

Um Estudo sobre a Aplicação de Algoritmos Genéticos Multiobjetivo para a Seleção de Atributos

Newton Spolaôr^{1,2}, Ana Carolina Lorena¹, Huei Diana Lee²

¹Centro de Matemática, Computação e Cognição/Universidade Federal do ABC
Santo André, São Paulo, Brasil

²Centro de Engenharias e Ciências Exatas/Universidade Estadual do Oeste do Paraná
Laboratório de Bioinformática/Parque Tecnológico Itaipu
Foz do Iguaçu, Paraná, Brasil

{newtonspolaor, aclorena, hueidianalee}@gmail.com

Abstract. *The occurrence of non-important attributes in databases may impair the performance of computational processes for knowledge discovery, which motivates the application of feature selection techniques. The combinatorial nature of this problem shows to be appropriate to the application of heuristic methods such as genetic algorithms to obtain or approximate optimal subsets of attributes. In many cases one wants to optimize conflicting goals, such as the predictive performance of a subset of attributes and the cardinality of that subset, which enables the formulation of this task as a multiobjective optimization problem. This paper aims to present a preliminary study on state-of-art work related to this context.*

Resumo. *A ocorrência de atributos não-importantes em Bases de Dados pode prejudicar o desempenho de processos computacionais de extração de conhecimento, motivando a realização da tarefa de Seleção de Atributos. A natureza combinatorial deste problema é apropriada para a aplicação de métodos heurísticos, como os Algoritmos Genéticos, de modo a obter ou aproximar subconjuntos de atributos ótimos. Em muitos casos deseja-se otimizar, em um subconjunto de atributos, objetivos conflitantes como o desempenho preditivo e a cardinalidade, o que possibilita a formulação dessa tarefa como um problema de otimização multiobjetivo. Este trabalho tem como objetivo apresentar um estudo preliminar sobre trabalhos do estado da arte relacionados a este contexto.*

1. Introdução

O progresso tecnológico tem possibilitado em distintas áreas a construção de conjuntos de dados cada vez maiores, sob a forma de Bases de Dados (BD). Essas bases podem ser submetidas a processos de análise inteligente, como a Mineração de Dados, com o intuito de gerar modelos computacionais que auxiliem à tomada de decisões nesses domínios [Han and Kamber 2006]. Esses modelos podem ser elaborados por meio de tarefas de Extração de Padrões, como classificação ou *clustering*, em que um algoritmo de indução constrói inferências (hipóteses) sobre os padrões presentes nos dados.

A precisão dos modelos pode ser prejudicada quando ocorrem atributos, ou características que descrevem cada exemplo registrado nas BD, irrelevantes ou redundantes

entre si. Tendo em vista estes atributos não-importantes, outro possível aspecto negativo é a relação proporcional entre complexidade computacional e a quantidade de atributos, inerente a muitos algoritmos de indução. Nesse sentido, um processo de análise inteligente, de modo geral, pode tornar-se mais complexo ou inviável. Estes e muitos outros aspectos estão relacionados ao problema da “maldição da dimensionalidade”, o qual representa uma das principais justificativas para a área de pesquisa em Seleção de Atributos (SA) [Liu and Motoda 2008b].

A tarefa de SA pode ser definida como um processo de busca por um subconjunto de atributos importantes, em termos de relevância e/ou não-redundância, que auxilie na identificação de padrões inerentes aos dados. A importância de subconjuntos de atributos pode ser estimada a partir de critérios como o desempenho de um algoritmo de indução e a correlação entre atributos [Liu and Motoda 1998]. Um atributo pode ser definido como relevante se, ao ser removido do conjunto de dados, implica em perda de desempenho do indutor. Dois atributos são considerados não-redundantes entre si quando seus valores não são significativamente correlacionados. A busca pelo subconjunto ótimo em geral é combinatorial (NP-difícil) [Charikar et al. 2000], o que motiva a aplicação de métodos heurísticos como os Algoritmos Genéticos (AG) [Yang and Honavar 1998].

