

Aplicando árvore de decisão para a localização de padrões no modelo Sugarscape do NetLogo

Vágner de Oliveira Gabriel¹, Cleo Zanella Billa¹, Diana Francisca Adamatti¹,
Fabiana Lorenzi²

¹Centro de Ciências Computacionais - Universidade Federal do Rio Grande - FURG
Av. Itália, Km 8 - Campus Carreiros – 96203-900 – Rio Grande, RS – Brasil

²Universidade Luterana do Brasil (ULBRA)
Av Farroupilha, 8001 Bairro São José – 92420280 – Canoas, RS – Brasil

vdeoliveiragabriel@gmail.com, {cleobilla, dianaadamatti}@furg.br
fabilorenzi@gmail.com

Abstract. *Sugarscape is a model that simulates populations in an environment with limited resources, where they must move around seeking resources for their survival. NetLogo software has an implemented model of Sugarscape, which simulates populations of agents who need to collect sugar to stay alive in the environment. This work aims to apply the decision tree technique on data collected in simulations carried out in the Sugarscape model of NetLogo. Thus, the purpose is to find patterns that define which attributes are important in determining whether an agent will live or die in the simulation environment.*

Resumo. *O Sugarscape é um modelo que simula populações em um ambiente com recursos limitados, onde as mesmas devem se locomover buscando recursos para a sua sobrevivência. O software NetLogo possui um modelo implementado do Sugarscape, o qual simula populações de agentes que necessitam coletar açúcar para se manterem vivos no ambiente. Este trabalho tem como objetivo, aplicar a técnica de árvore de decisão sobre dados coletados em simulações realizadas no modelo Sugarscape do NetLogo. Dessa forma, o propósito é localizar padrões que definam quais atributos são importantes para determinar se um agente irá viver ou morrer no ambiente de simulação.*

1. Introdução

O Sugarscape é um modelo amplamente utilizado em simulações baseadas em agentes. Este modelo apresenta como finalidade a simulação de sociedades que necessitam de recursos, os quais estão distribuídos em regiões de um ambiente em que a sociedade está localizada [Epstein and Axtell 1996]. Portanto, é de extrema relevância coletar estes recursos disponíveis no ambiente para que seja possível a sobrevivência dos agentes.

O software NetLogo possui implementado o Sugarscape¹, onde os agentes possuem a necessidade de coletar açúcar para sobreviverem no ambiente em que estão situados. O agente é modelado no sugarscape com as seguintes características: açúcar individual que o deixa vivo, metabolismo que consome o seu açúcar individual, visão de pontos onde estão distribuídos os açúcares e quantidade de açúcar existente no local.

¹<http://ccl.northwestern.edu/netlogo/models/Sugarscape1ImmediateGrowback>

A mineração de dados surge como uma ferramenta para localizar padrões significativos que não podem ser localizados apenas ao consultar as informações em um banco de dados [Tan et al. 2006]. Na mineração de dados existem distintas técnicas para a localização de padrões em base de dados com um número elevado de informações. Uma das técnicas mais conhecidas para classificação se chama árvore de decisão, a qual será utilizada neste trabalho.

Este trabalho tem como objetivo realizar uma mineração de dados utilizando o software Weka aplicando árvore de decisão sobre dados gerados no modelo de simulação de recursos emergentes Sugarscape. Com a aplicação da árvore de decisão sobre os dados que serão coletados no modelo Sugarscape, se torna possível encontrar um padrão nas simulações e identificar os atributos mais determinantes para a sobrevivência do agente no ambiente.

Este trabalho está organizado da seguinte forma. Na seção 2 é apresentada a fundamentação teórica do trabalho, onde são conceituados os principais componentes estudados para a realização deste trabalho. Na seção 3, é descrita a metodologia utilizada para o desenvolvimento do trabalho. Na seção 4 é apresentado o resultado obtido no trabalho. Na seção 5, são realizadas as considerações finais e na seção 6 são apresentadas as referências estudadas para o desenvolvimento deste trabalho.

2. Fundamentação Teórica

Nesta seção, serão apresentados os componentes que serviram como base teórica para o desenvolvimento deste trabalho. Esta seção está dividida em duas subseções, onde a primeira subseção aborda uma breve introdução sobre agentes, sistemas multiagente e o modelo Sugarscape do NetLogo. A segunda subseção é utilizada para a definição de árvore de decisão.

