

# Uma revisão de abordagens genético-difusas para descoberta de conhecimento em banco de dados

## A review of genetic-fuzzy approaches to knowledge discovery in databases

Wesley Romão<sup>1\*</sup>, Alex A. Freitas<sup>2</sup> e Roberto C. S. Pacheco<sup>3</sup>

<sup>1</sup> UFSC, PPGE e Universidade Estadual de Maringá, Departamento de Informática. <http://www.din.uem.br/~wesley>

\* Autor correspondente: Av. Morangueira, 1075, Apto. 2-502, CEP 87033-070 - Maringá, PR, Brasil. [romao@eps.ufsc.br](mailto:romao@eps.ufsc.br)

<sup>2</sup> PUC-PR, PPGIA-CCET. <http://www.ppgia.pucpr.br/~alex> [alex@ppgia.pucpr.br](mailto:alex@ppgia.pucpr.br)  
Rua Imaculada Conceição, 1155. CEP 80215-901 – Curitiba, PR, Brazil.

<sup>3</sup> UFSC, PPGE. C. P. 476 - CEP 88040-900 - Florianópolis, Brasil. [pacheco@eps.ufsc.br](mailto:pacheco@eps.ufsc.br)

**RESUMO:** O processo geral de descoberta de conhecimento em banco de dados é composto por diversas etapas destacando-se a etapa de Mineração de Dados (MD). As principais tarefas de MD são associação, agrupamento e descoberta de regras de classificação. A tarefa de classificação pode ser realizada por algoritmos convencionais (*e.g.*, estatísticos) ou por métodos de inteligência artificial (*e.g.*, redes neurais, algoritmos evolucionários, etc.). Nesta pesquisa o interesse é revisar algumas abordagens que utilizam algoritmos genéticos (AG) em combinação com conjuntos difusos (CD) de forma híbrida para realizar busca em espaços grandes e complexos. Este artigo mostra diversas abordagens híbridas (AG+CD), desenvolvidas para descoberta de regras de classificação, disponíveis na literatura e indica a possibilidade de adaptação das mesmas na descoberta de conhecimento em banco de dados de Ciência e Tecnologia (C&T).

**Palavras-chaves:** algoritmos genéticos; conjuntos difusos; descoberta de conhecimento.

**ABSTRACT:** The Knowledge Discovery in Databases (KDD) process consists of many stages where the Data Mining (DM) is the main one. There are many DM tasks but the discovery of classification rules is the most known. The classification task can be addressed by conventional algorithms (*e.g.*, statistics) or by artificial intelligence techniques (*e.g.*, neural networks, evolutionary algorithms, etc.). In this research we are interested in investigating the applicability of a hybrid combination of genetic algorithms and fuzzy sets to find rules in large and complex spaces. This paper reviews some hybrid Genetic-Fuzzy approaches for the extraction of classification rules found in the literature and discuss the possibility of adapting them to knowledge discovery in Science and Technology (S&T) databases.

**Key words:** fuzzy sets; genetic algorithms; knowledge discovery.

### Introdução

A grande quantidade de informações nos bancos de dados informatizados das organizações pode esconder conhecimentos valiosos e úteis para a tomada de decisão. O aumento no volume dos dados, associado à crescente demanda por conhecimento novo voltado para decisões estratégicas, tem provocado o interesse crescente em descobrir conhecimentos em banco de dados.

Uma área interdisciplinar específica, KDD (Knowledge Discovery in Databases – Descoberta de Conhecimento em Banco de Dados), surgiu em resposta à necessidade de novas abordagens e soluções para viabilizar a análise de grandes

bancos de dados. Particularmente, KDD tem obtido sucesso na área de marketing, onde a análise de banco de dados de clientes revela padrões de comportamento e preferências que facilitam a definição de estratégias de vendas. A empresa American Express, por exemplo, fez aumentar as vendas de cartão de crédito em 15 a 20% com a utilização de marketing auxiliado por técnicas de KDD (Berry, 1994; In: Fayyad *et al.*, 1996a).

Uma seqüência natural no processo de KDD é:

Dados → Conhecimento → Decisão

Para extrair conhecimento, são necessárias ferramentas de exploração, hoje conhecidas como

*Mineração de Dados* (MD), que podem incorporar técnicas estatísticas e/ou de inteligência artificial, capazes de fornecer respostas a várias questões ou mesmo de descobrir novos conhecimentos em grandes bancos de dados. MD é especialmente útil em casos onde não se conhece a pergunta, mas, mesmo assim, existe a necessidade de respostas.

Um dos primeiros passos em um processo de KDD é a definição da tarefa de MD a ser realizada, que determina o tipo de conhecimento a ser descoberto.

Nesta pesquisa, o primeiro estudo empírico realizado envolveu a tarefa de descoberta de regras de associação; ou seja, aplicou-se um algoritmo padrão de extração de regras de associação (Apriori) à base de dados do *Directorio dos Grupos de Pesquisa no Brasil* (CNPq, 1999), versão 3.0. Os resultados daquele experimento serviram para motivar o aprofundamento dos estudos sobre técnicas de MD aplicáveis neste domínio (Romão *et al.*, 1999b).

No entanto, na fase atual desta pesquisa o interesse é investigar técnicas voltadas à descoberta de regras de previsão. Para isso, segundo Freitas (Freitas, 2000) é necessário utilizar a tarefa de classificação ou modelagem de dependência – ou outra tarefa relacionada à aprendizagem de máquina - mas não à tarefa de associação padrão.

Neste artigo, o objetivo é investigar a possibilidade da aplicação de algumas abordagens genético-difusas (*i.e.*, técnicas fundamentadas na combinação de Algoritmos Genéticos (AG's) e Conjuntos Difusos), para extração de regras de previsão no contexto de ciência e tecnologia (C&T). Nos segmentos responsáveis pelo fomento a C&T, dados representam informações sobre pesquisadores, grupos, projetos e instituições de pesquisa. O desafio é transformar estas informações em subsídios para esta mesma comunidade. Neste contexto, o desafio que se apresenta pode ser simplificado como a resolução da seguinte questão básica: *Como extrair conhecimento destes dados?*

Muitas das técnicas tradicionais de MD têm sido aplicadas com êxito e outras esbarram em limitações, tanto no desempenho como na qualidade do conhecimento gerado. Pesquisas recentes têm demonstrado que técnicas da área de IA, tais como AG e Conjuntos Difusos, podem ser utilizadas com sucesso.

Este artigo está organizado da seguinte forma: as próximas seções apresentam uma visão geral

sobre o processo de descoberta de conhecimento, mineração de dados e a tarefa de classificação, respectivamente. A seguir, é apresentada uma visão geral de algoritmos genéticos (AG's) para a tarefa de classificação, com foco em AG's que descobrem regras de previsão. Na sequência, é apresentada uma visão geral de conjuntos difusos. As seções mencionadas acima têm o objetivo de fornecer ao leitor subsídios para o entendimento das seções seguintes, que constituem a principal contribuição deste trabalho. Mais precisamente, as seções seguintes discutem vários sistemas híbridos genético-difusos para descoberta de regras de previsão. A discussão é apresentada com foco na aplicabilidade desses sistemas à descoberta de conhecimento em ciência e tecnologia.