Os AG permitem a resolução/aproximação de problemas de otimização e busca, como a SA, por meio da simulação computacional dos mecanismos de genética e dos processos de seleção e evolução natural em uma população de indivíduos (soluções) [Mitchell 1998]. Esta simulação, fundamentada basicamente nos operadores de *crossover* (cruzamento), mutação e seleção, possibilita a formulação do problema como a otimização de uma função de aptidão, a qual mede a qualidade de potenciais soluções. De modo similar aos processos biológicos, a tendência é de que, ao longo das gerações, a aptidão média da população seja maximizada e uma ou mais soluções aproximem ou converjam para um ótimo local ou global.

Em determinados problemas de SA, a importância dos atributos é definida em termos de objetivos ou critérios distintos e geralmente conflitantes [Zhu et al. 2009]. Em muitos casos deseja-se, por exemplo, minimizar simultaneamente a quantidade (cardinalidade) de atributos em um subconjunto e a taxa de erro obtida com este subconjunto, estimada por meio de um algoritmo de indução. Em algoritmos de classificação, por exemplo, esta estimativa pode ser obtida por meio da identificação da porcentagem de exemplos classificados com um rótulo (atributo classe) não condizente aos padrões inerentes. O complemento desta porcentagem corresponde à taxa de acerto ou acurácia.

Relações de *tradeoff* como essas permitem formular a SA como um problema de Otimização Multiobjetivo (OM), a qual consiste em um tópico de interesse em teoria da computação [Ishibuchi et al. 2008], biologia [Handl et al. 2007], economia [Tapia and Coello 2007] e engenharia [Andersson 2000]. Na literatura são propostos distintos métodos para realizar a SA por meio de AG multiobjetivo, o que pode ser justificado, por um lado, pelos AG oferecerem suporte à otimização conjunta de múltiplas soluções em uma população. Por outro lado, a OM pode prover aos AG uma formulação representativa do problema sob investigação, em termos de *tradeoffs*, tornando possível obter uma ou mais soluções equivalentes para auxiliar em processos de tomada de decisão.

Este trabalho em andamento tem como objetivo apresentar um estudo preliminar

sobre alguns dos métodos descritos na literatura neste contexto. O trabalho constitui parte do Projeto de Análise Inteligente de Dados (AID), o qual é desenvolvido em uma parceria entre o Centro de Matemática, Computação e Cognição da Universidade Federal do ABC (UFABC), o Laboratório de Bioinformática da Universidade Estadual do Oeste do Paraná (UNIOESTE), o Laboratório de Inteligência Computacional da Universidade de São Paulo e o Serviço de Coloproctologia da Universidade Estadual de Campinas.

O restante do trabalho está organizado do seguinte modo. Na Seção 2 são abordados os temas de SA, AG e OM. Na Seção 3, são descritas algumas aplicações relacionadas a este contexto. Considerações relacionadas a este estudo são apresentadas na Seção 4.

2. Conceitos Preliminares

Nesta seção serão introduzidos alguns dos principais conceitos relacionados a Seleção de Atributos, Algoritmos Genéticos e Otimização Multiobjetivo. Uma descrição mais detalhada desses assuntos pode ser encontrada em [Liu and Motoda 2008a, Coello 2006, Dy et al. 2004, Freitas 2004, Everitt et al. 2001, Liu and Motoda 1998, Mitchell 1998].

2.1. Seleção de Atributos

O objetivo da tarefa de SA para dados supervisionados, em termos de precisão, consiste na identificação de um subconjunto mínimo de atributos importantes, os quais apresentam distribuição de probabilidade condicional¹ semelhante à obtida com todos os atributos [Yu and Liu 2004]. A importância desse subconjunto pode ser definida em relação a diferentes medidas de avaliação, como informação e dependência (correlação), aplicáveis também a dados sem a presença do atributo classe (não-supervisionados) [Dy et al. 2004].

