

Estimativa do Peso de Abate a Partir de Variáveis de Cria em Bovinos

Thales V. Maciel¹, Vinícius do N. Lampert², Denizar S. Souza³, Rodrigo R. da Silva¹

¹Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSUL)
Campus Bagé – Av. Leonel de Moura Brizola, 2501 – 96.418-400 – Bagé – RS – Brasil

²Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Unidade Pecuária Sul – Bagé – RS – Brasil

³Centro de Ciências Exatas e Aplicadas (CCEA)
Universidade da Região da Campanha (URCAMP) – Bagé, RS – Brasil

thalesmaciel@ifsul.edu.br, vinicius.lampert@embrapa.br,
denizarsouza@urcamp.edu.br, orki2008@gmail.com

***Abstract.** This paper describes a method for obtaining decision trees for predicting the weight at slaughter for bovine based on their breeding data. For such, data mining classification tasks were performed after data preprocessing. All numeric attributes were discretized by cluster discovery. Obtained results showed that clustering techniques as means for discretization may generate classes in better balancing conjecture when in comparison to the non-equal frequency binning method, allowing the discovery of models that may be applied to real world problems.*

***Resumo.** Este artigo descreve uma metodologia para obtenção de árvores de decisão para previsão do peso de abate com base em variáveis de cria em bovinos. Para tal, procedeu-se a tarefas de mineração de dados com classificação após pré-processamento com discretização dos atributos numéricos por descoberta de agrupamentos. Os resultados obtidos mostraram que a descoberta de agrupamentos como forma de discretização pode gerar classes com balanceamento de melhor qualidade em comparação ao método tradicional, permitindo a indução de modelos utilizáveis em problemas reais.*

1. Introdução

O sistema de produção de gado de corte compreende uma das principais atividades de exploração econômica no Brasil, onde, há décadas, tem-se afastado o cenário de resistência ao emprego tecnológico, de modo a permitir estudos para o melhoramento dos índices de qualidade na produção de carne, por exemplo, através de computação aplicada [Barbosa 1999].

Em [Costa 2016], foram analisados dados zootécnicos de 401 animais bovinos da raça Hereford com vistas em prever o peso de fazenda e bonificação dos indivíduos. Para tal, foram empregadas redes neurais artificiais e obtidos resultados satisfatórios, segundo o autor, embora não tenha considerado a praticidade da utilização do método por produtores pecuários em meio às tarefas cotidianas, tampouco apresentou comparações com outros métodos para descoberta de conhecimento em bancos de dados.

Em [Da Mota et al. 2017], foram empregadas tecnologias de armazém de dados, consultas analíticas e mineração de dados para 1142230 registros de abates bovinos. O objetivo foi o de prever o grau de acabamento e o rendimento das carcaças, onde os resultados foram classificados, pelos autores, como promissores, embora tenham considerado apenas as acurácias dos modelos descobertos na análise dos resultados.

O problema de pesquisa abordado no presente estudo é fundamentado em “quais variáveis podem ser coletadas, pelos criadores, sobre os indivíduos de rebanhos bovinos em etapa de desenvolvimento de cria e que explicam a obtenção de peso de carcaça ótimo?” O objetivo é obter um modelo gráfico, de fácil interpretação, capaz de orientar os criadores bovinos sobre o desempenho de seus rebanhos, ainda em etapa anterior ao desmame, com previsões do futuro peso de abate dos animais.

2. Metodologia

Procedeu-se à descoberta de conhecimento em bancos de dados (DCBD), especificamente com as tarefas de mineração de dados descritas nesta seção.

O processo de DCBD pode ser dividido em três etapas [Maciel et al. 2015]: o pré-processamento, onde o conjunto de dados original é preparado para as próximas etapas do processo através de tarefas de filtragem conforme necessário; o processamento, onde algoritmos de mineração de dados são aplicados sobre o conjunto de dados pré-processado e; o pós-processamento, onde os padrões descobertos no processamento são analisados e transformados em conhecimento útil sobre o domínio estudado.