2.1. Sistemas Multiagente

Agentes podem ser definidos como entidades reais ou virtuais, as quais estão situadas em um ambiente, onde se comunicam e interagem com outros agentes visando atingir seus objetivos [Rezende 2003]. Os agentes podem trabalhar isoladamente ou formar sociedades, resultando assim um sistema multiagente. Segundo [Wooldridge and Jennings 1994], agentes inteligentes possuem as seguintes características:

- **Autonomia:** os agentes possuem controle total de suas ações e estado interno, uma vez que, trabalham sem ligação direta com seres humanos;
- **Habilidade sociais:** se comunicam com outros agentes através do uso de um agente de comunicação de linguagem;
- **Reatividade:** Percebem alguma alteração no ambiente e agem sobre a mesma em um curto espaço de tempo;
- **Pró-atividade:** o agente pode raciocinar e tomar uma iniciativa.

2.1.1. Modelo Sugarscape

O modelo Sugarscape foi apresentado por [Epstein and Axtell 1996], onde é utilizado amplamente na comunidade de simulações baseadas em agentes. O Sugarscape simula

populações em um ambiente com recursos limitados, no caso do modelo introduzido por Epstein e Axtell, os agentes necessitam de açúcar para a sua sobrevivência.

No ambiente do Sugarscape (Figura 1) existem regiões com um nível elevado de açúcar e outras com um nível baixo, ou até mesmo sem açúcar. Os agentes se locomovem no ambiente e são capazes de coletar açúcar devido as suas propriedades como visão e metabolismo, sendo a visão utilizada para enxergar os lugares que possuem açúcar e o metabolismo para consumir o açúcar coletado. A visão e o metabolismo variam em cada agente, tornando dessa forma a população da simulação heterogênea. Conceituando o modelo sugarscape de um forma geral, os agentes dependem do açúcar para a sua sobrevivência, porém o metabolismo pode consumir todo o estoque de açúcar individual levando o agente a morte. Dessa forma, se torna necessário que o agente esteja posicionado em uma região com níveis altos de açúcar.

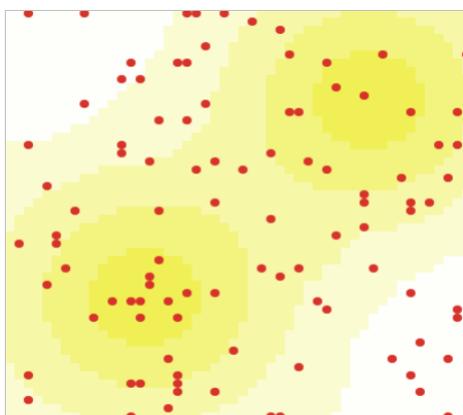


Figura 1. Modelo Sugarscape (NetLogo)

2.2. Árvore de Decisão

A mineração de dados pode ser definida como o processo de descoberta automática de informações relevantes ao usuário em grande depósito de dados [Tan et al. 2006]. Em mineração de dados são introduzidos os conceitos de funcionalidade, sub-funcionalidade, tarefas e o conceito de técnicas. Tarefas são as especificações do que o usuário está pesquisando, tipo de regularidades e padrões, como exemplo: a compra elevada de um produto em um supermercado apenas no mês de abril. Já técnicas, são caracterizadas na especificação dos métodos que garantem a descoberta de padrões que o indivíduo está buscando. As principais técnicas em mineração de dados são: estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento-podavalidação [De Amo 2004].

As funcionalidades, em mineração de dados, definem como os tipos de padrões ou relacionamentos que interligam os registros e suas variáveis podem ser utilizados na mineração. Como exemplos de funcionalidades pode-se citar a análise descritiva e a análise de prognóstico. Este trabalho utiliza apenas uma sub-funcionalidade da análise de prognóstico, a sub-funcionalidade de classificação, focando apenas na sua definição.

[De Amo 2004] define classificação como um processo para encontrar um conjunto de modelos que definem e distinguem classes ou conceitos visando prever as classes de objetos que ainda não foram classificadas. Árvores de decisão e redes neurais são técnicas usualmente utilizadas em tarefas de classificação.