Existem duas abordagens principais para a SA, de acordo com a interação realizada com o algoritmo de indução. A abordagem filtro utiliza medidas de importância associadas a propriedades inerentes aos dados, enquanto a abordagem *wrapper* necessita do indutor para avaliar a qualidade do subconjunto [Yang and Honavar 1998]. Neste contexto, podem ser utilizados algoritmos de classificação, como o C4.5 [Quinlan 1993], as Redes Neurais Artificiais (RNA) [Haykin 1994] e as *Support Vector Machines* (SVM) [Cristianini and Shawe-Taylor 2000], e de *clustering*, como o *Single-Link* (SL) [Florek et al. 1951] e o *K-Means* (KM) [MacQueen 1965].

2.2. Algoritmos Genéticos

Possíveis soluções do problema sob investigação podem ser modeladas através de indivíduos (cromossomos) em um AG. A codificação mais usual em um AG básico corresponde a uma sequência binária de h bits (genes), embora também seja possível representar sequências ordenadas de elementos e valores reais. No problema de SA os cromossomos binários representam, em geral, os atributos selecionados e os eliminados por meio dos valores 1 e 0, respectivamente. A população de indivíduos de um AG básico é submetida iterativamente aos operadores genéticos até que um determinado critério de parada, como a quantidade de iterações ou o valor específico de aptidão, seja atingido.

A seleção dos melhores indivíduos para reprodução pode ser realizada por meio de técnicas como roleta, *ranking* e torneio binário. Enquanto a técnica da roleta define

¹Neste contexto, corresponde à probabilidade de se obter os distintos atributos classe, considerando os valores dos atributos do subconjunto em investigação.

a probabilidade de seleção de um indivíduo proporcionalmente à aptidão associada, a técnica de *ranking* determina a probabilidade em termos da ordenação dos cromossomos por aptidão. O torneio binário seleciona dois indivíduos da população e gera um valor randômico $r \in [0, 1]$. Se r for maior que um limiar específico, atribui-se ao cromossomo de maior aptidão o direito de reproduzir; caso contrário, seleciona-se o outro indivíduo.

O operador básico de *crossover* define nos cromossomos selecionados para reprodução u pontos ou *locus*, em que $(1 \leq u \leq (h - 1))$. As subsequências situadas nos intervalos entre estes pontos são alocadas alternadamente entre os descendentes. O *crossover* uniforme atribui, para cada um dos *bits* dos descendentes, o *bit* na posição correspondente em um dos pais. A mutação, na maioria das aplicações, consiste na substituição do valor de um *bit* randomicamente selecionado pelo outro valor possível (*flip*). Na Figura 1 é apresentada uma exemplificação destes operadores, considerando $u = 1$ para o *crossover*. Em algumas aplicações utiliza-se também a técnica de elitismo, a qual permite manter os e melhores indivíduos intactos para as próximas gerações.

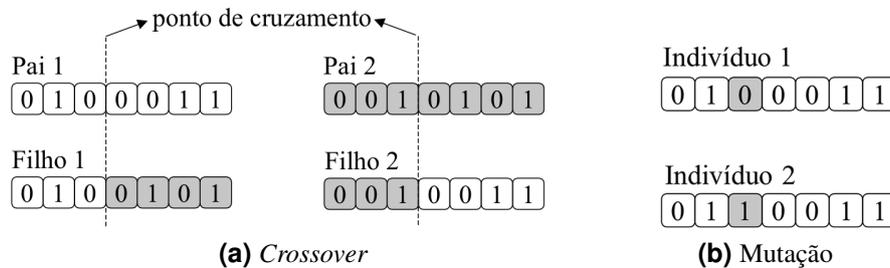


Figura 1: Ilustração da aplicação de operadores genéticos.

