UTILIZAÇÃO DE MÉTODOS COMPUTACIONAIS NA SELEÇÃO DE UNIDADES DE CONSERVAÇÃO

Sibelius Lellis Vieira, **Ana Carolina Correia Rézio**

Resumo: *este artigo apresenta uma solução para o problema de seleção de reservas ambientais, utilizando Mineração de dados e a metaheurística Busca Tabu. O objetivo do trabalho consiste em maximizar o número de espécies levando em conta a relação existente entre elas. Os resultados previamente obtidos pela solução foram comparados com a aplicação em um problema real de conservação no Cerrado.*

Palavras-chave: *metaheurísticas, Busca Tabu, mineração, regras, associação*

A redução da biodiversidade, devido sobretudo à degradação ambiental e ao crescente número de espécies ameaçadas, tem sido um problema crítico e vem despertando a atenção dos meios científicos nos últimos anos. Algumas questões vêm sendo analisadas e uma que tem tido destaque é a necessidade de alocação de áreas geográficas específicas para o desenvolvimento das espécies a fim de garantir a proteção delas. Porém, é importante ressaltar que, com o crescimento da urbanização e a necessidade de áreas agricultáveis, essas áreas devem ser selecionadas para a conservação de acordo com os recursos limitados. O objetivo esperado deste sistema de reservas é uma biodiversidade expressiva, relacionada tanto ao número de tipos de vegetação, quanto à população de espécies que habitam na região especificada.

Existem diversos métodos capazes de solucionar este problema de seleção de reservas. Os métodos de enumeração *branch-and-bound* são

métodos de divisão e avaliação progressiva, capazes de fornecer soluções ótimas para uma pequena massa de dados. Para problemas reais muito grandes, encontrar a solução ótima em tempo aceitável pode ser algo inviável. O fato é que, na prática, existe a necessidade de se visualizar diversas soluções que podem estar próximas da solução ótima, mas que traduzem restrições subjetivas do problema (POSSINGHAM; BALL; ANDELMAN 2000). Portanto, a ênfase atual aplicada no tratamento do problema prioriza a flexibilidade da localização e análise de soluções sub-ótimas que podem ser obtidas rapidamente e, assim, podem servir como base para o processo de tomada de decisão por parte dos gestores ambientais (LESLIE *et al.*, 2003).

Dentre as características importantes existentes nas heurísticas, uma das triviais é tentar obter a complementaridade e evitar redundância ou excesso de representação, levando-se em conta as informações já cobertas em iterações anteriores (PRESSEY; COWLING, 2001). Porém, os métodos heurísticos geram soluções aproximadas, sendo isso algumas vezes caracterizado como uma de suas desvantagens, pois não garante que as soluções ótimas sejam encontradas, podendo, inclusive, encontrar soluções indesejáveis (ruins). Além de não garantir a solução ótima, não garantem quão distante seus resultados estão da mesma, o que sugere a necessidade de se empregar outros métodos formais na busca da solução (LESLIE *et al.*, 2003).

Nesse sentido, como a garantia da solução ótima pode não ser conveniente, o foco para o tratamento do problema não é a exatidão da solução, possibilitando soluções aproximadas com a utilização das metaheurísticas. Tais metaheurísticas podem lidar de forma relativamente rápida em uma variedade de problemas e gerar soluções melhores do que aquelas geradas por heurísticas conhecidas, mesmo para problemas mais complexos e com tempo de processamento elevado.

Inúmeros problemas de otimização têm sido solucionados com sucesso pela aplicação de metaheurísticas. O Busca Tabu (BT), por exemplo, é um método muito comum empregado com alto desempenho em problemas complexos e que apresentam a característica da explosão combinatória (BLUM; ROLI, 2003). Este método utiliza técnicas de busca aleatória no espaço de soluções, baseada na intensificação da exploração de regiões com boas soluções e na diversificação da procura. O método parte do princípio de que os problemas de otimização são mais bem resolvidos quando incorporadas a este algumas estratégias inteligentes baseadas em processos implícitos e explícitos de aprendizagem (GLOVER; LAGUNA, 1997).

Diante do problema proposto e da restrição considerada, no qual o número de unidades de conservação que deverão ser preservadas é pre-

viamente definido, beneficiar certas espécies, atribuindo-lhes graus de prioridade, pode ser uma forma de identificar uma unidade de conservação. A solução apresentada neste artigo consiste em indicar quais as melhores áreas para as espécies quando existe relação entre elas. No entanto, tais relações não estão claramente especificadas nos dados fornecidos pelas agências de preservação e, para determinar estas relações e as possíveis formas de classificação das áreas em termo de sua importância ou prioridade, é necessária a análise dos dados fornecidos pela agência de preservação. Esta análise pode ser realizada utilizando técnicas fornecidas pelo *Data Mining*, cujo objetivo consiste em descobrir novas informações em função dos padrões de relacionamento entre os itens de dados do domínio que ocorre com uma determinada freqüência (WITTEN; FRANK, 2005).

O presente objetivo é formular e analisar a aplicação da metaheurística Busca Tabu ao problema de seleção de reserva a fim de maximizar a biodiversidade e, ao mesmo tempo, satisfazer restrições de recursos disponíveis. Para tanto foi considerado que as restrições dos recursos imponham um número máximo de reservas, e assim busca-se maximizar a representação das espécies, levando-se em conta sua prioridade, capturada empiricamente através da relação de ocorrência das espécies nas áreas estudadas. Para garantir a relação entre as espécies, foi utilizada junto ao método Busca Tabu a regra de associação.