## O Processo de Descoberta de Conhecimento

*“KDD é o processo não trivial de identificação de padrões, a partir de dados, que sejam válidos, novos, potencialmente úteis e compreensíveis”* (Fayyad, 1996b).

Na definição de Fayyad, KDD é descrito como um processo geral de descoberta de conhecimento composto por várias etapas, incluindo: preparação dos dados, busca de padrões, avaliação do conhecimento e refinamentos.

Os padrões devem ser novos, compreensíveis e úteis, ou seja, deverão trazer algum benefício novo que possa ser compreendido rapidamente pelo usuário para tomada de decisão.

Para descobrir conhecimento que seja relevante, é importante estabelecer metas bem definidas. Segundo Fayyad *et al.* (1996b), no processo de descoberta de conhecimento as metas são definidas em função dos objetivos na utilização do sistema, podendo ser de dois tipos básicos: *verificação* ou *descoberta*.

Quando a meta é do tipo *verificação*, o sistema está limitado a verificar hipóteses definidas pelo usuário, enquanto que na *descoberta* o sistema encontra novos padrões de forma autônoma. A meta do tipo *descoberta* pode ser subdividida em: *previsão* e *descrição*, conforme a Figura 1.

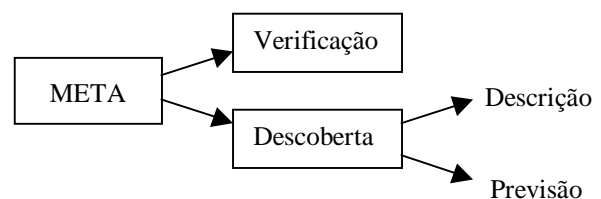


Figura 1. Tipos de metas no processo de KDD.

A *descrição* procura encontrar padrões,

interpretáveis pelos usuários, que descrevam os dados. A *previsão* parte de diversas variáveis para prever outras variáveis ou valores desconhecidos (Fayyad *et al.*, 1996a).

As metas de previsão e descrição são alcançadas através de alguma das seguintes tarefas de MD: classificação, regressão, agrupamento, sumarização, modelagem de dependência e identificação de mudanças e desvios, sendo a tarefa de classificação a mais empregada.

Um trabalho relevante no contexto de C&T, envolvendo a tarefa de agrupamento, foi desenvolvido recentemente por Gonçalves (Gonçalves, 2000), o qual efetuou a extração de conhecimento sobre C&T no Diretório 3.0 (CNPq, 1999). O conhecimento extraído com base em técnicas de MD, como neste trabalho, deve ser avaliado quanto à relevância para seu contexto de decisão. Neste artigo, o interesse é determinar a viabilidade de extração de conhecimento, preservando-se a noção de sua usabilidade de contexto.

Apesar da MD ser a etapa principal, o processo de descoberta de conhecimento em banco de dados, não se resume a minerar os dados. Exige-se a construção de mais dois estágios: pré-processamento e pós-processamento, conforme ilustra a Figura 2.

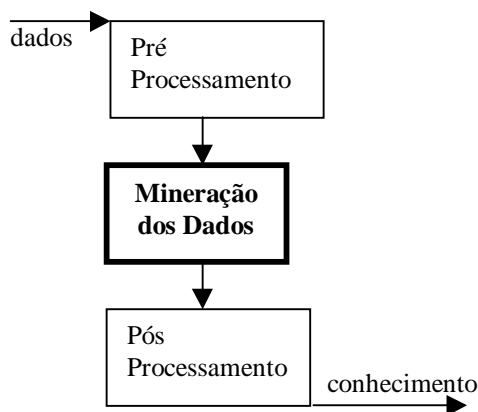


Figura 2. Etapas básicas para descoberta de conhecimento.

Portanto, o processo de KDD utiliza banco de dados para realizar: seleção de atributos e transformações necessárias sobre os dados (pré-processamento); aplicação de métodos (algoritmos) de MD para extrair padrões dos dados; e avaliação do produto da MD para identificar os padrões julgados como “conhecimento” (pós-processamento).

Dentre as etapas mencionadas acima, um dos maiores desafios é a seleção de atributos

relevantes para realizar uma tarefa de MD específica. Existem duas abordagens principais utilizadas para este fim: *Processo Envolvório* (Wrapper) e *Processo por Filtro*. Em geral, as técnicas do tipo envoltório tendem a ser mais efetivas, ou seja, resultam em uma menor taxa de erro de classificação se comparadas com as técnicas do tipo filtro, mas estas últimas normalmente são mais eficientes, uma vez que consomem menor tempo de processamento.

O pós-processamento é utilizado principalmente para avaliar o processo de descoberta, melhorar a compreensão e/ou selecionar conhecimento descoberto que seja mais interessante. Quando são geradas muitas regras, é importante remover algumas regras e/ou condições para facilitar a compreensão do conhecimento extraído.

Existem diversas abordagens para avaliar o processo de descoberta de conhecimento, incluindo-se: exatidão dos resultados (*e.g.*, alguma medida da taxa de acerto), eficiência (tempo de processamento), facilidade de compreensão do conhecimento extraído, etc.. A maior parte da literatura utiliza exatidão (taxa de acerto) como principal meio para avaliar as técnicas de KDD (Freitas, 1997), principalmente no contexto da tarefa de classificação.

## Mineração de Dados

Técnicas de MD utilizam dados históricos para aprendizagem objetivando realizar alguma tarefa particular. Esta tarefa tem como meta responder alguma pergunta específica de interesse do usuário. Portanto, é necessário informar qual problema se deseja resolver.

*“Não há um método de Mineração de Dados ‘universal’ e a escolha de um algoritmo particular para uma aplicação particular é de certa forma uma arte (Fayyad et al., p. 86, 1996b).”*

Além da área de negócios, MD tem sido também utilizado na área científica (*e.g.*, biologia molecular, modelagem de mudanças climáticas globais, etc.).

Para encontrar respostas, ou extrair conhecimento interessante, existem diversas técnicas de MD disponíveis na literatura (Chen *et al.*, 1996; Cheung *et al.*, 1996). As principais podem ser agrupadas em:

- Indução e/ou Extração de Regras;
- Redes Neurais;
- Algoritmos Evolucionários;

- Técnicas estatísticas (classificadores e redes Bayesianas, etc.)
- Conjuntos Difusos (“Fuzzy”), etc.

Para a escolha da técnica mais adequada é estratégico saber alguma coisa a respeito do domínio da aplicação de MD: quais são os atributos importantes, quais os relacionamentos possíveis, o que é uma função útil para o usuário, que padrões já são conhecidos e assim por diante.

Na seção seguinte, discute-se a principal tarefa de mineração de dados.

### A Tarefa de Classificação

Classificação é uma das tarefas mais referenciadas na literatura de MD e a mais importante para a presente pesquisa. Neste tipo de tarefa, o objetivo é descobrir um relacionamento entre um atributo meta (cujo valor, ou classe, será previsto) e um conjunto de atributos de previsão. O sistema deve descobrir este relacionamento a partir de exemplos com classe conhecida. O relacionamento descoberto será usado para prever o valor do atributo meta (ou a classe) para exemplos cujas classes são desconhecidas (Fertig *et al.*, 1999).