2.3. Otimização Multiobjetivo

Uma grande quantidade de problemas podem ser resolvidos adequadamente apenas se muitos objetivos, em geral conflitantes, forem satisfeitos. O problema pode se tornar mais complexo se restrições forem adotadas. Nesse sentido, distintas estratégias para relacionar vários objetivos em um problema de otimização foram propostas. Uma das mais usuais consiste na combinação linear entre os objetivos por meio de uma função agregadora, a qual pode ser exemplificada pela Equação 1, considerando um problema de SA definido pelo *tradeoff* entre acurácia (Obj_1) e cardinalidade (Obj_2) [Roy and Bhattacharya 2008].

$$Obj = Obj_1 \times w_1 + Obj_2 \times w_2 \quad (1)$$

$$\sum w_i = 1 \quad (2)$$

Desse modo, os dois objetivos são representados por um objetivo Obj , o que possibilita a aplicação de métodos tradicionais de otimização sem muitas adaptações. Outra estratégia comumente adotada se fundamenta na dominância de Pareto [Pareto 1896], a qual possibilita a comparação entre soluções considerando todos os objetivos equivalentemente. Na Figura 2 é ilustrado um exemplo de aplicação desta estratégia em um AG para um problema de minimização de três objetivos. Dados dois cromossomos z_1 e z_2 em um AG, os vetores v_1 e v_2 indicam em $v_i[j]$ a qualidade de z_i para o j -ésimo objetivo.

Uma solução z_1 é denominada dominante (ou não-dominada) em Pareto, em relação a z_2 , se e somente se nenhum elemento de v_1 é pior do que v_2 , e pelo menos

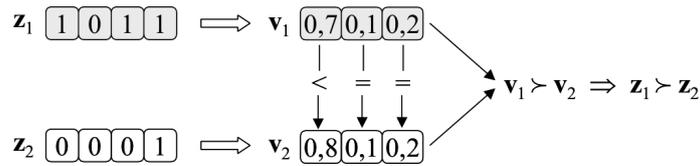


Figura 2: Aplicação da dominância de Pareto a um problema de minimização.

um elemento $v_1[j]$ é melhor que $v_2[j]$. O conjunto de todas as soluções não-dominadas em Pareto define uma região denominada fronteira de Pareto.

3. Aplicações de Algoritmos Genéticos Multiobjetivo em Seleção de Atributos

O estado da arte apresenta distintas propostas para realizar a otimização da tarefa SA por meio de AG multiobjetivo. Algumas propõem hibridismos entre as abordagens filtro e *wrapper*, enquanto outras aplicam métodos clássicos de AG como o *Nondominated Sorting Genetic Algorithm* (NSGA) [Srinivas and Deb 1994] e o NSGA-II [Deb et al. 2002] associados a modificações específicas. Nesta seção são apresentados sucintamente alguns dos estudos mais recentes e relevantes neste contexto. Exemplos de outros trabalhos relacionados a este contexto incluem [Suguna and Thanushkodi 2009, Sánchez et al. 2008, Geng et al. 2007, Oliveira et al. 2006, Emmanouilidis 2002, Emmanouilidis et al. 2000].

3.1. Frameworks Fundamentados em Algoritmos Meméticos (Trabalho #1)

Em [Zhu et al. 2009] são descritos dois algoritmos meméticos, os quais consistem em um hibridismo entre Algoritmos Genéticos e um procedimento de otimização local. Este procedimento atua como um método de SA pertencente à abordagem filtro. Neste trabalho são adotadas duas medidas de importância propostas em [Zhu et al. 2008], as quais são baseadas na relevância de cada atributo em relação à discriminação de quaisquer classes² e em relação à discriminação de um determinado subconjunto de classes³. Os experimentos realizados, utilizando o classificador C4.5 para avaliação da qualidade dos subconjuntos, demonstraram que os métodos propostos são competitivos ou superiores a alguns métodos para SA, como *Fast Correlation-Based Filter* (FCBF) [Yu and Liu 2004] e NSGA-II.

No primeiro método (*Trabalho #1a*), o *tradeoff cardinalidade × erro de generalização* é reformulado como um problema de minimização do objetivo único $J(s)$, com maior prioridade para a taxa de erro e utilização da cardinalidade para desempate em subconjuntos de atributos s de mesma precisão. O procedimento de busca local adotado realiza uma otimização no melhor indivíduo (subconjunto de atributos) da população, por meio de operadores meméticos específicos baseados no conceito de aprendizado Lamarckiano [Ong and Keane 2004]. Esta estratégia tem como intuito adicionar atributos que apresentam alta relevância e remover os que são redundantes ou irrelevantes no subconjunto em investigação, o qual posteriormente deve disseminar esta melhoria nos outros indivíduos da população por meio do processo evolucionário. Especificamente neste *framework* é possível adotar restrições aos operadores de *crossover* e mutação para não gerar descendentes com mais de m atributos selecionados, em que m é definido pelo usuário.