A árvore de decisão pode ser caracterizada como um fluxograma que possui similaridade com uma árvore. Este fluxograma se constitui de nós internos, os quais representam testes em atributos e cada folha representa a distribuição dos registros [da Costa Côrtes et al. 2002]. Em [Monard and Baranauskas 2003], é apresentado um exemplo que demonstra uma árvore de decisão para diagnosticar um paciente.

No exemplo (Figura 2), as elipses servem como um teste em um atributo para um conjunto de dados de pacientes. Os retângulos representam classes, os quais referenciam os diagnósticos.

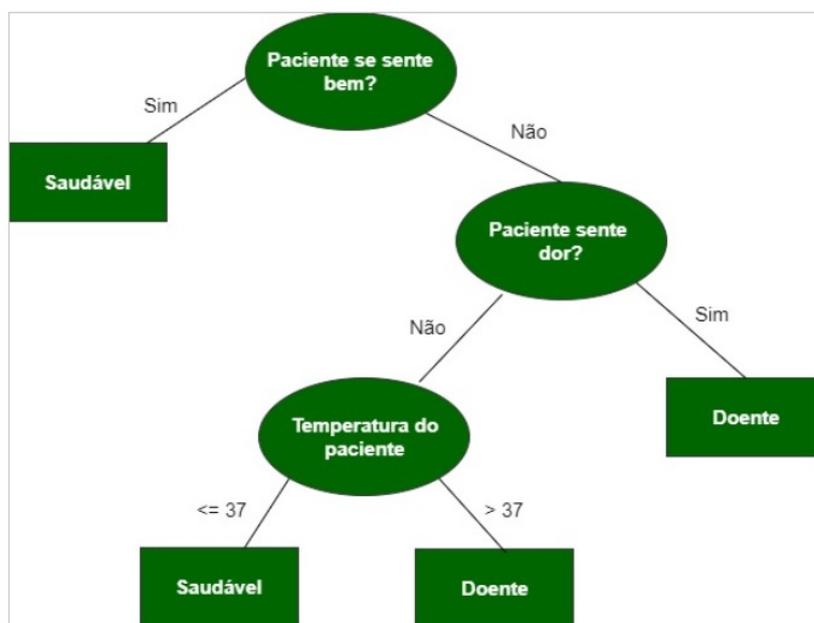


Figura 2. Árvore de Decisão de diagnóstico de um paciente (Adaptado de Monard e Baranauskas (2003))

No exemplo apresentado por Monard e Baranauskas (2003) (Figura 2), existem duas classes que são a classificação do paciente (Saudável/Doente). Existem também os conjuntos de dados que vão classificar se o paciente está saudável ou doente. Este conjuntos são:

- Paciente se sente bem?
- Paciente sente dor?
- Paciente sente dor?

3. Materiais

Nesta subseção serão conceituados os softwares utilizados para o desenvolvimento desse trabalho.

3.1. NetLogo

A ferramenta NetLogo ² (Figura 3), é um ambiente de modelagem programável onde é possível simular fenômenos naturais e sociais. Esta ferramenta se adequa bem para modelar sistemas que apresentam uma complexidade maior em seu desenvolvimento, sendo

²<https://ccl.northwestern.edu/netlogo/>

composta ainda por uma linguagem de programação simples voltada para a modelagem e para simulação de fenômenos naturais e sociais [Wilensky 1999]. O NetLogo se adapta bem para a modelagem de sistemas que apresentam uma complexidade elevada e que evoluem ao longo do tempo. A ferramenta possibilita aos desenvolvedores passar instruções a centenas ou milhares de ‘agentes’, que operam de forma independente, interagindo entre si ou com o ambiente.

O NetLogo possui uma extensa documentação, trazendo alguns modelos em sua biblioteca, sendo uma coleção de simulações pré-escritas que podem ser usadas e modificadas. NetLogo em um geral é uma ferramenta poderosa e simples o suficiente para possibilitar que estudantes e professores simulem suas pesquisas abrangendo diversos campos de conhecimento.