Na área de aplicação considerada (Gestão de C&T), pode-se definir classificação como sendo a tarefa de prever corretamente a classe de um exemplo da unidade de análise (*i.e.*, *pesquisador*, *grupo de pesquisa*, *projeto de pesquisa*, *etc.*), a partir de alguns atributos da unidade de análise, chamados *atributos previsores*, cujos valores são conhecidos. Uma das possibilidades é a descoberta de regras que representem as correlações entre os atributos que definem a unidade de análise (*e.g.*, *sexo*, *idade e titulação para pesquisadores*).

A literatura apresenta diversas técnicas de classificação (King *et al.*, 1995; Hand, 1997). Segundo Michie *et al.* (1994), as principais propostas são originárias de três campos de pesquisa: estatística, aprendizagem de máquina simbólica e redes neurais.

Neste trabalho, o interesse é investigar principalmente o campo chamado de *aprendizagem de máquina simbólica*, que abrange procedimentos computacionais automáticos, baseados em operações lógicas ou binárias, capazes de aprender a realizar uma tarefa a partir de uma série de exemplos. Focalizando a tarefa de classificação, muita atenção tem sido dada a técnicas baseadas em árvores de decisão (Quinlan, 1993; *etc.*). Outras técnicas, tais como AG's e PLI (Programação Lógica Indutiva) têm sido alvo de

mais interesse por parte de pesquisadores recentemente. As técnicas de aprendizagem de máquina simbólica para a tarefa de classificação possuem a vantagem de gerar expressões simples o suficiente para a compreensão humana (Michie *et al.*, 1994).

Cabe ressaltar que nenhum algoritmo é “o melhor” em todas as aplicações. A performance de um algoritmo de classificação depende muito do domínio da aplicação (Freitas, 2000).

### Estrutura Geral da Tarefa de Classificação

No escopo desta pesquisa, assume-se que o problema é projetar um procedimento para ser aplicado em um banco de dados onde as classes são pré-definidas e cada novo dado deve ser associado a uma destas classes. Este processo é conhecido como *reconhecimento de padrões*, *discriminação*, *aprendizagem supervisionada* ou *classificação*. Na literatura de estatística, a aprendizagem supervisionada usualmente é referenciada como *discriminação*.

Existem diversas formas de representar o conhecimento em um sistema de aprendizagem. No contexto da tarefa de classificação, o conhecimento descoberto muitas vezes é expresso como um conjunto de regras de classificação do tipo SE-ENTÃO, uma vez que este tipo de representação do conhecimento é intuitivo para o usuário (Carvalho & Freitas, 2000).

Regras do tipo SE-ENTÃO são também chamadas *regras de produção*, constituem uma forma de representação simbólica e possuem a seguinte forma:

SE <antecedente> ENTÃO <conseqüente>

O *antecedente* é formado por expressões condicionais envolvendo atributos do domínio da aplicação existentes nos bancos de dados.

O *conseqüente* é formado por expressões que indicam a previsão de algum valor para um atributo meta, obtido em função dos valores encontrados nos atributos que compõem o antecedente.

Portanto, a tarefa é descobrir regras de classificação capazes de prever o valor de um atributo meta a partir dos valores de atributos previsores. As regras de previsão, portanto, objetivam auxiliar o planejamento de ações futuras.

Como ilustração, considere-se um banco de dados sobre C&T cujo atributo meta, escolhido por um especialista de uma agência, é uma

determinada região do país (*e.g.*, sul, sudeste, etc.) candidata a receber investimentos para pesquisa, associada às classes “bom” ou “ruim”, e considere-se, ainda, que os atributos previsores são: produção científica, idade e titulação dos pesquisadores.

Seguindo esta ilustração, pode-se obter regras como:

SE (prod\_científ. = “alta”) E (Idade = “média”)  
ENTÃO (Região.Pot\_Invest. = “boa”)

SE (titulação = “baixa”)  
ENTÃO (Região.Pot\_Invest. = “ruim”).

### Algoritmos de Classificação Convencionais

Existem vários métodos já consolidados na tarefa de classificação. Uma revisão geral sobre este tema pode ser encontrada em (Han & Kamber, 2000).

Pesquisas comparando diversos algoritmos de classificação convencionais foram publicadas por (King *et al.*, 1995). Estes trabalhos descrevem resultados de um projeto (denominado de Statlog) que compara 17 (dezesete) algoritmos de classificação. King e seus colegas utilizaram 12 (doze) conjuntos de dados: 5 (cinco) de análise de imagens; 3 (três) de medicina; 2 (dois) de engenharia e 2 (dois) de finanças. Foram avaliados 7 (sete) algoritmos de aprendizagem simbólica, sendo cinco árvores de decisão e dois métodos baseados em regras; 6 (seis) da estatística tradicional; 3 (três) da estatística moderna e 2 (dois) RNA (Redes Neurais Artificiais).

Os autores concluíram que o melhor algoritmo para classificar um conjunto particular de dados depende primordialmente das características dos dados utilizados. Uma das principais conclusões foi que dados com distribuição muito desbalanceada e muitos atributos binários/categóricos favorecem a utilização de algoritmos de aprendizagem simbólica.

O projeto StatLog gerou resultados de pesquisa, na avaliação de algoritmos de classificação, que culminaram na publicação de um livro (Michie *et al.*, 1994).

Portanto, considerando a análise acima, conclui-se que os algoritmos simbólicos seriam adequados para a tarefa de classificação dos dados sobre C&T. Entretanto, os algoritmos simbólicos tradicionais apresentam algumas limitações, principalmente quando ocorre interação entre atributos, como é o caso nos dados sobre C&T. Uma alternativa seria a utilização de AG's.

### Algoritmos Genéticos

Os algoritmos genéticos (AG's) são algoritmos de busca e otimização baseados na analogia com os processos de seleção natural e genética evolucionária (Goldberg, 1989). A essência do método consiste em manter uma população de indivíduos (cromossomos), os quais representam possíveis soluções para um problema específico. A melhor solução é atingida através de um processo de seleção competitiva, envolvendo cruzamentos e mutações (Herrera *et al.*, 1996).

A constituição hereditária de um indivíduo é dada pelo seu *genótipo*, enquanto que *fenótipo* é o nome dado àqueles que têm o mesmo aspecto geral de outros da mesma espécie, diferindo dela apenas por certos caracteres exteriores resultantes de condições mesológicas (meio ambiente).

AG simula uma população genética organizada, composta por agentes individuais que competem entre si para sobreviver e cooperam para alcançar uma adaptação melhor. Estes agentes também são chamados de cromossomos ou indivíduos. Similares aos cromossomos originais, os indivíduos de um AG são compostos por genes (genótipo). O significado de um indivíduo é definido externamente pelo usuário e fornece a solução aproximada de um problema específico.

Os AG's utilizam dois mecanismos adaptativos: seleção e herança. A seleção, ou competição, é um processo estocástico de sobrevivência de um indivíduo proporcional à sua capacidade de adaptação. A adaptação é medida através da avaliação do *fenótipo* no ambiente em que está inserido, ou seja, sua capacidade de resolver o problema. Esta seleção promove a sobrevivência dos melhores indivíduos (melhores soluções) que irão gerar descendentes (novas soluções, baseadas nas soluções “pais”).

A cooperação é obtida através de *crossover* (cruzamento) do material genético entre dois indivíduos selecionados, com a expectativa de produzir indivíduos melhor adaptados ou melhores soluções.