²Full Class Relevant (FCR).

³Partial Class Relevant (PCR).

O segundo *framework* (*Trabalho #1b*) modela o problema de SA em termos de minimização multiobjetivo da função $F(s) = f_1(s) \times \dots \times f_k(s)$, em que cada objetivo $f_i(s) \mid i \in (1 \dots k)$ corresponde à taxa de erro obtida pelo subconjunto de atributos s na discriminação entre uma classe c_i e todas as classes restantes. Esta formulação permite identificar um conjunto de soluções, importantes de acordo com a medida PCR, por meio da aplicação do procedimento de busca local em cada solução não-dominada. A qualidade da solução s neste caso é estimada por um objetivo randomicamente selecionado. O melhor subconjunto de atributos para a medida FCR é definido pela interseção das soluções não-dominadas identificadas. Todas as soluções identificadas ao final do método podem ser utilizadas para treinar um aglomerado de classificadores (*ensemble*), considerando um subconjunto por classificador. Outra alternativa corresponde à construção de um único classificador em que os atributos são definidos pela interseção dessas soluções.

3.2. Algoritmo Genético Híbrido Aplicados a Dados *Microarray* (*Trabalho #2*)

O método proposto por [Dessi and Pes 2009] utiliza um AG para SA *wrapper* em dados *microarray* [Causton et al. 2003]. Entretanto, este algoritmo é aplicado em um subconjunto reduzido de atributos, definido previamente por meio de uma combinação de técnicas de *ranking*⁴ de atributos para a seleção de subconjuntos de genes relevantes. Desse modo, os autores também propõem um hibridismo, por meio da integração entre um método *wrapper* e várias técnicas da abordagem filtro. Posteriormente realizam-se experimentos para avaliar o desempenho, em termos da tarefa de classificação, do melhor subconjunto de atributos identificado em relação a sete outros trabalhos do estado da arte. O *tradeoff* definido para o problema multiobjetivo da SA corresponde ao balanceamento *cardinalidade* \times *taxa de acerto*, o qual é tratado por meio de uma função agregadora.

Esse método é organizado basicamente em duas etapas. Na primeira etapa define-se o *score* dos atributos para cada uma das M técnicas de *ranking* adotadas. Os T melhores atributos para cada técnica definem um subconjunto específico (*pool*). Cada *pool* representa os atributos que estão presentes em pelo menos i subconjuntos ($1 \leq i \leq M$). Na última etapa, aplica-se o AG conjuntamente ao SVM para reduzir a cardinalidade dos *pools*, identificando o subconjunto que melhor atende ao *tradeoff* estabelecido.

3.3. Outras Aplicações Relacionadas

Em [Roy and Bhattacharya 2008] (*Trabalho #3*) se propõe um método que contempla todo o processo de reconhecimento de imagens de íris, desde o pré-processamento até a classificação. Nos procedimentos relacionados à seleção dos atributos extraídos das imagens, aplica-se um AG para minimizar simultaneamente a cardinalidade e a taxa de erro, estimada por um SVM com parâmetros refinados, caracterizando a abordagem de SA *wrapper*. A estratégia de função agregadora é utilizada para relacionar esses critérios em um único objetivo, por meio da soma ponderada. Os experimentos realizados demonstram que o método proposto supera trabalhos relacionados, atingindo resultados satisfatórios mesmo em conjuntos de dados com classes desbalanceadas.