Neste artigo, será apresentada a descrição do problema de maximização da biodiversidade, priorizando a relação entre as espécies e considerando as restrições de custo do sistema de unidades através do formalismo da programação inteira e, além disso, serão apresentadas as possíveis formas de resolução do mesmo. Para tanto, o método BT será descrito, realçando os parâmetros que devem ser ajustados para cada tipo de problema. O ajuste dos parâmetros é realizado a partir da aplicação do método em exemplos hipotéticos para extrair as informações necessárias. Além dos exemplos hipotéticos, destaca-se o cenário real de conservação do cerrado para o qual esta abordagem foi desenvolvida. Além da discussão dos resultados aplicados, tanto para os cenários hipotéticos quanto para o cenário real, serão apresentadas as considerações sobre os desafios correntes e futuras direções.

MATERIAIS

Os problemas de seleção de reservas podem ser formulados através de um modelo de programação inteira (0-1) e podem ser classificados de duas formas. Uma delas consiste em minimizar o uso de recurso em relação

ao custo da reserva e, ao mesmo tempo, assegurar metas de conservação de biodiversidade adequadas, baseadas em algumas informações ou considerações. A outra classificação consiste em maximizar a preservação e conservação da biodiversidade em relação a um dado custo. Este custo pode ser associado à área da reserva, ao número de reservas, entre outras variáveis.

A primeira classificação é formulada como um problema de cobertura de conjunto dentro do campo da pesquisa operacional (CHURCH; STOMS; DAVIS, 1996, GAREY; JOHNSON, 1979). Essa formulação, embora comum, não fornece a flexibilidade necessária para as várias opções presentes na solução de um problema de seleção de reservas.

Na prática, o problema é formulado segundo um conjunto de recursos financeiros disponíveis, que permitem a seleção de um número de reservas restringido por questões muitas vezes de ordem socioeconômica. Sendo assim, tem-se a segunda classificação para a formulação do problema, conhecida como formulação da representação das espécies (WILLIAMS; REVELLE, 1997). Para a pesquisa operacional, as formulações são equivalentes, e estão associadas à solução de problemas lineares inteiras com restrições lineares. No entanto, esta última formulação permite que as soluções sejam aplicadas de forma a planejar a seleção em períodos, e não de uma única vez.

Além dessas formulações, tem havido um interesse cada vez maior por situações que indiquem uma possível relação das reservas entre si e/ou das espécies entre si. Outro ponto importante é a relação espacial entre as espécies no sistema de reservas, indicando uma possível relação biológica entre espécies diferentes que compartilham um determinado *habitat* e cooperam na sustentação da diversidade deste *habitat.* Para garantir que essa relação possa estar presente na solução do sistema de reservas, a formulação do problema deve levar em conta a distribuição *in situ* das espécies no ambiente e procurar manter esta relação na solução encontrada.

Formulação do Problema

O problema consiste em maximizar uma função objetivo que representa o número de espécies que deverão ser mantidas a um dado custo, cujo valor está vinculado ao número de reservas disponíveis. Para solucionar o problema, a função objetivo foi ponderada pela prioridade das espécies. Para isso, foi adicionado um fator que leva em conta a relação entre elas. Essa prioridade, derivada da relação das espécies, foi obtida pela regra de associação, uma das técnicas presentes no *Data Mining*. Seu valor é determinado por uma variável que representa a confiança da espécie e é proporcional

ao número de ocorrências das espécies associadas entre si, ou seja, o número de vezes em que uma espécie aparece com as outras. Em resumo, essa variável, chamada de prioridade, caracteriza a importância de se manter em uma reserva as espécies que aparecem juntas e é obtida pela soma das confianças das espécies relacionadas duas a duas. Este problema é apresentado nas Equações 1, 2, 3 e 4 e suas notações básicas estão descritas na Tabela 1.

$$\text{Maximize } \Sigma_{i\in\eta} P_i Y_i \tag{1}$$

$$\text{Sujeito à } \Sigma_{k\in\eta} A_{ik} X_i Y_i \text{ para } i = 1,..,M \tag{2}$$

$$\Sigma_{i\in\eta} X_j = Q \text{ para } i = 1,..,M \tag{3}$$

$$Y_i, X_{j,a} \{0,1\} \tag{4}$$

A Eq. (1) expressa a necessidade de maximizar o número de espécies a partir de suas relações. Nesta equação, Y_i indica se a espécie *i* é adequadamente representada no sistema de reservas e P_i indica a prioridade para se preservar a espécie *i.* A Eq. (2) apresenta a restrição de que a espécie *i* só pode ser preservada no sítio *k* se ocorrer neste sítio. A Eq. (3) indica o número máximo de sítios que se tornarão reservas. Esse número é limitado pelo valor *Q* e as reservas estão representadas em X_j, variável que indica se o sítio j será ou não selecionado para reserva. Já a Eq. (4) restringe o problema à variável binária (0-1), caracterizando um problema de maximização inteira com restrições lineares.