Além de cruzamento, pode-se empregar *mutação*, em menor proporção, para introduzir variação adicional. A mutação possibilita a geração de genes não presentes na população corrente, aumentando a robustez do algoritmo.

Os AG's têm sido utilizados com sucesso na otimização de parâmetros em diversos tipos de aplicações, como pode ser visto em Romão (1999a) e Janikow (1995).

Na literatura existem diversas abordagens desta natureza. Os operadores genéticos de reprodução, crossover e mutação são apresentados como poderosos mecanismos para descoberta de regras interessantes em aplicações como otimização (Janikow, 1995), aproximação de funções (Delgado *et al.*, 1999), obtenção de resumos (Kacprzyk & Strykowski, 1999), diagnóstico médico (Peña-Reyes & Sipper, 1999), robótica (Deb *et al.*, 1998) e, mais especificamente, classificação (Lee, 1998; Mota *et al.*, 1999; Xiong & Litz, 1999).

**Pittsburgh x Michigan.** Para utilização de AG's na tarefa de classificação, existem duas abordagens básicas que receberam os nomes das universidades de Pittsburgh e Michigan onde foram desenvolvidas.

Na primeira abordagem, conhecida como Pittsburgh, a codificação é complexa (Figura 3), implicando em operadores mais complexos, uma vez que cada indivíduo irá corresponder a um conjunto de regras solução para o problema.

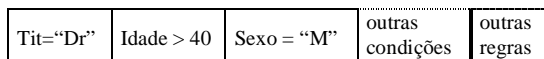


Figura 3. Indivíduo representando um conjunto de regras.

Na segunda abordagem, Michigan, cada regra é representada por um indivíduo, sendo que um conjunto de regras (uma solução do problema) é representada por um conjunto de indivíduos.

A abordagem de Michigan facilita a codificação dos indivíduos (ver Figura 3) permitindo a construção de indivíduos simples e pequenos, , mas apresenta o problema de que é mais difícil lidar com interações entre regras.

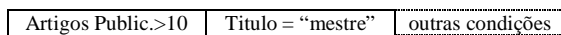


Figura 4. Indivíduo representando uma regra.

**Codificação do Indivíduo.** Quando o assunto é MD, há necessidade de algumas considerações no projeto dos AG's. A representação do indivíduo, os operadores genéticos e as funções de aptidão devem ser adaptadas para extração de conhecimento de alto nível compreensível para o usuário.

Se o AG é aplicado à tarefa de classificação, os indivíduos podem representar apenas os antecedentes das regras de classificação, desde que os consequentes das regras sejam determinados por algum critério pré-definido. Sendo todos os atributos categóricos (os valores contínuos seriam discretizados), pode-se utilizar

codificação binária, onde o número de bits de determinado atributo seja igual à quantidade de valores possíveis para este atributo. Nesse caso, em cada condição mais de um bit pode estar "ligado" (tendo valor 1), indicando uma disjunção implícita dos valores do respectivo atributo. Por exemplo, o valor "0 0 1 1 0" pode representar a condição: SE (titulação = "mestre" OU "doutor"), onde o atributo *titulação* assume os valores "graduado", "especialista", "mestre", "doutor" ou "pós-doutor".

Este esquema pode ainda ser estendido para representar regras com várias condições, ligadas implicitamente pelo operador de conjunção, através da simples inclusão de mais bits no genoma.

Outra alternativa é representar o genoma codificando as condições das regras diretamente em uma linguagem de alto nível.

Seja qual for a forma de codificação, em geral é necessário utilizar uma estrutura de dados que permita variar o tamanho do vetor que representa o indivíduo, uma vez que não se sabe previamente quantas condições serão produzidas em cada regra.

Em AG, utilizando a abordagem de Michigan, cada indivíduo pode representar uma regra candidata para solução de um dado problema (por exemplo, classificação), e o indivíduo é composto por diversos genes, conforme Figura 5, onde cada gene pode representar uma condição da regra.

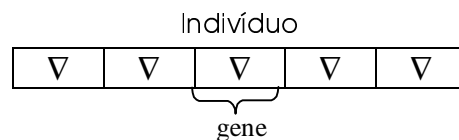


Figura 5. Exemplo de indivíduo com 5 genes.

**Operadores Genéticos.** A escolha dos operadores genéticos, juntamente com a determinação da função de *fitness* (aptidão) e da apropriada representação dos indivíduos, é determinante para o sucesso de um AG. Os operadores genéticos de seleção, crossover e mutação fornecem o mecanismo básico de busca e são usados para criar novas soluções baseadas nas melhores soluções existentes na população atual do AG.

**Função de Aptidão (Fitness).** Todo AG exige a definição de uma função de aptidão, denominada função de Fitness. Ela é responsável em fornecer um valor que irá indicar a qualidade do indivíduo (solução candidata) avaliado. Esta função é específica para cada aplicação do AG e portanto

deve ser definida para cada tipo de problema tratado. No contexto de mineração em banco de dados de grande porte, o maior tempo de processamento dos AG's é gasto na computação da função de Fitness (Freitas, 2001).

**Operador de Seleção.** A partir dos resultados da função de Fitness é possível comparar diversos indivíduos e escolher os melhores através do operador de Seleção.

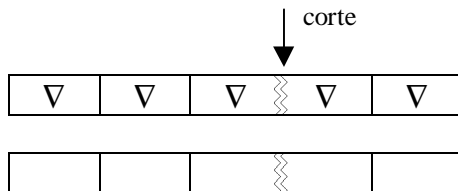


Figura 6. Exemplo de dois indivíduos selecionados.

Este operador é responsável em selecionar dois indivíduos, dentre aqueles que obtiveram os melhores valores da função de Fitness, para serem usados pelos operadores de crossover e mutação. Supondo que os indivíduos sejam formados por cinco genes, os dois indivíduos selecionados teriam a estrutura mostrada na Figura 6.

**Operador de Crossover.** Com os dois indivíduos selecionados é possível aplicar o operador de crossover, o qual realiza a troca de material genético entre estes indivíduos selecionados para reprodução. Para isto, aleatoriamente é escolhido um ponto entre dois genes onde se efetua um corte, conforme mostra a Figura 6. Supondo que o ponto sorteado para o corte seja entre o terceiro e quarto genes, o primeiro indivíduo receberia os genes 4 e 5 do segundo indivíduo e vice versa. Após o crossover, os indivíduos ficariam conforme mostra a Figura 7.

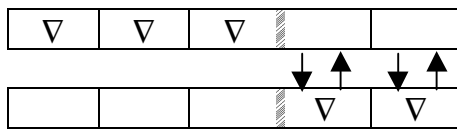


Figura 7. Indivíduos após crossover.

Na tarefa de classificação, o operador de crossover deve ser adaptado para trabalhar com indivíduos de tamanho variável. Além disso, para evitar a criação de filhos “deformados”, o indivíduo deverá ser programado para manter, na sua representação interna, a mesma ordem dos atributos que se encontra no conjunto de dados de treinamento.

**Operador de Mutação.** O operador de mutação pode ser aplicado individualmente sobre cada indivíduo selecionado, realizando a alteração de

algum gene escolhido aleatoriamente. Este operador permite que novas combinações de genes, que ainda não existam dentro dos indivíduos da população, sejam viabilizadas.