A metodologia proposta por [Oliveira et al. 2003] (*Trabalho #4*) aplica o algoritmo NSGA para otimizar o *tradeoff* *complexidade* \times *taxa de acerto* no contexto de reconhecimento de dígitos manuscritos em uma BD com milhares de imagens. A aptidão

⁴Baseadas no conceito de separabilidade das classes [Liu and Yu 2002].

dos subconjuntos de atributos, em termos de acurácia, é definida com o auxílio de uma RNA. Em uma etapa posterior, realiza-se a validação da fronteira de Pareto utilizando um conjunto de dados não considerado para a SA. O subconjunto de atributos que apresenta a melhor capacidade de generalização é selecionado como solução final pela metodologia.

Não existem muitas aplicações de otimização multiobjetivo para SA em dados não-supervisionados [Mierswa 2009]. Alguns destes métodos consideram a minimização da cardinalidade simultaneamente à maximização da qualidade dos *clusters* [Kim et al. 2002]. O método proposto em [Mierswa and Wurst 2006] (*Trabalho #5*) formula o problema multiobjetivo de SA de um modo distinto, considerando a maximização da quantidade de atributos ao invés da minimização em termos da dominância de Pareto. O algoritmo NSGA-II é aplicado no problema considerando *crossover* uniforme e mutação simples (*flip*). Os experimentos realizados demonstram que o método proposto por [Mierswa and Wurst 2006] possibilita a identificação de mais pontos na fronteira de Pareto. Algumas das soluções identificadas apresentam uma mudança de inclinação significativa em relação à fronteira de Pareto, o que possibilita a segmentação automática dessa fronteira e a definição de um conjunto menor de soluções ótimas.

3.4. Discussão

As principais diferenças entre os trabalhos descritos, em termos de operadores genéticos e formulações multiobjetivo, são respectivamente apresentadas nas Tabelas 1 e 2. Na Tabela 2 especificamente se descreve o algoritmo de indução utilizado em cada trabalho.

Trabalho	Seleção	Crossover	Mutação
#1a	Ranking linear	Uniforme restritivo	Restritiva
#1b	Torneio binário	Uniforme	Simples (<i>flip</i>)
#2	Roleta	Um ponto	Simples (<i>flip</i>)
#3	Roleta	-	-
#4	Roleta	Um ponto	Simples (<i>flip</i>)
#5	Torneio binário	Uniforme	Simples (<i>flip</i>)

Tabela 1: Operadores genéticos utilizados.

Trabalho	Estratégia Multiobjetivo	Objetivos	Indutor
#1a	Erro de generalização (prioritária)	$J(s)$ (<i>Erro</i>)	C4.5
#1b	Seleção randômica de objetivo $f_i(s)$	$f_1(s) \times \dots \times f_k(s)$	C4.5
#2	Função agregadora	<i>Precisão</i> \times <i>Cardinalidade</i>	SVM
#3	Função agregadora	<i>Erro</i> \times <i>Cardinalidade</i>	SVM
#4	Dominância de Pareto	<i>Erro</i> \times <i>Cardinalidade</i>	RNA
#5	Dominância de Pareto	<i>Qualidade</i> \times <i>Cardinalidade</i>	KM e SL

Tabela 2: Formulação multiobjetivo e algoritmo para estimativa de aptidão.

Na Tabela 1 é possível constatar que a maioria dos trabalhos abordados adotaram mutação simples, com exceção do Trabalho #1a, no qual optou-se em restringir a cardinalidade representada nos cromossomos por meio da mutação e do *crossover*. Nos demais casos, há uma diversidade de variantes em cada operador, o que pode ser justificado pelo fato de não existir uma configuração de AG aplicável para solucionar todos os problemas.

Ao analisar a Tabela 2, constata-se que houve predominância na formulação da SA como um problema de otimização do *tradeoff cardinalidade × erro de generalização*, com algumas variações. No caso do Trabalho #5, por exemplo, optou-se por otimizar uma medida de qualidade dos *clusters* definida no trabalho, tendo em vista que o conjunto de dados considerado possui natureza não-supervisionada. Em relação às estratégias multi-objetivo, constata-se que não há preferência específica, o que pode ser explicado, dentre outras razões, pelas vantagens e desvantagens inerentes a cada estratégia [Freitas 2004].