Quadro 1: Notação Básica

μ	Conjunto dos sítios
η	Conjunto das espécies
M	Número de sítios
N	Número de espécies
Q	Número de unidades de conservação
X_k	Variável de decisão 0-1 que indica se o sítio k é uma reserva
Y_i	Variável de decisão 0-1 que indica se a espécie i será preservada
P_i	Parâmetro que indica o nível de prioridade da espécie i
A_{ik}	Parâmetro que indica a ocorrência da espécie i no sítio k

Cenários e Conjuntos de Dados

A eficácia da metaheurística é ilustrada inicialmente por uma base de dados com parâmetros aleatórios que estabelecem o número de sítios, espécies, a presença e a ausência da espécie em cada sítio e o número de sítios que devem ser preservados.

No algoritmo de BT os parâmetros foram ajustados a partir da comparação da solução encontrada por este com uma solução ótima encontrada por um algoritmo enumerativo. Para determinar a solução ótima e com bom desempenho, o algoritmo foi executado em instâncias com o volume de dados (espécies e sítios) reduzido.

Depois de ajustados os parâmetros, o BT foi aplicado para analisar os padrões da biodiversidade de 131 espécies de anfíbios anuros presentes em diferentes partes do bioma do Cerrado, dividido em 181 sítios com 1º de latitude e longitude (DINIZ-FILHO *et al.*, 1996). Estes sítios se estendem pela região norte, nordeste, centro-oeste e sudeste, englobando cerca de 9 estados destas regiões, indo desde a latitude 5,5º até a latitude 23,5º e longitude 41,5º até 57,5º.

MÉTODOS

Em geral, as metaheurísticas para tratar problemas de combinatória procuram explorar locais com boas soluções no espaço de busca, processo definido como intensificação. Em contrapartida, as metaheurísticas também buscam diversificar o espaço de busca quando este se encontra circunscrito em um ótimo local. Este procedimento é conhecido como mecanismo de diversificação.

Para solucionar problemas de natureza combinatória, tal como o problema de seleção de reservas apresentado, existem várias metaheurísticas. Dentre as mais utilizadas encontra-se a Busca Tabu, o S*imulated Annealing* e a Computação Evolutiva (BLUM; ROLI, 2003). Neste artigo, o algoritmo selecionado é baseado no BT. Este método foi escolhido por apresentar um bom desempenho nas soluções desses problemas, tanto de natureza linear quanto quadrática (GLOVER, 1990). Este algoritmo tem sido utilizado com sucesso em vários campos, podendo inclusive ser utilizado de forma integrada, nas ferramentas de auxílio à tomada de decisão ou, também, para análise e calibração dos parâmetros.

Descrição do *Data Mining*

O *Data Mining* é uma das principais técnicas utilizadas no processo de descoberta de conhecimento em base de dados (*Knowledge Discovery in Databases* – KDD), que tem como objetivo identificar as informações importantes e utilizá-las no processo de tomada de decisões. Ou seja, *Data Mining* refere-se à mineração ou descoberta de novas informações em função de padrões ou regras que descrevem o comportamento em uma ampla base de dados (OLIVEIRA, 2001, WITTEN; FRANK, 2005).

Existem diferentes tipos de informações que podem ser mineradas em uma base de dados em diferentes formas. Dentre elas encontram-se as regras de associação, hierarquias de classificação, *clusters* de dados, padrões seqüenciais e os padrões em séries temporais (GONÇALVES, 2004). Neste artigo, o objetivo do estudo são as regras de associação, também conhecido como afinidade de grupos. Estas regras visam combinar itens importantes, tal que a presença de um item em uma transação *t* pressupõe, na mesma transação, a presença de outro (OLIVEIRA, 2001).

Em resumo, uma regra de associação representa um padrão de relacionamento entre os itens de dados do domínio da aplicação que ocorre com uma determinada freqüência na base de dados. Esta regra demonstra a presença de um determinado conjunto que implica na presença de algum outro conjunto distinto de itens (GILLMEISTER; CAZELLA, 2007).

Uma aplicação muito comum das técnicas de associação encontra-se na área de *marketing,* cujo objetivo consiste em descobrir as associações existentes entre os produtos vendidos e, assim, determinar quais vendas normalmente são realizadas em conjunto. O resultado dessa descoberta possibilita definir como os produtos devem ser expostos para aumentar o volume de vendas (OLIVEIRA, 2001).

Em uma regra associativa, o número de instâncias que são preditas corretamente é chamada de suporte e sua precisão é conhecida como confiança. Esta confiança é determinada a partir da proporção das instâncias corretamente preditas sobre todas as instâncias analisadas (GILLMEISTER; CAZELLA, 2007).

O suporte de um conjunto de itens *Y* (espécies), *Sup(Y)*, representa a porcentagem de transações da base de dados que contêm os itens de *Y*. Em uma regra de associação, o suporte das espécies $Y_i \rightarrow Y_j$, $Sup(Y_i \rightarrow Y_j)$, é dado por $Sup(Y_i \cup Y_j)$. Já a confiança desta regra, $Conf(Y_i \rightarrow Y_j)$, representa, dentre as transações que contêm Y_i, a porcentagem de tran-

sações que também contêm Y_j, ou seja, $Conf(Y_i \rightarrow Y_j) = Sup(Y_i \cup Y_j)/Sup(Y_i)$ (GONÇALVES, 2004).

