordagens para indução de regras de decisão para classificação ordenada, com aplicações em diversas áreas como política, medicina e o setor industrial [Slowinski, Greco e Matarazzo, 2015]. Contudo, é fato que o desenvolvimento de algoritmos desta natureza ainda é incipiente na literatura [Szelag, Greco e Slowinski, 2014].

Assim, este trabalho apresenta o desenvolvimento do algoritmo *Permutation-based Pattern Yielding* (PREPPY) para indução de regras de decisão para aplicação em modelos de decisão baseados no método DRSA. Em linhas gerais, o PREPPY possui um processo estruturado de indução baseado na permutação de componentes possíveis para regras de uma determinada classe, possuindo como direcionadores de seleção destes componentes os indicadores de precisão e cobertura. As etapas de processamento do PREPPY serão discutidas a seguir.

2. Algoritmo PREPPY

O PREPPY integra, num único arcabouço, critérios quantitativos e qualitativos (inclusive nominais) com preferências flexíveis do decisor, evitando discretizações. O algoritmo opera sobre uma matriz de decisão com alternativas, critérios e classes. Os critérios são rotulados como ganho (maior é melhor), custo (menor é melhor) ou qualitativo. Em qualitativos ordenáveis, as preferências são representadas por relações parciais. Já os atributos nominais, categorias são tratadas como rótulos distintos sem relações de dominância entre elas.

As relações de dominância e as aproximações inferiores das uniões de classe geradas com o DRSA fornecem um núcleo de exemplos positivos “certos”, garantindo regras consistentes com as preferências. O indicador de ganho de informação é aplicado para priorização de critérios ao medir a redução de incerteza da decisão. Em outras palavras, considera divisões coerentes com a direção ganho ou custo; para atributos qualitativos, segue a partição por categorias ou ordens parciais, tornando-os comparáveis em diferentes escalas. As etapas de geração de regras desenvolvidas pelo PREPPY podem ser observadas no pseudocódigo no Quadro 1:

Quadro 1 – Pseudocódigo do algoritmo PREPPY

Para um problema de decisão com conjunto de alternativas U, atributos $A = \{a_1, a_2, \dots, a_m\}$ e classes ordenadas $\{C_1, C_2, \dots, C_n\}$ (da mais baixa para a mais alta), cada objeto $x \in U$ é descrito por $(x_1, x_2, \dots, x_m, d)$, em que d trata-se de sua classe decisória. Também se associa uma

preferência $P[c] \in \{ganho, custo, quantitativo\}$ a cada atributo. Para este pseudocódigo é assumida como conhecida a relação de dominância e, para cada t, a união ascendente $C_t^{\geq} = \bigcup\{s \geq t\} C_s$ e sua aproximação inferior $Lower(C_t^{\geq})$ (DRSA).

Passo 1: Priorizar critérios por informação
Calcular a *entropia da decisão* $H(d)$ e a *entropia condicional* $H(d|a)$ para cada atributo a. Definir o *ganho de informação* $IG(a) = H(d) - H(d|a)$. Ordenar os critérios por IG em ordem decrescente e gerar um *conjunto de permutações* Π priorizadas, possivelmente limitado por um máximo de permutações estabelecido.

Passo 2: Definir alvos por classe
Para cada classe $t \in \{2, \dots, n\}$ (exclui-se a classe mais inferior por definição), define-se o conjunto-alvo $T_t = Lower(C_t^{\geq})$ e inicializa-se o conjunto de regras $\mathcal{R} \leftarrow \emptyset$.

Passo 3: Cobertura sequencial por permutação
Para cada $\pi \in \Pi$ e para cada $t = n, n-1, \dots, 2$, enquanto $T_t \neq \emptyset$, construir uma *regra candidata* r adicionando condições na ordem de π:
(i) se $P[c] = qualitativo$, registrar os valores presentes das categorias em T_t;
(ii) se $P[c] \neq qualitativo$, inicializar um limiar v_0 coerente ($ganho$: $v_0 = min\ T_t$; $custo$: $v_0 = max\ T_t$) e fixar a direção de ajuste ($ganho$: $\downarrow$, $custo$: $\uparrow$).

Passo 4: Otimização bifásica do limiar
Para cada critério quantitativo incluído, ajustar o limiar v_0 em duas fases:
(i) maximizar a precisão movendo v_0 na direção definida até não haver melhora;
(ii) expandir a cobertura mantendo a precisão constante, avançando v_0 enquanto a precisão não se reduz.

Passo 5: Avaliar, registrar e retirar cobertos
Calcular cobertura e precisão de r sobre T_t. Acrescentar r a $\mathcal{R}$ com seus indicadores e remover de T_t os objetos corretamente cobertos. Encerrar antecipadamente se todos os T_t ficarem vazios.