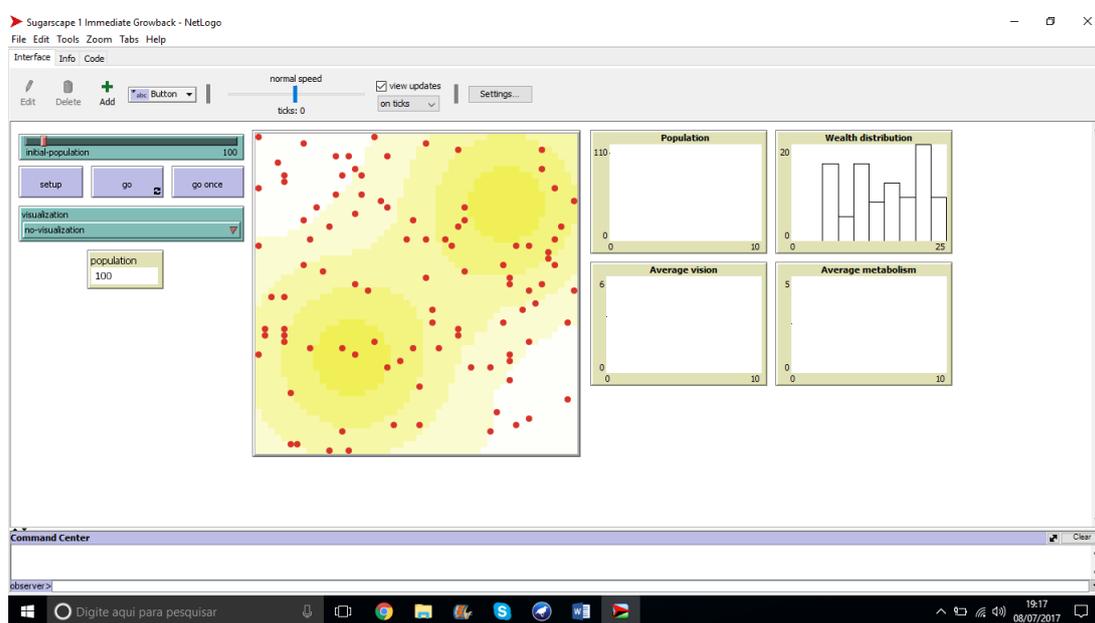


Figura 3. Ferramenta NetLogo

3.2. Weka

O software Weka ³ (Figura 4), possui uma coleção de algoritmos de aprendizagem de máquinas, estes usados para a mineração de dados buscando localizar padrões entre outras características em uma gigantesca base de dados. O Weka foi desenvolvido na Universidade de Waikato na Nova Zelândia e fornece métodos para a mineração de dados, como classificação, regressão, agrupamento, regras de associação e visualização [Hall et al. 2009].

Neste trabalho foi adotada a utilização do software Weka pelo fato desta ferramenta ser de fácil manuseio e possibilitar a utilização de árvore de decisão para a classificação dos dados através do uso do algoritmo J48.

4. Metodologia

Na primeira etapa do trabalho foi utilizado o software NetLogo para realizar as simulações do modelo Sugarcape, o qual já está instanciado como um modelo de exemplo do soft-