Todos os trabalhos descritos aplicaram SA *wrapper*, com destaque para os híbridos realizados nos trabalhos #1 e #2 com técnicas da abordagem filtro, os quais possibilitaram agregar vantagens de ambas abordagens e reduzir de modo geral a complexidade computacional. Uma avaliação experimental comparativa pode evidenciar qual o método que permite a identificação dos subconjuntos de atributos mais importantes.

4. Considerações Finais

Neste trabalho foi apresentado um estudo inicial sobre conceitos e propostas da literatura relacionados à otimização do problema de SA por meio de AG multiobjetivo. Foram abordados sucintamente alguns dos trabalhos mais recentes neste contexto, os quais apresentam desempenho superior ou equivalente ao obtido por métodos do estado da arte e por propostas tradicionais. O estudo realizado pode ser útil, por exemplo, para auxiliar na contextualização e na motivação para trabalhos futuros relacionados a este contexto.

A aplicação de AG multiobjetivo no auxílio à tarefa de SA pode ser interessante em aplicações futuras, tendo em vista os resultados promissores obtidos pelo estado da arte. Este trabalho tem como intuito desenvolver futuramente um método, baseado nas aplicações abordadas, que possibilite a otimização e a avaliação conjunta de múltiplos subconjuntos de atributos por meio de AG e OM, respectivamente.

Tópicos como a otimização local e o hibridismo podem ser mais explorados por meio de interações com outras áreas de pesquisa. Em [Meisel and Mattfeld 2007], por exemplo, descrevem-se intersecções entre as áreas de mineração de dados e programação matemática, incluindo aplicações de metaheurísticas para o problema de SA.

Agradecimentos À UFABC e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio recebido para a realização deste trabalho.

Referências

- Andersson, J. (2000). A survey of multiobjective optimization in engineering design. Technical report, Linkoping University - Sweden.
- Causton, H. C., Quackenbush, J., and Brazm, A. (2003). *Microarray gene expression data analysis: a beginner's guide*. Wiley-Blackwell.
- Charikar, M., Guruswami, V., Kumar, R., Rajagopalan, S., and Sahai, A. (2000). Combinatorial feature selection problems. In *Proc. Annual Symp. Foundations Computer Science (FOCS '00)*, pages 631–640.
- Coello, C. A. C. (2006). Evolutionary multi-objective optimization: a historical view of the field. *Comp. Intellig. Magazine, IEEE*, pages 28–36.
- Cristianini, N. and Shawe-Taylor, J. (2000). *Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge University Press.

- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Trans. Evolut. Comp.*, pages 182–197.
- Dessì, N. and Pes, B. (2009). An evolutionary method for combining different feature selection criteria in microarray data classification. *J. Artif. Evol. and Applic.*, pages 1–10.
- Dy, J. G., Brodley, C. E., and Wrobel, S. (2004). Feature selection for unsupervised learning. *J. Machine Learning Res.*, pages 845–889.
- Emmanouilidis, C. (2002). Evolutionary multiobjective feature selection and roc analysis with application to industrial machinery fault diagnosis. In *Evolut. Methods for Design Optimisation and Control*, pages 1–6.
- Emmanouilidis, C., Hunter, A., and MacIntyre, J. (2000). A multiobjective evolutionary setting for feature selection and a commonality-based crossover operator. In *Proc. Congr. Evolut. Comp. (CEC00)*, pages 309–316.
- Everitt, B. S., Landau, S., and Leese, M. (2001). *Cluster analysis*. A Hodder Arnold.
- Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloq. Mathemat.*, pages 282–285.
- Freitas, A. A. (2004). A critical review of multi-objective optimization in data mining: a position paper. *SIGKDD Explor. Newsl.*, pages 77–86.
- Geng, X., Liu, T.-Y., Qin, T., and Li, H. (2007). Feature selection for ranking. In *Proc. Annual Int. ACM SIGIR Conf. Res. and Develop. in Inform. Retrieval*, pages 407–414.
- Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. M. Kaufmann.
- Handl, J., Kell, D. B., and Knowles, J. (2007). Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Trans. Comp. Biol. and Bioinf.*, pages 279–292.
- Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Ishibuchi, H., Hitotsuyanagi, Y., and Nojima, Y. (2008). Scalability of multiobjective genetic local search to many-objective problems: Knapsack problem case studies. In *Proc. IEEE Congr. Evolut. Comp.*, pages 3586–3593.
- Kim, Y., Street, W. N., and Menczer, F. (2002). Evolutionary model selection in unsupervised learning. *Intell. Data Anal.*, pages 531–556.
- Liu, H. and Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers.
- Liu, H. and Motoda, H. (2008a). *Computational Methods of Feature Selection*. Chapman & Hall/CRC.
- Liu, H. and Motoda, H. (2008b). Less is more. In *Computational Methods Feature Selection*, pages 3–17. Chapman & Hall/CRC.
- Liu, H. and Yu, L. (2002). Feature selection for data mining. <http://www.public.asu.edu/~huanliu/sur-fs02.ps>.
- MacQueen, J. B. (1965). On convergence of k -means and partitions with minimum average variance. *Mathematical Statistics*, pages 1084–1084.