Tanto o operador de crossover como o de mutação são aplicados conforme probabilidade definida pelo usuário, sendo a probabilidade de crossover normalmente definida bem maior do que a probabilidade de mutação.

## Conjuntos Difusos

A lógica binária comum trabalha com dois valores: verdadeiro ou falso. No entanto, no mundo real nem sempre esta representação corresponde à realidade, pois em geral proposições são verdadeiras com um certo grau de veracidade. Conjuntos Difusos, juntamente com a Lógica Difusa e seu raciocínio aproximado, são uma alternativa para contornar estes problemas. Eles podem representar informação de acordo com seu grau de verdade e permitem codificar expressões do tipo “mais ou menos”, “aproximadamente”, “quase”, “pouco”, “muito”, etc..

Na teoria de conjuntos difusos uma mesma informação pode ser representada por mais de um conjunto difuso. Por exemplo, um dado indicando que a temperatura foi medida como sendo de 25°C pode simultaneamente ser expresso como um “pouco frio” e um “pouco quente”. Esta dualidade da informação reflete-se nas funções de pertinência ( $\mu$ ) de cada temperatura em cada conjunto. Os graus de pertinência indicam o quanto uma temperatura corresponde a cada conjunto. As formas mais populares de funções de pertinência são a trapezoidal e a triangular.

Quando os conjuntos difusos representam palavras da linguagem natural, a variável mapeada no conjunto é chamada *variável lingüística* (no exemplo acima, temperatura é uma variável lingüística).

A Figura 8 ilustra esta interação. Neste exemplo diz-se que a temperatura está mais fria do que quente, ou seja, 0.3 quente e 0.7 fria. Em geral, a soma dos valores das funções de pertinência (FP) deve ser sempre igual a 1.

Para a criação de expressões lógicas, de forma análoga aos conjuntos clássicos, a Teoria dos Conjuntos Difusos forma a base científica para a *Lógica Difusa*. Nesta, uma implicação de fatos difusos é representada por regras difusas. Os sistemas de lógica difusa, em geral, são mapeamentos não lineares de um vetor de entrada

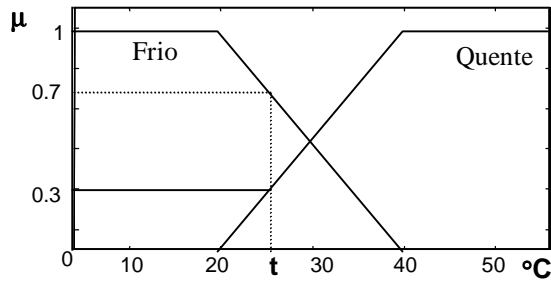


Figura 8. Conjuntos difusos para temperatura.

(dados) em um escalar de saída, formando uma coleção de sistemas independentes de múltiplas entradas/única saída (Mendel, 1995).

A cada variável lingüística corresponde uma série de conjuntos difusos denominados "termos lingüísticos", que descrevem os diversos estados das variáveis lingüísticas e que possuem formas variadas.

A qualidade do sistema difuso depende principalmente da escolha dos seguintes fatores: operadores difusos; tipo das funções de pertinência (FP's); método de defuzzificação (*i.e.*, procedimento de extração do valor mais característico de um conjunto difuso); variáveis relevantes; número de FP's e quantidade de regras. Uma vez construída a base de conhecimento de um sistema difuso, pode-se descrever a essência do método como sendo composta por quatro estágios:

- determinação do grau de pertinência ( $\bullet$ ) no antecedente da regra (fuzzificação). Nesta etapa o valor informado é confrontado com cada conjunto difuso do sistema (ex. temperatura = 23°:  $\bullet_{\text{QUENTE}} = 0.31$  e  $\bullet_{\text{FRIO}} = 0.69$ );
- cálculo dos conseqüentes das regras. Nesta etapa, emprega-se métodos do raciocínio aproximado (implicação) para derivar conclusões a partir da intersecção dos fatos apurados na primeira etapa com os antecedentes das regras difusas;
- agregação dos conseqüentes das regras no conjunto difuso. Nesta etapa, os fatos apurados são conjugados para a construção do conjunto difuso resultante; e
- defuzzificação. Esta etapa ocorre no estágio final quando se necessita retornar o valor típico do conjunto difuso derivado do sistema difuso (e.g., problemas de controle).

O método mais utilizado para agregação é a implicação mínima (Mendel, 1995), dada por:

$$\bullet_{A \rightarrow B}(x,y) = \min[\bullet_A(x), \bullet_B(y)] \quad (1)$$

onde:  $\bullet_{A \rightarrow B}(x,y)$  mede o grau de verdade da relação de implicação entre a entrada  $x$  e a saída  $y$ .

Fertig *et al.* (1999) vêem dois ingredientes que motivam a utilização da abordagem difusa em mineração de dados:

- a lógica difusa é um método flexível e poderoso para tratar com incertezas;
- regras difusas são um meio natural para representar regras com atributos contínuos.

### Sistemas Híbridos Genéticos-Difusos

Um sistema é identificado como híbrido quando incorpora duas ou mais técnicas diferentes em sua estrutura, tais como: AG, redes neurais, lógica difusa, aproveitando a potencialidade de cada uma das técnicas de forma a realizarem um trabalho sinérgico.

Para que um sistema de classificação seja difuso, ele deve encontrar um conjunto de regras com termos lingüísticos onde um objeto desconhecido possa ser classificado através do raciocínio difuso. A forma direta de se obter este sistema é dividir o espaço de exemplos (ou amostras) em uma grade difusa. Entretanto, quando há muitos atributos a combinação aumenta exponencialmente, tornando o problema complexo.

Os AG's aparecem como uma alternativa eficiente para tratar problemas com muitos atributos. Foi provado teórica e empiricamente que AG's são eficientes e robustos para encontrar soluções ótimas ou desejáveis em espaços complexos (Xiong & Litz, 1999).

Aplicações de AG na modelagem de sistemas difusos surgiram por volta de 1990, principalmente em controle de processos, seguidos por aplicações em química, medicina, telecomunicações, biologia e geofísica (Peña-Reyes & Sipper, 1999). Na atualidade, sistemas híbridos têm sido utilizados em muitas aplicações nas mais diferentes áreas.

Focalizando a tarefa de classificação, a combinação de AG com sistemas difusos apresenta qualidades marcantes, como a busca global e a facilidade de compreensão dos resultados, motivando diversos pesquisadores da área de MD a usufruírem destas técnicas.

### Métodos

Nesta pesquisa, o objetivo é discutir a viabilidade de adaptar alguma abordagem genético-difusa para realizar a tarefa de



classificação aplicada à análise de pesquisadores e grupos de pesquisa, considerando a possibilidade desta abordagem ser um instrumento valioso no auxílio ao planejamento de C&T.

Na seqüência, é apresentado um resumo de algumas abordagens genético-difusas analisadas, para extração de regras, onde a maioria apresenta as FP's (funções de pertinência) codificadas no indivíduo juntamente com as regras. Uma exceção foi o trabalho de Delgado *et al.*, discutido a seguir.

Delgado *et al.* (1999) utilizou AG para evoluir uma população de modelos difusos baseados em regras, através de uma abordagem intermediária entre Pittsburgh e Michigan, empregada na aproximação de funções.