³<https://www.cs.waikato.ac.nz/ml/weka/>

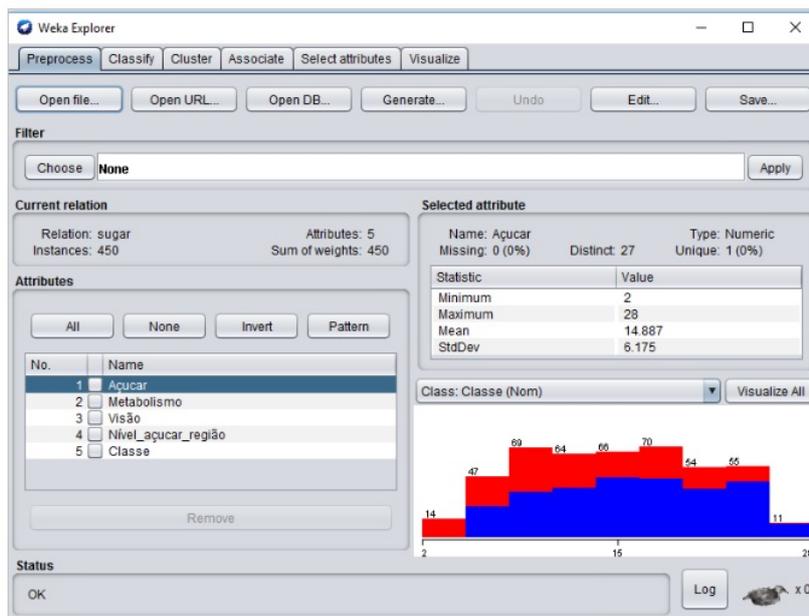


Figura 4. Software Weka

ware. Algumas alterações foram realizadas no modelo incluso no NetLogo, objetivando facilitar a coleta dos dados das simulações. A Figura 5 ilustra o código implementado para salvar em um arquivo txt os seguintes dados: id do agente, seu nível de açúcar, seu nível de metabolismo, seu nível de visão e a quantidade de açúcar existente na região onde o agente está posicionado.

```

to-report crialista
  set lista []
  set lista lput who lista
  set lista lput sugar lista
  set lista lput metabolism lista
  set lista lput vision lista
  set lista lput max-sugar lista
  report lista
end

to escrever [lista2]
  file-open "dados.txt"
  file-print reduce [ ( word ?1 "," ?2 ) ] lista2;;
  show reduce [ ( word ?1 "," ?2 ) ] lista2;;
  file-flush
  file-close
end

```

Figura 5. Código implementado no NetLogo

A primeira parte do código da Figura 5 foi desenvolvida com a finalidade de criar uma lista (vetor) e salvar as seguintes informações:

- Who: o *Id* (identificador) do agente;
- Sugar: nível de açúcar individual do agente;
- Metabolismo: nível de metabolismo que o agente possui;

- Vision: nível de visão do agente;
- Max-sugar: nível de açúcar existente na região onde o agente está posicionado;

O segundo trecho do código apresentado na Figura 5 foi desenvolvido com a finalidade de salvar em um arquivo txt, este chamado de “dados”, os arquivos salvos na lista.

Foram realizadas 20 simulações contendo 50 agentes no ambiente, onde foram coletados inicialmente os dados descritos na Figura 5 dos 50 agentes no ambiente, estes dados foram gerados de forma aleatória pelo software.

Das 20 simulações, a média de ticks (medida de tempo no software NetLogo) foi de 100. Após a finalização de cada simulação foram coletadas as informações de quais agentes sobrevivem no decorrer da simulação e colocadas na linha correspondente a cada agente classificando ele como “vive” ao final da simulação ou “morre”. Ao final da primeira etapa foi possível coletar uma base de dados contendo 1.000 dados. A Figura 6 ilustra os dados salvos no arquivo txt.

```
7,3,5,3,vive
19,4,1,0,morre
22,4,3,4,vive
18,4,6,4,vive
11,3,1,2,morre
20,1,2,1,vive
16,2,1,2,vive
10,1,4,0,morre
13,1,5,3,vive
23,2,2,2,vive
```

Figura 6. Dados após finalização de cada simulação

A segunda etapa consiste em preparar o arquivo de dados para um padrão que o software Weka consiga reconhecer. Dessa forma, foi necessário alterar o arquivo dados.txt para a extensão arff, esta reconhecida pelo software Weka. Para a leitura do arquivo foi necessária a criação de um cabeçalho, este ilustrado na Figura 7.

No cabeçalho tem a relação chamada sugar, e os atributos correspondentes a cada coluna do arquivo dados. A primeira coluna corresponde ao açúcar do agente, a segunda corresponde ao metabolismo do agente, a terceira corresponde a visão do agente, a quarta corresponde ao nível de açúcar do agente. Estas quatro primeiras colunas são do tipo numérico, uma vez que, os dados inclusos nelas são números. A coluna cinco é a classe do agente que representa a definição se ele vive ou morre ao final na simulação.