Para fins de realização das tarefas e experimentos descritos neste estudo, foi empregado o Waikato Environment for Knowledge Analysis (WEKA), um ambiente para análise de conhecimento desenvolvido pela Universidade de Waikato [Hall et al. 2009]. Trata-se de uma coleção de implementações de algoritmos para utilização em atividades de mineração de dados [Witten et al. 2017].

O conjunto de dados analisado teve sua apresentação original em 167 instâncias de animais bovinos e 5 atributos, conforme descrição na tabela 1.

Tabela 1. Descrição do conjunto de dados analisado.

#	Nome do Atributo	Significado	Tipo de Dado
1	nascimento_mes	mês de nascimento (01-12)	nominal
2	desmame_mes	mês de desmame (01-12)	nominal
3	desmame_idade	idade de desmame em meses	numérico
4	desmame_peso	peso de desmame em quilogramas	numérico
5	abate_peso	peso de abate em quilogramas	numérico

No pré-processamento, os atributos numéricos foram discretizados, conforme a escala de Likert [Likert 1932], de forma a criar segmentos nominais dentre o intervalo numérico com as denominações: muito baixo, baixo, intermediário, alto e muito alto.

No âmbito deste estudo, foi proposta a descoberta automatizada de 5 agrupamentos com base na distância de Manhattan em aplicação do algoritmo Simple k-means [Arthur e Vassilvitskii 2007] como método de discretização. Isto ocorreu em alternativa ao método tradicional de discretização, que ocorre pelo fracionamento uniforme de intervalos numéricos, que ocasionou baixo balanceamento entre as classes descobertas em experimentos com o mesmo conjunto de dados.

A figura 1 apresenta os histogramas referentes às distribuições de frequência das instâncias de bovinos nas categorias propostas pela escala de Likert nos atributos discretizados pelos métodos tradicional (a) e com a descoberta automatizada dos agrupamentos baseados na distância de Manhattan (b).

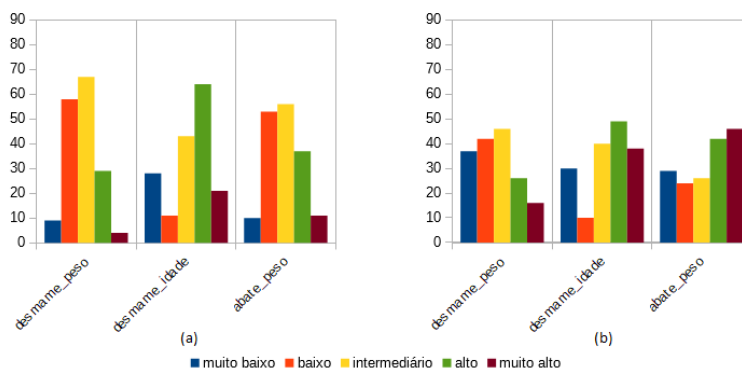


Figura 1. Histogramas referentes aos atributos discretizados por fracionamento (a) e por descoberta de agrupamentos (b).

A etapa de processamento foi realizada com o algoritmo J48 [Quinlan 1996] configurado para permitir apenas divisões binárias em galhos formados por atributos nominais e desconsiderar limites inferiores de ocorrências de instâncias em folhas para critérios de poda em seu treinamento. Os demais parâmetros do algoritmo foram mantidos em conformação padrão. A etapa de testes do modelo descoberto foi realizada por validação cruzada em 5 frações sobre o mesmo conjunto de dados de entrada para treinamento.

3. Resultados Obtidos

Os resultados obtidos na tarefa de classificação descrita na seção 2 foram apresentados na forma de árvore de decisão, matriz de confusão e acurácia dos modelo. Também é discutida a praticidade dos modelo descoberto. A tabela 2 apresenta a matriz de confusão encontrada no teste, que apresentou acurácia de 54%.

Tabela 2. Matriz de confusão resultante do experimento de classificação.

classe verdadeira \ prevista	mb	b	i	a	ma
muito baixo (mb)	16	1	5	2	5
baixo (b)	2	6	3	1	14
intermediário (i)	3	3	20	2	14
alto (a)	1	0	6	7	10
muito alto (ma)	1	3	4	1	37

A árvore de decisão resultante do experimento é apresentada na figura 2. Nela é possível observar um modelo gráfico capaz de prover orientação instrutiva sobre a influência das variáveis de cria, categorizadas de acordo com a escala de Likert, sobre o peso de abate em bovinos.