Foram utilizados sete termos lingüísticos para representar cada variável numérica e a forma de cada FP foi codificada no gene através de 5 parâmetros: Tp, L, C1, C2 e R, onde Tp é um parâmetro adaptativo que indica o tipo da FP, podendo ser trapezoidal, triangular ou gaussiana. Os demais parâmetros definem a distribuição da FP, conforme Figura 9.

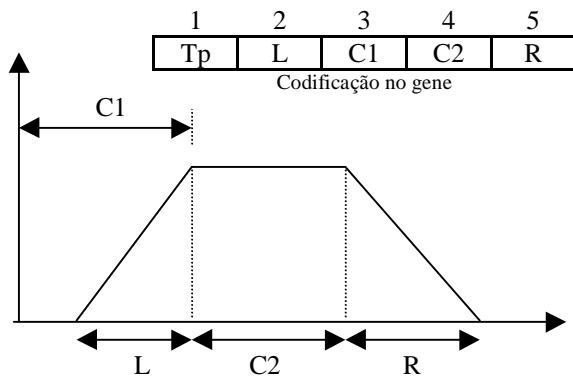


Figura 9. Representação das FP em Delgado *et al.* (1999).

Janikow (1995) apresentou um método, baseado em AG, que otimiza componentes de árvores de decisão difusas, o qual é incorporado na rotina de construção da própria árvore. Foram identificadas diversas restrições usadas para reduzir o espaço de busca. Ele utilizou FP's na forma de trapézios codificados no indivíduo através de 4 parâmetros correspondentes aos 4 vértices do trapézio, conforme Figura 10.

A definição da representação de um indivíduo do AG incluiu restrições nos valores dos parâmetros  $\beta_1 \dots \beta_4$ . Além disso, restrições foram utilizadas para provocar a produção apenas de descendentes viáveis a partir de pais viáveis. O autor demonstrou que esta abordagem garante uma solução possível e melhora a eficiência da

busca, podendo ser estendida para otimização de sistemas baseados em regras.

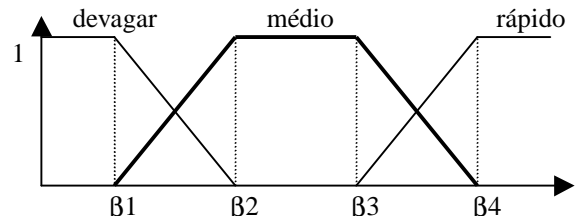


Figura 10. Conjuntos difusos utilizados em Janikow (1995).

Lee (1998) propôs um método de classificação que difere dos demais por otimizar não só as regras e a forma das FP's, mas também a quantidade de FP's, que nos demais métodos é fixa. Além disso, o conseqüente das regras é obtido através do uso do método do gradiente descendente, que minimiza o número de regras resultante.

As FP's, utilizadas na forma triangular, foram representadas por dois parâmetros, a saber: **a** (centro do triângulo) e **b** (largura da base), conforme Figura 11.

A saída do sistema (*y*) é obtida por:

$$y = \frac{\sum_{i=1}^n (\mu_i \cdot \omega_i)}{\sum_{i=1}^n \mu_i} \quad (2)$$

onde:  $\mu_i$  = valor da FP;

$\omega_i$  = valor contínuo do conseqüente.

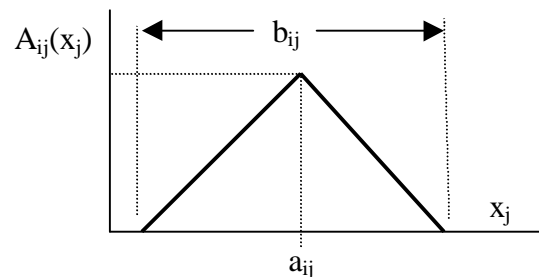


Figura 11. Representação da FP em Lee (1998).

A função de Fitness (*E*) é então obtida por:

$$E = \frac{1}{2} (y - y^p)^2 \quad (3)$$

onde:  $y^p$  = saída desejada.

Mota *et al.* (1999) desenvolveram um sistema difuso de classificação para automatizar um estágio moroso, conhecido no contexto clínico

como *Sleep*, de análise de 2000 páginas de registros poligráficos. O sistema é composto por quatro classificadores independentes e um módulo para integrar suas saídas.

A entrada do sistema, constituída por 12 variáveis, e a saída são representados por três FP's, sendo duas trapezoidais e uma triangular, conforme mostra a Figura 12.

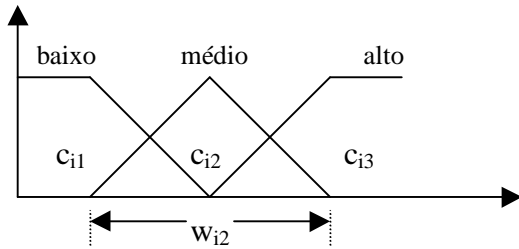


Figura 12. Conjuntos difusos em Mota et al. (1999).

Foi utilizado o método da soma e do produto nas fases de composição e inferência. A defuzzificação foi realizada pela média dos máximos e o número máximo de regras foi limitado em 10.

Cada indivíduo foi formado por dois genomas: um para os parâmetros da FP e outro para as regras. O genoma das FP's foi codificado pelos parâmetros  $c_{ij}$  e  $w_{ij}$  (veja Figura 12), onde:  $c_{ij}$  = centro da FP  $j$  da variável  $i$ ; e  $w_{ij}$  = largura da FP  $j$  da variável  $i$ .

A escolha dos indivíduos foi executada pelo método da roleta (roulette wheel), mantendo-se o elitismo. O crossover foi uniforme e a mutação foi realizada de acordo com uma distribuição normal. Identificou-se a necessidade de eliminar as características de entrada irrelevantes para aumentar o desempenho do AG.

Peña-Reyes & Sipper (1999) mostraram resultados da aplicação de técnicas genético-difusas no diagnóstico do câncer de mama a partir de características de tumores.

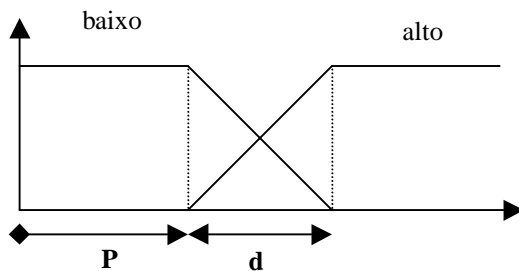


Figura 13. Representação das FP's (Peña-Reyes, 1999).

Os autores seguiram a abordagem de Pittsburgh, utilizando AG para realizar a busca de três parâmetros:  $\mathbf{P}$ ,  $\mathbf{d}$  e  $A_j^i$ , onde os parâmetros  $\mathbf{P}$  e  $\mathbf{d}$  definem a forma e distribuição das FP's, conforme Figura 13, e  $A_j^i$  representa as regras.

Cada indivíduo foi codificado segundo a ilustração da Figura 14.

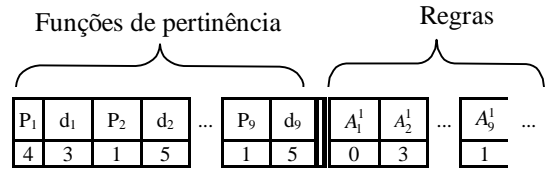


Figura 14. Exemplo de indivíduo contendo apenas uma regra.