Após a adaptação dos dados coletados para um tipo de arquivo, o qual tornou-se reconhecido pelo software Weka, foi possível gerar a árvore decisão, a qual é apresentada na seção resultados.

```
@relation sugar
@attribute Açucar numeric
@attribute Metabolismo numeric
@attribute Visão numeric
@attribute Nível_açucar_região numeric
@attribute Classe {vive,morre}
@data
7,3,5,3,vive
19,4,1,0,morre
22,4,3,4,vive
18,4,6,4,vive
```

Figura 7. Alteração do arquivo para extensão arff

5. Resultados

Aplicou-se a técnica de árvore de decisão sobre os dados coletados como saída nas simulações realizadas no modelo *Sugarscape*. Dessa forma, localizando os seguintes Padrões:

- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for ≤ 1 , o metabolismo ≤ 1 , o nível de açúcar da região ≤ 0 , o agente morre.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for ≤ 1 , o metabolismo ≤ 1 , o nível de açúcar da região > 0 , o agente vive.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for ≤ 1 , o metabolismo > 1 , o agente morre.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for ≤ 1 , o metabolismo ≤ 2 , o agente vive.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for > 1 , o metabolismo > 2 , nível de açúcar da região for ≤ 2 , a visão for ≤ 5 , o agente morre.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for > 1 , o metabolismo > 2 , nível de açúcar da região for ≤ 2 , a visão for > 5 , o agente vive.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for ≤ 3 , o nível de açúcar da região for > 1 , o metabolismo > 2 , nível de açúcar da região for > 2 , o agente vive.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for > 3 , visão ≤ 5 , o agente morre.

- Se o nível de açúcar da região for ≤ 3 , o metabolismo for > 3 , visão > 5 , o nível de açúcar da região for ≤ 1 , o agente morre.
- Se o nível de açúcar da região for ≤ 3 , o metabolismo for > 3 , visão > 5 , o nível de açúcar da região for > 1 , o agente vive.
- Se o nível de açúcar da região for > 3 , o agente vive.

A Figura 8, ilustra o resultado obtido em formato de árvore, o qual foi gerado pela ferramenta Weka tendo como entrada os resultados coletados das simulações no NetLogo sobre o modelo *Sugarscape*.

A árvore (Figura 9), mostra que o atributo mais determinante para a sobrevivência de um agente é região onde ele está posicionado. Se a região possuir mais que três de nível ele vive, caso o contrário sua sobrevivência depende de um conjunto de outras situações.

```

J48 pruned tree
-----
Nível_açúcar_região <= 3
|  Metabolismo <= 3
|  |  Nível_açúcar_região <= 1
|  |  |  Metabolismo <= 1
|  |  |  |  Nível_açúcar_região <= 0: morre (7.0)
|  |  |  |  Nível_açúcar_região > 0: vive (14.0)
|  |  |  |  Metabolismo > 1: morre (32.0)
|  |  |  Nível_açúcar_região > 1
|  |  |  |  Metabolismo <= 2: vive (141.0/6.0)
|  |  |  |  Metabolismo > 2
|  |  |  |  |  Nível_açúcar_região <= 2
|  |  |  |  |  |  Visão <= 5: morre (40.0/4.0)
|  |  |  |  |  |  Visão > 5: vive (7.0/2.0)
|  |  |  |  |  Nível_açúcar_região > 2: vive (39.0/1.0)
|  |  |  Metabolismo > 3
|  |  |  |  Visão <= 5: morre (82.0)
|  |  |  |  Visão > 5
|  |  |  |  |  Nível_açúcar_região <= 1: morre (4.0)
|  |  |  |  |  Nível_açúcar_região > 1: vive (7.0/2.0)
Nível_açúcar_região > 3: vive (77.0)

Number of Leaves   :    11
Size of the tree   :    21

```

Figura 8. Árvore de decisão contendo os padrões localizados sobre os resultados das simulações do modelo *Sugarscape*

6. Considerações Finais

Com a aplicação da árvore de decisão se tornou possível localizar os atributos que são mais relevantes para a sobrevivência do agente no modelo de *Sugarscape* disponível no software NetLogo, além de apresentar um conjunto de situações que podem determinar a sobrevivência ou morte de um agente no ambiente.

A aplicação de árvores de decisão pode ser uma grande ferramenta para a localização de padrões gerados em simulações. Atualmente, existem diversos modelos de simulação baseados em agentes que geram um grande número de resultados. A utilização de mineração de dados para analisar e encontrar padrões nestes resultados pode ser de grande valia.