A utilização da regra de associação no problema proposto consiste em associar as espécies duas a duas. Esta associação, calculada pela soma das confianças das espécies relacionadas com as outras, garante as prioridades das espécies. O sítio com espécies de maior confiança possui a maior prioridade. O cálculo da prioridade dos sítios através do somatório das confianças das espécies presentes em cada sítio é apresentado na Figura 1.

```
Algoritmo CalcularPrioridade

  T[i] <- SomaDasEspecies(i);
  Para i de 1 até N faça
   C[i][i] <- 1;
   Para j de i+1 até N faça
    Para k de 1 até M faça
     Se A[k][i] = 1 e A[k][j] = 1 então
      C[j][i] <- C[j][i] + 1;
     FimSe
    FimPara
    C[i][j] <- C[j][i];
   FimPara
  FimPara
  Para i de 1 até N faca
   Para j de 1 até N faça
    C[j][i] <- C[j][i]/T[i];
   FimPara
  FimPara
  P[i]<- SomaDasConfiancas(i);
FimAlgoritmo
```

Figura 1: Descrição Algorítmica para o Cálculo da Prioridade

O resultado da simulação do algoritmo da Figura 1 está ilustrado na Figura 2. Essa simulação foi realizada com base na matriz de ocorrência *A* com 20 espécies e 20 sítios cuja distribuição está apresentada na Tabela 1 e o resultado da matriz de confiança de cada espécie está representado na Tabela 2.

Tabela 1: Distribuição das Espécies *Y* nos Sítios *X*

A[X][Y]	Y01	Y02	Y03	Y04	Y05	Y06	Y07	Y08	Y09	Y10
X01	0	0	0	0	0	0	0	0	0	0
X02	0	0	0	0	1	0	1	0	0	1
X03	0	0	0	0	0	0	1	0	1	1
X04	0	0	0	0	0	0	1	0	0	1
X05	0	0	0	0	0	0	1	0	0	0
X06	0	0	0	1	0	0	0	0	0	0
X07	0	0	1	0	0	0	0	0	0	0
X08	0	0	0	0	0	0	1	0	0	1
X09	0	0	0	0	0	0	1	0	0	1
X10	0	0	0	0	0	0	1	0	0	0
X11	0	0	0	1	0	0	0	0	0	0
X12	0	0	0	0	0	0	0	0	0	0
X13	1	0	0	0	0	0	1	0	0	1
X14	1	0	0	0	0	0	1	0	0	1
X15	0	0	0	0	0	0	1	0	0	0
X16	1	1	0	0	0	1	0	0	0	1
X17	1	1	0	0	0	1	0	0	0	1
X18	1	0	0	0	0	0	1	0	0	1
X19	1	0	0	0	0	0	1	1	0	1
X20	0	0	0	0	0	0	1	0	0	1

A[X][Y]	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
X01	1	1	0	1	0	0	0	0	0	0
X02	0	0	0	0	1	0	0	1	0	0
X03	0	0	0	0	0	0	0	1	0	0
X04	0	0	1	0	0	0	0	1	0	0
X05	0	0	0	0	0	0	0	1	0	0
X06	0	0	0	0	0	0	0	1	0	0
X07	0	0	0	0	0	0	0	0	0	0

Continua...

...Conclusão

A[X][Y]	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
X08	0	0	0	0	0	0	0	1	0	0
X09	0	0	0	0	0	0	0	1	0	0
X10	0	0	0	0	0	0	0	1	0	0
X11	0	0	0	0	0	0	1	0	0	0
X12	0	0	0	0	0	0	1	0	0	0
X13	0	0	0	0	1	1	0	1	0	1
X14	0	0	0	0	1	1	0	1	0	0
X15	0	0	0	0	0	0	1	1	1	0
X16	0	0	0	0	1	0	0	1	0	1
X17	0	0	0	0	1	1	0	1	0	1
X18	0	0	0	0	1	1	0	1	1	1
X19	0	0	0	0	0	1	0	1	1	0
X20	0	0	0	0	0	1	0	1	1	0

Tabela 2: Matriz de Confiança

C[Y][Y]	Y01	Y02	Y03	Y04	Y05	Y06	Y07	Y08	Y09	Y10
Y01	1	1	0	0	2	4	0,31	1	0	0,5
Y02	0,33	1	0	0	0	1	0	0	0	0,17
Y03	0	0	1	0	0	0	0	0	0	0
Y04	0	0	0	1	0	0	0	0	0	0
Y05	0,33	0	0	0	1	0	0,08	0	0	0,08
Y06	0,67	1	0	0	0	1	0	0	0	0,17
Y07	0,67	0	0	0	1	0	1	1	1	0,83
Y08	0,17	0	0	0	0	0	0,08	1	0	0,08
Y09	0	0	0	0	0	0	0,08	0	1	0,08
Y10	1	1	0	0	1	1	0,77	1	1	1
Y11	0	0	0	0	0	0	0	0	0	0
Y12	0	0	0	0	0	0	0	0	0	0
Y13	0	0	0	0	0	0	0,08	0	0	0,08
Y14	0	0	0	0	0	0	0	0	0	0
Y15	0,83	1	0	0	1	1	0,31	0	0	0,5
Y16	0,83	0,5	0	0	0	0,5	0,38	1	0	0,5