Para a avaliação dos resultados foram combinados três critérios:

- a)  $F_c$  = percentagem de casos classificados corretamente;
- b)  $F_e = (\text{valor estimado} - \text{valor correto})^2$ ;
- c)  $F_v = n^\circ$  médio de variáveis/regras ativas.

A função de Fitness foi definida como:

$$F = F_c - \alpha F_v - \beta F_e \tag{4}$$

onde:  $\alpha = 0.05$  e  $\beta = 0.01$  (estimados empiricamente).

Xiong & Litz (1999) apresentaram um método baseado em AG para classificação a partir de dados numéricos. Embora tenham aplicado apenas nos dados sobre a Iris, o método serve para classificação em geral. Eles seguiram a abordagem de Pittsburgh, codificando regras e FP's no mesmo indivíduo, conforme mostra a Figura 15 ( $m = n^\circ$  de atributos;  $k = n^\circ$  de FP's).

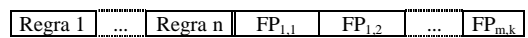


Figura 15. Representação das regras e das FP no indivíduo.

As regras foram codificadas por representação binária. Por exemplo, a regra: SE [ $x_1 =$  (pequeno OU grande)] E [ $x_3 =$  médio] E [ $x_4 =$  (médio OU grande)] foi codificada por: (101; 111; 010; 011). O padrão "111" na segunda condição indica que ela é efetivamente removida da regra (já que o respectivo atributo pode assumir qualquer um de seus valores).

Foram utilizadas FP's triangulares com distribuição determinada pelo parâmetro  $\mathbf{p}$ , conforme mostra a Figura 16.

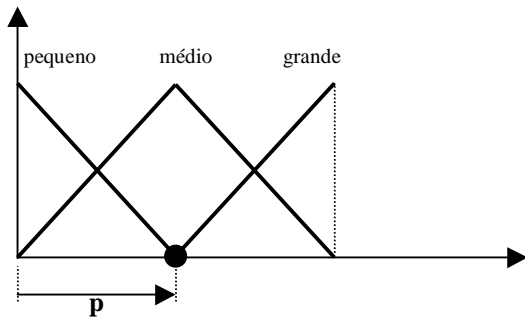


Figura 16. CD definidos em Xiong & Litz (1999).

O parâmetro  $p$  das FP's foi codificado utilizando número inteiro (e não binário como normalmente é codificado), o que permite reduzir o tamanho do indivíduo.

A operação de crossover foi realizada considerando a distinta natureza de cada uma das duas partes do indivíduo (regras e FP's). Para isto, foram aplicados cortes em três pontos: um corte fixo, para separar as duas substrings (uma

contendo as regras, e outra contendo as FP's), e dois cortes correspondentes a um corte em cada substring, caracterizando os pontos reais de corte para cruzamento.

Devido ao mesmo motivo (a diferença entre as duas partes do indivíduo), a operação de mutação também foi realizada de forma diferente em cada parte. Como os valores do parâmetro  $p$  das FP's são contínuos, foi utilizado uma pequena mutação com alta probabilidade e uma perturbação em cada bit gerada por uma função Gaussiana. Na parte das regras, a mutação foi inserida através da simples inversão de bits com pequena probabilidade.

A função de Fitness foi calculada com base na taxa de acerto na classificação dos padrões de treinamento.

Um resumo das abordagens genético-difusas analisadas, com as principais características de cada uma, encontra-se na Tabela 1.

**Tabela 1.** Características de algumas abordagens Genético-Difusas para descoberta de regras.

AUTORES	Objetivo	Tipo dos dados	Nº de FP por variável	Forma da FP	Abordagem	Conteúdo do Indivíduo
Delgado, M. <i>et al.</i> , 1999	Aproximação de funções	Num	7	Trap, Triang, Gauss	Mich-Pitt	Regra + FP FP = (Tp,L, C1,C2,R)
Janikow, C. Z., 1995	Otimização de árvore de decisão difusa	Num	3	Trap	Pitt	Regras + FP FP= 4 vértices dos trapézios
Lee, M-R., 1998	Classificação	Num	variável	Triang	Pitt	Regras + FP FP = (três parâmetros)
Mota, C. <i>et al.</i> , 1999	Classificação	Num	3	Triang ou Trap	Pitt	Regras + FP FP = (centro; largura)
Peña-Reyes, C.A. & Sipper, M., 1999	Diagnóstico médico	Cat	2	Trap	Pitt	Regras + FP = (dois parâmetros)
Xiong, N. & Litz, L., 1999	Classificação	Num	3	Triang	Pitt	Regras + FP FP = único ponto

Legenda: Cat = Categórico; Num = Numérico; Trap = Trapezoidal; Triang = Triangular; Gauss = Gaussiana; Pitt = Pittsburgh; Mich = Michigan.

## Discussão

Nesta seção é apresentada uma visão crítica dos principais pontos positivos e negativos dos métodos genético-difusos discutidos acima. Essa

visão crítica é apresentada no contexto de *mineração de dados*, culminando em sugestões sobre como adaptar os métodos acima para que eles se tornem mais adequados à descoberta de regras de previsão em C&T.

Quase todos os métodos analisados utilizam a abordagem de Pittsburgh (veja 6ª coluna da Tabela 1), mas os dados utilizados para efetuar a extração de regras eram de pequeno volume. Para adaptar alguma abordagem genético-difusa sobre banco de dados de maior porte, como os dados sobre C&T, é recomendável utilizar a abordagem de Michigan para reduzir o espaço de busca e, conseqüentemente, o tempo de processamento.

As restrições dentro do AG, utilizadas por Janikow (1995), também reduzem o espaço de busca e poderão ser empregadas em casos, de extração de regras de previsão, onde se tem conhecimento do domínio, utilizando a experiência de um especialista para determinar restrições para cada atributo. Por exemplo, em alguns dados de C&T, o atributo NÍVEL\_DE\_FORMAÇÃO de um pesquisador possui um domínio contendo 15 (quinze) valores possíveis, mas apenas cinco ou seis destes valores são relevantes para a busca.

Apesar de quase todos os métodos analisados apresentarem as FP's codificadas dentro do indivíduo junto com as regras (veja 7ª coluna da Tabela 1), a otimização destas FP's é pouco significativo perante a importância do processo de evolução das regras. Em casos de banco de dados de maior porte é recomendável otimizar estas FP's em separado, utilizando algum método rápido de busca local, ou utilizar conhecimento do usuário para especificar a distribuição destas funções.

A retirada das FP's de dentro do indivíduo simplifica a construção dos operadores genéticos e melhora o desempenho do AG, uma vez que o espaço de busca se torna menor. Além disso, no caso de extração de regras de previsão, o formato das FP's pode ser fixo, sendo o trapezoidal o mais comum (veja 5ª coluna da Tabela 1). A distribuição dos trapézios pode ser semelhante à utilizada por Janikow onde quatro parâmetros são suficientes para codificar a distribuição de três FP's.

A quantidade de FP's também não precisa necessariamente ser otimizada, como proposta por Lee (1998). Três FP's são suficientes (veja 4ª coluna da Tabela 1) e facilita a compreensão por parte do usuário.

A representação do indivíduo deve ser de mais alto nível possível. Isto aumenta a compreensão das regras e facilita a especificação de restrições para redução do espaço de busca. Uma alternativa seria a utilização de variáveis

linguísticas (*e.g.*: IDADE) com termos simples, tais como, “baixo”, “médio” e “alto”.