Em termos operacionais, a indução de regras é incremental, top-down e parcimoniosa: uma condição é incorporada apenas quando melhora a precisão ou mantém a precisão e aumenta a cobertura. A especialização cessa quando precisão e cobertura atingem o máximo de 100% ou quando não há mais ganho; se a cobertura não atinge o máximo, seguem-se novas permutações até o esgotamento ou limite estabelecido. No pós-processamento, aplica-se: (1) escolher regra com cobertura máxima (se houver); (2) caso contrário, ordenar por precisão; (3) remover duplicatas; (4) eliminar regras dominadas, preferindo as mais gerais.

Os conflitos entre regras são tratados de modo hierárquico: prevalece primeiro a regra de maior precisão, depois a de maior cobertura e, se o empate persistir, aplica-se a classe prioritária — tipicamente a superior. Casos com valores faltantes não são ignorados: embora não contribuam como evidência direta para o critério específico, as observações incompletas ainda podem ser avaliadas pelas regras parciais formadas.

Por fim, a eficácia do modelo induzido é avaliada, são reportadas métricas de desempenho como acurácia, F1 Score, MCC e Kappa, além da matriz de confusão e da verificação de consistência em relação às aproximações do DRSA. Essa etapa garante não apenas qualidade preditiva, mas também confiabilidade interpretativa, mantendo o equilíbrio entre precisão, concisão e generalização.

3. Conclusão

Os testes preliminares do PREPPY evidenciaram a capacidade de induzir regras que equilibram cobertura e precisão, produzindo modelos concisos e interpretáveis. No entanto, novos testes e validações ainda são necessários, dado que a exploração de permutações pode elevar substancialmente o tempo de processamento em bases extensas, além de impor limites práticos de escalabilidade quando há muitos critérios e objetos. Tratando-se de dados reais, com possíveis ruídos, a aplicação adequada do algoritmo requer uma prévia estruturação dos exemplos de decisão para evitar o desbalanceamento entre classes e o risco de obtenção de regras excessivamente específicas ao conjunto analisado. Assim, torna-se fundamental estender os experimentos com bases de naturezas distintas e validações mais robustas para verificar a estabilidade dos resultados e a capacidade de generalização. Esses cuidados fundamentam as melhorias propostas a seguir.

Como novas direções, sugere-se: (i) otimizações algorítmicas como vetorização, paralelismo e gestão eficiente de memória; (ii) validações adicionais em conjuntos de dados diversos, com comparativos sistemáticos a métodos estabelecidos, para calibrar relações de preferência; (iii) refatoração e modularização do código com boas práticas de engenharia e implantação em nuvem para escalar desempenho e disponibilidade; (iv) adoção de mecanismos de consistência variável (VC-DRSA) para lidar com ruídos e inconsistências [Blaszczynski , Slowinski e Szelag, 2011].

Referências

1. BŁASZCZYŃSKI, J.; GRECO, S.; MATARAZZO, B.; SZELĄG, M. Dominance-based rough set approach, basic ideas and main trends. *arXiv*, 2022.
2. BŁASZCZYŃSKI, J.; GRECO, S.; SŁOWIŃSKI, R.; SZELĄG, M. On Variable Consistency Dominance-Based Rough Set Approaches. In: INTERNATIONAL CONFERENCE ON ROUGH SETS AND CURRENT TRENDS IN COMPUTING (RSCTC), 2006, Kobe. Anais. Berlin: Springer-Verlag, 2006. p. 191–202. (Lecture Notes in Artificial Intelligence, v. 4259).
3. BŁASZCZYŃSKI, J.; SŁOWIŃSKI, R.; SZELĄG, M. Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences*, v. 181, p. 987–1002, 2011.
4. GRECO, S.; MATARAZZO, B.; SŁOWIŃSKI, R.; STEFANOWSKI, J. An algorithm for induction of decision rules consistent with the dominance principle. In: *Rough Sets and Current Trends in Computing*, 2000. p. 304–313, 2000.
5. GRECO, S.; MATARAZZO, B.; SŁOWIŃSKI, R. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, v. 129, n. 1, p. 1–47, 2001.
6. GRZYMALA-BUSSE, J. W. LERS — A system for learning from examples based on rough sets. In: *Intelligent Decision Support*, 1992. Dordrecht: Kluwer Academic Publishers, 1992. p. 3–18.
7. PAWLAK, Z. Rough sets. *International Journal of Computer & Information Sciences*, v. 11, n. 5, p. 341–356, 1982.
8. SŁOWIŃSKI, R.; GRECO, S.; MATARAZZO, B. Rough set methodology for decision aiding. In: *Springer Handbook of Computational Intelligence*. Springer, 2015. p. 349–370.
9. SZELAG, M.; GRECO, S.; SLOWINSKI, R. Variable consistency dominance-based rough set approach to preference learning in multicriteria ranking. Information Sciences, v. 277, p. 525–552, 2014.