...Conclusão

C[Y][Y]	Y01	Y02	Y03	Y04	Y05	Y06	Y07	Y08	Y09	Y10
Y17	0	0	0	0,5	0	0	0,08	0	0	0
Y18	1	1	0	0,5	1	0,5	1	1	1	12
Y19	0,33	0	0	0	0	0	0,31	1	0	0,25
Y20	0,5	1	0	0	0	0,5	0,15	0	0	0,33

C[Y][Y]	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
Y01	0	0	0	0	0,83	0,83	0	0,4	0,5	0,75
Y02	0	0	0	0	0,33	0,17	0	0,13	0	0,5
Y03	0	0	0	0	0	0	0	0	0	0
Y04	0	0	0	0	0	0	0,33	0,07	0	0
Y05	0	0	0	0	0,17	0	0	0,07	0	0
Y06	0	0	0	0	0,33	0,17	0	0,13	0	0,5
Y07	0	0	1	0	0,67	0,83	0,33	0,87	1	0,5
Y08	0	0	0	0	0	0,17	0	0,07	0,25	0
Y09	0	0	0	0	0	0	0	0,07	0	0
Y10	0	0	1	0	1	1	0	0,8	0,75	1
Y11	1	1	0	1	0	0	0	0	0	0
Y12	1	1	0	1	0	0	0	0	0	0
Y13	0	0	1	0	0	0	0	0,07	0	0
Y14	1	1	0	1	0	0	0	0	0	0
Y15	0	0	0	0	1	0,67	0	0,4	0,25	1
Y16	0	0	0	0	0,67	1	0	0,4	0,75	0,75
Y17	0	0	0	0	0	0	1	0,07	0,25	0
Y18	0	0	1	0	1	1	0,33	1	1	1
Y19	0	0	0	0	0,17	0,5	0,33	0,27	1	0,25
Y20	0	0	0	0	0,67	0,5	0	0,27	0,25	1

Como ilustrado na Figura 1, o cálculo da prioridade de cada espécie é realizado pela Equação 5:

$$P[Y_i] = \text{“}\Sigma_{i \in \eta}\ C[j][i]/T[i] \qquad (5)$$

Para a simulação, o resultado obtido foi o estampado na Tabela 3.

Tabela 3: Prioridade das Espécies

P[Y]	7,7	7,5	1	2	7	9,5	4,6	7	4	17	3	3	4	3	6,8	6,8	2,3	5,1	6	7,3

Descrição da Busca Tabu

Fred Glover propôs, na década de 1980, o método chamado *Tabu Search (TS),* também conhecido como BT. Este método parte do princípio de que os problemas de otimização possuem um maior desempenho em suas resoluções quando incorporado à metaheurística, ou seja, algumas estratégias inteligentes baseadas em processos implícitos de aprendizagem.

A BT é uma técnica que possui um conjunto de funções, e quando integrada à inteligência artificial, permite resolver problemas complexos, incorporando o uso de memória adaptativa e de exploração sensível. A memória adaptativa (*adaptive memory*) é utilizada para prover a resolução de problemas através de um processo inteligente para diversificar ou intensificar a busca. Assim, quando incorporada à BT, possibilita que as informações das escolhas ruins ou configurações com qualidade inferior sejam aproveitadas, para que, posteriormente, não ocorra a mesma seleção, além de permitir que a estratégia busque, a cada processo, melhores configurações. Isso permite escapar de situações em que se tem um ótimo local, fazendo com que a solução tenha a possibilidade de convergir em alguns passos para as escolhas menos favorecidas, conforme as especificações da memória adaptativa, podendo ter condições de prosperar para um ótimo global.

Com o objetivo de evitar a realização de movimentos desnecessários, utiliza-se o critério tabu baseado em seu histórico de ocorrências recentes ou freqüentes. Assim, este critério pode restringir a realização de movimentos que tenham sido feitos recentemente ou freqüentemente, ou que possuam qualidade semelhante à situação atual. No entanto, o critério de aspiração possibilita movimentos que no critério tabu não são permitidos. Com isso, em busca de qualidade, este critério possibilita soluções que determinem a melhor vizinhança, similar/dissimilar à solução já existente.

Na formulação de algoritmos BT eficientes, um aspecto importante é a definição e caracterização da vizinhança de uma determinada configuração. No BT, uma transição consiste em passar de uma configuração corrente para uma configuração vizinha. Idealmente, essa configuração vizinha

deve ser uma configuração que ainda não foi visitada e que seja a melhor entre as configurações vizinhas. Caso não exista uma configuração vizinha com melhor qualidade ou a configuração vizinha de melhor qualidade não satisfaz o critério tabu e não satisfaz o critério de aspiração, deve-se passar para a configuração vizinha menos ruim.

```
Algoritmo Busca Tabu
  X <- X_INIC
  Para w de 1 até ITERACOES faça
    Para vizinho de 1 até NVIZ faça
      X_VIZ <- X;
      Gerar VIZ(X_VIZ)
      Calcular OBJ (X_VIZ)
      Ordenar OBJ
    FimPara
    valorFO <-Calcular OBJ (X)
    Para j de 1 até NVIZ faça
      X_VIZ <- X;
      Gerar VIZ(X_VIZ) para OBJ[j];
      Se X_VIZ ∉ LTABU ou (X_VIZ ∈ LTABU e valorFO < OBJ[j])
        então X <- X_VIZ;
        AtualizaLT(LTABU, X_VIZ);
        Sai do laço
      FimSe
    FimPara
  FimPara
  Return X;
FimAlgoritmo
```