Com o objetivo de pesquisar a viabilidade de se obter conhecimento interessante a partir de banco de dados sobre C&T, procurou-se investigar as técnicas de MD utilizadas na tarefa de classificação, em especial as técnicas híbridas genético-difusas.

Entre as diversas tarefas de MD, uma alternativa menos explorada é a tarefa de modelagem de dependência que se caracteriza como uma generalização da tarefa de classificação, já que em modelagem de dependências pode haver vários atributos meta.

A performance de muitos algoritmos pode ser melhorada através da remoção de atributos irrelevantes, tarefa que pode ser realizada manualmente ou através de algum método automático de seleção de atributos.

Os AG possuem comprovada eficiência no trato com interação entre atributos, o que é uma característica dos dados em C&T, apesar de serem duas a três ordens de magnitude mais lentos do que algoritmos de indução de regras convencionais (Dhar *et al.*, 2000). No entanto, utilizando as restrições e simplificações discutidas acima, o AG torna-se uma alternativa promissora.

Os conjuntos difusos incorporam os termos linguísticos, intrínsecos da linguagem natural, e o raciocínio aproximado, que o torna um método flexível e poderoso para tratar com incertezas.

A abordagem genético-difusa, para realizar a tarefa de extração de regras de previsão, deverá fornecer conhecimento novo, compreensível e exato, representado por regras de tamanho e quantidade variáveis, a partir de banco de dados que contém atributos que se interagem (*e.g.* C&T).

Apesar da existência de várias técnicas convencionais de MD para realizar a tarefa de classificação, as discussões acima indicam que a aplicação de AG's, em combinação híbrida com conjuntos difusos, é uma alternativa viável e promissora.

## Referências

- Carvalho, D.R.; Freitas, A.A. A hybrid decision tree/genetic algorithm for coping with the problem of small disjuncts in data mining. *Proc. Genetic and Evolutionary Computation (GECCO-2000)*, 1061-1068, Las Vegas, NV, USA. July 2000.
- Chen, M-S.; Han, J.; Yu, P.S. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, v. 8, n. 6, p.886-883, 1996.
- Cheung, D.W.; Ng, V.T.; Fu, A.W. Efficient mining of association rules in distributed databases. *IEEE Transactions on Knowledge and Data Eng.*, v.8, n.6, p.911-922, 1996.
- CNPq. *Diretório dos Grupos de Pesquisa no Brasil. Versão 3.0*. Disponível na Internet. <http://www.cnpq.br/gpesq3>.
- Deb, K. *et al.* Learning to Avoid Moving Obstacles Optimally for Mobile Robots Using a Genetic-Fuzzy. In: Eiben, A. E. et al. (Eds.) *PPSN-V*. Lecture Notes in Comp. Science 1498, pp. 583-592. Springer-Verlag, 1998.
- Delgado, M. *et al.* Modular and Hierarchical Evolutionary Design of Fuzzy Systems. In: Banzhaf, W. et al. (Eds.) *Proc. Genetic and Evolutionary Computation Conference (GECCO'99)*, pp. 180-187, Vol. 1, Morgan Kaufmann Publishers, 1999.
- Dhar, V. *et al.* Discovering Interesting Patterns for Investment Decision Making with GLOWER – A Genetic Learner Overlaid With Entropy Reduction. To appear in: *Data Mining and Knowledge Discovery journal*, 4(4), 251-280, 2000.
- Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P. *Advances in knowledge discovery & data mining*. Chapter 1: From data mining to knowledge discovery: an overview. AAAI/MIT, 1996a.
- . Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Second International Conference on KD & DM*. Portland, Oregon, 1996b.
- Fertig, C.S.; Freitas, A.A.; Arruda, L.V.R.; Kaestner, C. A Fuzzy Beam-Search Rule Induction Algorithm. Principles of Data Mining and Knowledge Discovery: Proc. *3rd European Conf. (PKDD-99)* Lecture Notes in Artificial Intelligence 1704, 341-347. Springer-Verlag, 1999.
- Freitas, A.A. Generic, Set-Oriented Primitives to Support Data-Parallel Knowledge Discovery in Relational Database Systems. **Ph.D. Thesis**, University of Essex, UK, July 1997.
- . Understanding the Crucial Differences Between Classification and Discovery of Association Rules – A Position Paper. To appear in: *SIGKDD Explorations*. V.2, No.1, pp. 65-69. 2000.
- . Evolutionary Algorithms. To appear in: *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, 2001.
- Goldberg, D.E. *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley Publishing, Inc., 1989.
- Gonçalves, A.L. *Utilização de Técnicas de Mineração de Dados na Análise dos Grupos de Pesquisa no Brasil*. Florianópolis, 2000. Dissertação (Mestrado em Engenharia de Produção) – Engenharia de Produção e Sistemas, UFSC.
- Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 300 p., 2000.
- Hand, D.J. *Construction and Assessment of Classification Rules*. John Wiley & Sons, 1997. (Capítulos 6, 7 e 8)
- Herrera, F. *et al.* Tackling real-coded genetic algorithms: operators and tools for behavioral analysis, *mimeo*, 1996.
- Janikow, C.Z. A Genetic Algorithm for Optimizing Fuzzy Decision Trees. *6<sup>th</sup> Int. Conf. GA*, p. 421-428, 1995.
- Kacprzyk, J.; Strykowski, P. Linguistic Summaries of Sales Data at a Computer Retailer via Fuzzy Logic and a Genetic Algorithm. *Congress of Evol. Comp. CEC-99*, p. 937-943, 1999.
- King, R.D. *et al.* STATLOG: Comparison of classification algorithms on large real-world problems. *Applied Art. Int.* 9(3), pp. 289-333, May/June 1995.

- Lee, M-R. Generating fuzzy rules by genetic method and its application. *International Journal on Artificial Intelligence Tools*, Vol. 7, No. 4, pp. 399-413, 1998.
- Mendel, J.M. Fuzzy logic systems for engineering: a tutorial. *Proceedings of the IEEE*. v.83, n.3, p. 345-377, 1995.
- Michie, D. *et al.* (Eds.) *ML, Neural and Statistical Classification*, Chapters 1, 2 and 11, 1994.
- Mota, C. *et al.* Independent and Simultaneous Evolution of Fuzzy Sleep Classifiers by Genetic Algorithms. *GECCO-99*, p. 1622-1629, 1999.
- Peña-Reyes, C.A.; Sipper, M. Designing Breast Cancer Diagnostic Systems via a Hybrid Fuzzy-Genetic Methodology. *Fuzzy-IEEE*, 1999.
- Quinlan, J.R. **C4.5: Programs for Machine Learning**. Morgan Kaufmann, 1993.
- Romão, W. *et al.* Algoritmos genéticos e conjuntos difusos aplicados ao controle de um processo térmico. *Revista Tecnológica*, n.8, p. 7-21, 1999a.
- , Extração de regras de associação em C&T: O algoritmo Apriori. *XIX Encontro Nacional em Engenharia de Produção*, 1999b.
- Xiong, N.; Litz, L. Generating Linguistic Fuzzy Rules for Pattern Classification with Genetic Algorithms. *PKDD-99*, p. 574-579, 1999.