- Meisel, S. and Mattfeld, D. C. (2007). Synergies of data mining and operations research. In *Proc. Annual Hawaii Int. Conf. Syst. Sciences*, pages 1–10.
- Mierswa, I. (2009). *Non-Convex and Multi-Objective Optimization in Data Mining*. Phd thesis, University of Dortmund - Germany.
- Mierswa, I. and Wurst, M. (2006). Information preserving multi-objective feature selection for unsupervised learning. In *Proc. Genetic and Evolut. Comp. Conf.*, pages 1545–1552.
- Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT Press.
- Oliveira, L. S., Morita, M., and Sabourin, R. (2006). Feature selection for ensembles using the multi-objective optimization approach. In *Multi-Objective Machine Learning*, pages 49–74. Springer-Verlag.
- Oliveira, L. S., Sabourin, R., Bortolozzi, F., and Suen, C. Y. (2003). A methodology for feature selection using multiobjective genetic algorithms for handwritten digit string recognition. *Int. J. Pattern Recognition and Artif. Intellig.*, pages 903–929.
- Ong, Y.-S. and Keane, A. J. (2004). Meta-lamarckian in memetic algorithm. *IEEE Trans. Evolut. Comp.*, pages 99–110.
- Pareto, V. (1896). *Cours d’Economie Politique*. Droz.
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. M. Kaufmann.
- Roy, K. and Bhattacharya, P. (2008). Optimal features subset selection and classification for iris recognition. *J. Image and Video Processing*, pages 1–20.
- Sánchez, L., Villar, J. R., and Couso, I. (2008). Genetic feature selection for fuzzy discretized data. In *Proc. Inform. Processing and Managem. Uncertainly in Knowl.-Based Syst.*, pages 1159–1166.
- Srinivas, N. and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolut. Comp.*, pages 221–248.
- Suguna, N. and Thanushkodi, K. (2009). Genetic algorithm based feature ranking in multi-criteria optimization. *Int. J. Computer Science and Netw. Sec.*, pages 132–141.
- Tapia, M. G. C. and Coello, C. A. C. (2007). Applications of multi-objective evolutionary algorithms in economics and finance: A survey. In *Proc. IEEE Congr. Evolut. Comp.*, pages 532–539.
- Yang, J. and Honavar, V. (1998). Feature subset selection using a genetic algorithm. In *Proc. IEEE Intellig. Syst. and their Applications*, pages 44–49.
- Yu, L. and Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *J. Machine Learning Res.*, pages 1205–1224.
- Zhu, Z., Ong, Y.-S., and Kuo, J.-L. (2009). Feature selection using single/multi-objective memetic frameworks. In *Multi-Objective Memetic Algorithms*, pages 111–131. Springer-Verlag.
- Zhu, Z., Ong, Y.-S., and Zurada, M. (2008). Simultaneous identification of full class relevant and partial class relevant genes. *IEEE/ACM Trans. Comp. Biol. and Bioinf.* (Under Review).