Figura 2: Descrição Algorítmica da Busca Tabu Aplicado na Solução do Problema

O BT possui alguns parâmetros de controle, dentre eles o número de iterações ou critério de parada, a estrutura e o tamanho da vizinhança, a formação e manutenção da Lista Tabu (LTabu) e o critério de aspiração. No algoritmo apresentado na Figura 3, a solução inicial é denotada por X_{INIC} e foi denotado X e X_{VIZ} como a melhor solução e a solução corrente respectivamente. *OBJ* representa o conjunto de funções objetivo para o conjunto de soluções obtidas a partir da solução corrente e *valorFO* representa a Função

Objetivo para a vizinhança (*VIZ*) da solução corrente. O conjunto *VIZ* é composto por *NVIZ* elementos (*VIZ[i]*) e cada elemento é determinado a partir da solução corrente, que através de uma permutação aleatória obtêm uma nova solução que compõe *VIZ*. No problema descrito, cuja representação pode ser feita por bits, cada nova solução é gerada a partir da solução corrente através da permutação da posição entre dois bits. Este processo representa a mudança de uma reserva previamente selecionada por outra ainda não-selecionada.

Os valores de *OBJ* são calculados a partir do critério de prioridade estabelecido. Para nosso problema, esse valor é obtido pela soma dos produtos da prioridade com a variável que representa se a espécie será ou não preservada no sítio. Em *OBJ*, estes valores estão ordenados de forma decrescente. O primeiro vizinho da lista de vizinhos que não está na lista tabu ou que está na lista, mas satisfaz o critério de aspiração, é escolhido para entrar na lista tabu e é definido como a nova solução corrente. O critério de aspiração é baseado na função objetivo contida em *OBJ*. Se o vizinho corrente tiver o valor da função objetivo (*valorFO*) inferior ao da função objetivo do vizinho avaliado (*OBJ[j]*), o critério de aspiração é satisfeito.

Todo o procedimento citado é repetido, até o número de iterações especificado, *Iteracoes*. Ao final, o valor da melhor solução encontra-se em X.

RESULTADOS E DISCUSSÃO

Para demonstrar a importância da técnica de Busca Tabu na solução dos problemas de seleção de reservas na forma apresentada, primeiro é necessário investigar quais parâmetros são adequados, de forma que o algoritmo selecione soluções próximas da solução ótima. No entanto, como adequar o algoritmo de busca tabu com relação aos parâmetros do método, tais como o critério de parada, a exploração da vizinhança e o número de vizinhos, tamanho e formação da lista tabu? O algoritmo de busca tabu pode ser adequado a partir de sua a aplicação a um conjunto de dados gerado aleatoriamente e comparando o resultado a um procedimento que calcula a solução ótima.

Para a aplicação do método em problemas reais, pode-se fazer a seguinte pergunta: quais os resultados da aplicação desta metaheurística em problemas reais de conservação do cerrado? A solução para essa questão se faz a partir da aplicação do algoritmo em um conjunto de dados que representam a ocorrência de anuros na região do cerrado.

Ajuste do Busca Tabu

O ajuste do desempenho do algoritmo de BT para o problema apresentado foi realizado a partir da comparação do seu resultado com o resultado de um procedimento que gera a solução ótima para problemas com um número reduzido de sítios e espécies. Nestas instâncias, considera-se uma matriz de ocorrência A_{ij} com uma distribuição uniforme que representa um número aleatório de ocorrências das espécies nos sítios. Para cada sítio *i*, uma linha da matriz de ocorrência é gerada com probabilidade de ocorrência em cada coluna de O_i ($0 < O_i < 1$), cujo valor é fornecido na entrada do programa. Para cada espécie atribuí-se uma prioridade, essa prioridade é calculada pela soma das confianças da espécie *j* associada com as outras.

Um dos critérios a ser ajustado é o critério de parada. Este critério é baseado no número de iterações e varia entre 100 e 1000 com variação de 100. Outro parâmetro importante é a formação da vizinhança. No problema abordado, a solução é representada pelo vetor *X* que indica o número *Q* de um total de *M* sítios que foram escolhidos como reservas. Para cada possível configuração da solução *X* é estabelecido uma vizinhança através da permutação aleatória de 2 sítios, ou seja, um sítio previamente não escolhido terá seu estado alterado para reserva enquanto um que foi previamente escolhido passa para o estado de não-reserva. Ainda na formação da vizinhança, um critério importante é o tamanho da mesma. Para tanto, foi analisado o comportamento do algoritmo para vizinhanças com 5 e 10 vizinhos.