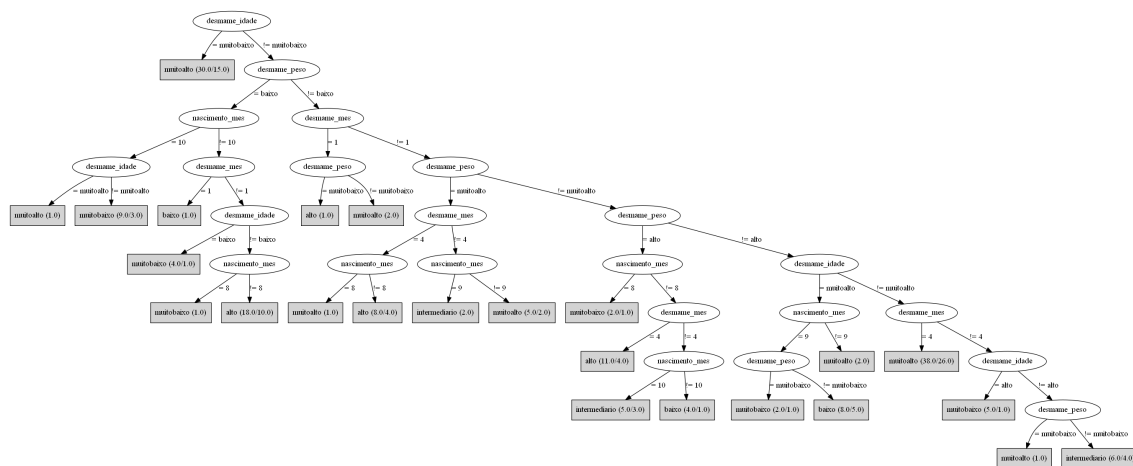


Figura 2. Árvore de decisão para predição do peso de abate

4. Conclusão

O presente trabalho buscou um método para descoberta de árvores de decisão capazes de auxiliar os produtores de gado de corte na previsão do peso de abate de bovinos com base nas respectivas variáveis de cria, ou seja, dados que podem ser coletados entre o nascimento e o desmame dos animais destinados ao abate.

Foi realizado experimento de classificação com o algoritmo J48 após pré-processamento do conjunto de dados para discretização dos atributos numéricos, tendo sido proposta uma alternativa à discretização pelo método tradicional e descoberto um modelo parcialmente capaz de explicar a influência das variáveis de cria analisadas no peso de abate dos animais.

Referências

Arthur, D. and Vassilvitskii S. (2007) k-means++: the advantages of carefull seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 1027-1035.

Barbosa, P. (1999) Raças e estratégias de cruzamento para produção de novilhos precoces. Embrapa Pecuária Sudeste. In: Simpósio de Produção de Gado de Corte, 1. Viçosa, Brasil.

Costa, C. L. (2016). Utilização de características zootécnicas e de manejo na pecuária para previsão do peso final e bonificação de bovinos empregando redes neurais artificiais. Trabalho de conclusão de curso, Universidade Federal do Pampa.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. (2009) The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1.

Likert, R. (1932) A Technique for the Measurement of Attitudes. Archives of Psychology. 140: 1–55.

Maciel, T., Seus, V., Machado, K. and Borges, E. (2015). Mineração de dados em triagem de risco de saúde. Revista Brasileira de Computação Aplicada, 7(2), 26-40.

Mota, F., Souza, K., Ishii, R. and Gomes, R. (2017) BovReveals: uma plataforma OLAP e data mining para tomada de decisão na pecuária de corte. In: Congresso Brasileiro de Agroinformática, 11. Campinas, Brasil.

Quinlan, R. (1993) C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. Journal of artificial intelligence research, 4, 77-90. Chicago, IL.

Witten, I., Frank, E., Hall, M. and Pal, C. (2017) Data mining: practical machine learning tools and techniques. Morgan Kaufmann.