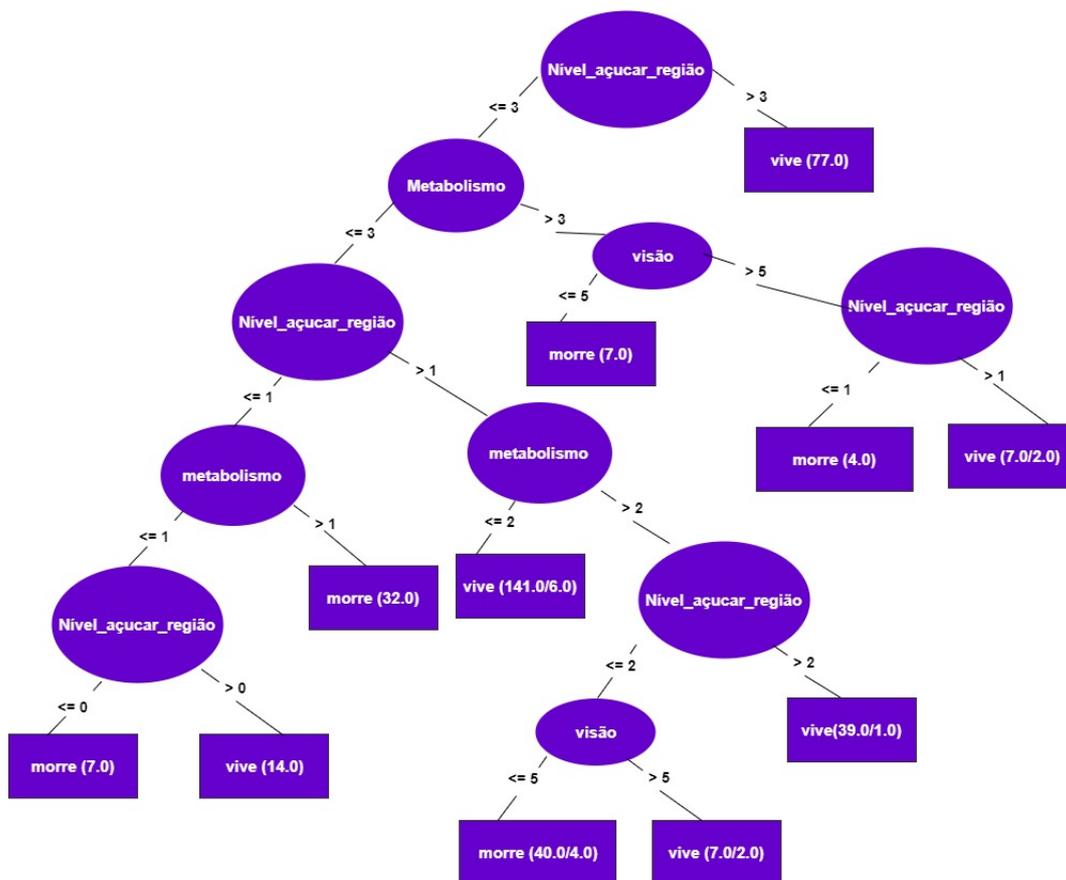


Figura 9. Fluxograma em formato de árvore contendo os padrões localizados sobre os resultados das simulações do modelo Sugarscape

Futuramente, se espera aplicar outras técnicas de mineração de dados sobre o modelo *Sugarscape*, visando testar sua eficiência e comparar os novos achados com os resultados obtidos neste trabalho.

Referências

- da Costa Côrtes, S., Porcaro, R. M., and Lifschitz, S. (2002). *Mineração de dados-funcionalidades, técnicas e abordagens*. PUC.
- De Amo, S. (2004). Técnicas de mineração de dados. *Jornada de Atualização em Informática*.
- Epstein, J. M. and Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- Monard, M. C. and Baranauskas, J. A. (2003). Indução de regras e árvores de decisão. *Sistemas Inteligentes. Rezende, SO Editora Manole Ltda*, pages 115–140.
- Rezende, S. O. (2003). *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda.

Tan, P.-N. et al. (2006). *Introduction to data mining*. Pearson Education India.

Wilensky, U. (1999). Netlogo.

Wooldridge, M. and Jennings, N. R. (1994). Agent theories, architectures, and languages: a survey. In *Intelligent agents*, pages 1–39. Springer.