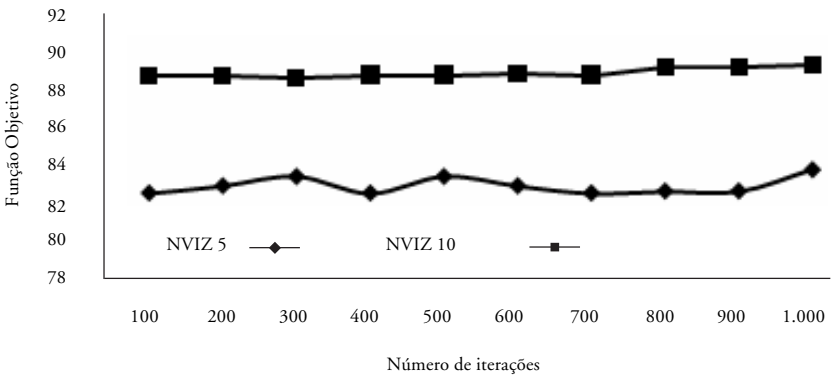

Figura 3: Valor da Função Objetivo para LT de 5

O critério tabu, representado por LTabu, tem o objetivo de evitar que movimentos desnecessários sejam realizados, baseando-se no histórico de movimentos recentes. Este critério está limitado pelo tamanho da lista (NTabu*)*, tendo valores 5, 10 e 20. Deste modo, os últimos NTabu movimentos não são permitidos a menos que satisfaça o critério de aspiração. A finalidade do critério de aspiração é justamente possibilitar movimentos que o critério tabu não permite. Assim, este critério possibilita soluções que determinem a melhor vizinhança, similar/dissimilar à solução já existente e é adotado sempre que ocorre uma melhor solução corrente.

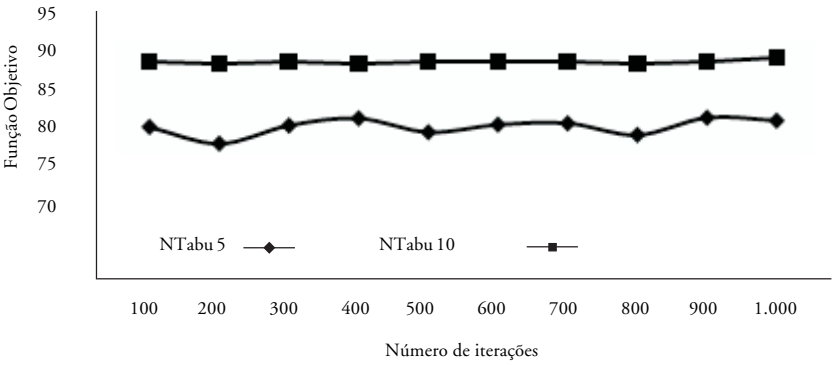

Figura 4: Valor da Função Objetivo para LT de 10

O cenário com 10 reservas (*Q*), 20 espécies (*N*) e 20 sítios (*M*) é usado para ilustrar o comportamento do BT em relação à solução ótima obtida para o problema (vide Figuras 4, 5 e 6). Para obter a solução ótima, em tempo razoavelmente pequeno, utiliza-se o método de enumeração *branch-and-bound.* Nas figuras citadas, o eixo ordenado representa o número de iterações (critério de parada), o eixo abscissa representa o valor médio da função objetivo, encontrada pela média de um total de 100 rodadas do algoritmo de busca tabu e a legenda ilustra a variação da vizinhança. Nas ilustrações, observa-se que o tamanho da vizinhança tem maior impacto no estabelecimento do resultado ótimo do que o tamanho da lista tabu e do número de iterações. Para os experimentos realizados, o tamanho da lista e a variação no número de iterações não interferem de forma tão determinante no comportamento da solução.

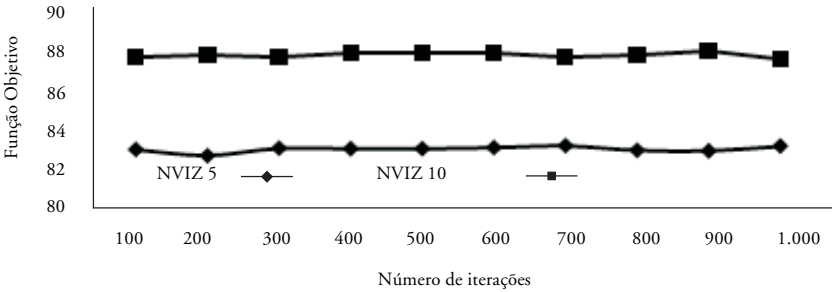

Figura 5: Valor da Função Objetivo para LT de 20

Características dos Sistemas de Reservas

Depois de realizados os testes com um cenário de ajuste, o resultado para o cenário real é apresentado na Figura 7, que considera a ocorrência de 131 espécies de anfíbios anuros presentes em diferentes partes do Cerrado, estimado como uma região com 181 sítios de 10.000km^2. Das 131 espécies consideradas, algumas delas aparecem em todos os sítios e outras aparecem em apenas um sítio, ou seja, as espécies estão bem distribuídas e isso gera uma matriz de ocorrência bastante variada (COUTO, DINIZ-FILHO, BINI, 2004; VIEIRA *et al.*, 2005).

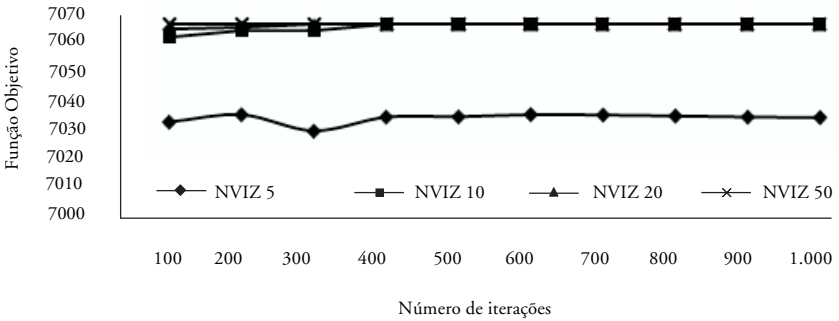

Figura 6: Resultado Obtido com Dados Reais

Neste cenário é considerado o número de reservas igual a 100, o NTabu igual a 20 e a vizinhança variando 5, 10, 20 e 50. Como pode ser observado na ilustração, o número de iterações em relação ao valor da Fun-

ção Objetivo, ponderada pela prioridade das espécies, nota-se que com 400 iterações para NVIZ maior do que 10, atingimos a maior função objetivo, ou seja, a melhor solução é obtida.

CONCLUSÃO

O problema de seleção de reservas foi formulado a fim de permitir a exploração empírica entre as espécies, na forma apresentada por uma matriz de ocorrência das mesmas. Este problema estabelece que a seleção realizada deva maximizar o número de espécies por meio de restrições de orçamento e permite extrair características do problema para o processo de tomada de decisão da gestão ambiental. O uso da BT, uma metaheurística bastante utilizada, aplicada juntamente com o *Data Mining* atende aos requisitos básicos do processo de solução, ou seja, permite obter soluções boas ou ótimas em um tempo relativamente baixo a partir do ajuste dos parâmetros de controle existente no método.

Esta abordagem permite que novas questões sejam investigadas, tais como a persistência da biodiversidade, levando-se em conta a dinâmica populacional e o comportamento das áreas não selecionadas.

Referências

BLUM, C.; ROLI, A. Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Computing Surveys*, v. 35, n. 3, p. 268-308, 2003.

CHURCH, R. L.; STOMS, D. M.; DAVIES, F. W. Reserve selection as a maximal covering location problem. *Biological Conservation*, v. 76, p. 105-112, 1996.

COUTO, M. S. D. S.; DINIZ-FILHO, J. A. F.; BINI L. M. Programação linear e seleção de áreas prioritárias para conservação da biodiversidade. *Caderno de Estudos*, Goiânia, 2004.

DINIZ-FILHO, J. A. F. et al. Conservation biogeography of anurans in Brazilian Cerrado. *Biodiversity Conservation*, v. 16, p. 997-1.008, 1996.

GAREY, M.; JOHNSON, D. *Computers and Intractability*: a guide to the theory of NP-Completeness. San Francisco: Ed. W. H. Freeman, 1979.

GILLMEISTER, P. R. G.; CAZELLA, S. C. *Uma análise comparativa de algoritmos de regras de associação*: minerando dados da indústria automotiva. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2007.

GLOVER, F. Tabu search: a tutorial. *Interfaces*, v. 20, n. 4, p. 74-94, 1990.

GLOVER, F.; LAGUNA, M. *Tabu search*. [S. L.]: Klewer Academic Publishers, 1997.

GLOVER, F.; MÉLIAN, B. Busqueda tabu. *Revista Iberoamericana de Inteligência Artificial*, v. 19, p. 29-48, 2003.

GONÇALVES, E. C. *Regras de associação e suas medidas de interesse objetivas e subjetivas*. Niterói: Instituto de Computação, Universidade Federal Fluminense (UFF), 2004.

LESLIE, H. et al. Using siting algorithms in the design of Marine Reserve Networks. *Ecological Applications*, v. 13, p. S185-S198, 2003.

OLIVEIRA, I. C. de. Aplicação de Data Mining na busca de um modelo de prevenção da mortalidade infantil. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis, 2001.

POSSINGHAM, H.; BALL, I.; ANDELMAN, S. Mathematical methods for identifying representative reserve networks. In: POSSINGHAM, H.; BALL, I.; ANDELMAN, S. *Quantitative methods for conservation biology*. New York: Spring-Verlag, 2000. p. 291-305,

PRESSEY, R.; COWLING, R. Reserve selection algorithms and the real world. *Conservation Biology*, v. 15, p. 275-277, 2001.

VIEIRA, S. et al. Utilização de metaheurísticas no processo de seleção de reservas ambientais. In: XXXVII SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, Gramado-RS, 2005.

WILLIAMS, J.; REVELLE, C. Applying mathematical programming to reserve selection. *Environmental Modeling and Assessment*, v. 2, p. 167-175, 1997.

WITTEN, I. H.; FRANK, E. Data Mining. *Pratical machine learning tools and techniques.* 2. ed. San Francisco: Elsevier, 2005.

Abstract*: this paper presents a solution to the reserve selection problem using the Data Mining and Tabu Search. The goal is to maximize the number of species concerning the relationship among them. The outcome previously obtained was compared with the application in a real conservation problem in the Brazilian Cerrado.*

Key words: *metaheuristics, Tabu Search, data mining and association rules*

SIBELIUS LELLIS VIEIRA
Pós-doutor em Ciência da Computação pela Universidade de Virgínia. Doutor em Engenharia Elétrica e Computação pela Unicamp. Graduado em Física pela Universidade Federal de Goiás. Professor no Departamento de Computação da Universidade Católica de Goiás (UCG).

ANA CAROLINA CORREIA RÉZIO
Graduanda no Curso de Ciências da Computação da UCG.