



Emprego de regras de associação para extração de padrões mercadológicos de touros Nelore com avaliação genética¹

Jeferson Nomelini^{2*}, Solange Oliveira Rezende³, Claudio Haruo Yamamoto³, Luiz Antônio Framartino Bezerra², Pedro Alejandro Vozzi⁴, Raysildo Barbosa Lôbo²

¹ CNPq.

² FMRP-USP.

³ ICMC-USP.

⁴ INTA-EEA Chubut.

* *In memoriam.*

RESUMO - Objetivou-se com este trabalho utilizar regras de associação para identificar forças de mercado que regem a comercialização de touros com avaliação genética pelo programa Nelore Brasil. Essas regras permitem evidenciar padrões implícitos nas transações de grandes bases de dados, indicando causas e efeitos determinantes da oferta e comercialização de touros. Na análise foram considerados 19.736 registros de touros comercializados, 17 fazendas e 15 atributos referentes às diferenças esperadas nas progênes dos reprodutores, local e época da venda. Utilizou-se um sistema com interface gráfica usuário-dirigido que permite geração e seleção interativa de regras de associação. Análise de Pareto foi aplicada para as três medidas objetivas (suporte, confiança e *lift*) que acompanham cada uma das regras de associação, para validação das mesmas. Foram geradas 2.667 regras de associação, 164 consideradas úteis pelo usuário e 107 válidas para $lift \geq 1,0505$. As fazendas participantes do programa Nelore Brasil apresentam especializações na oferta de touros, segundo características para habilidade materna, ganho de peso, fertilidade, precocidade sexual, longevidade, rendimento e terminação de carcaça. Os perfis genéticos dos touros são diferentes para as variedades padrão e mocho. Algumas regiões brasileiras são nichos de mercado para touros sem registro genealógico. A análise de evolução de mercado sugere que o mérito genético total, índice oficial do programa Nelore Brasil, tornou-se um importante índice para comercialização dos touros. Com o uso das regras de associação, foi possível descobrir forças do mercado e identificar combinações de atributos genéticos, geográficos e temporais que determinam a comercialização de touros no programa Nelore Brasil.

Palavras-chave: comércio de touros, forças de mercado, melhoramento genético, mineração de dados, tecnologia da informação

Use of association rules for extracting trade patterns of Nelore bulls using genetic evaluation

ABSTRACT - The aim of this work was to use association rules to identify forces that govern bull trade with genetic evaluation by Programa Nelore Brasil. Those rules permit to highlight patterns implicit in the transactions of large databases, which indicate determining causes and effects of supply and trade of bulls. In the analysis, it was considered 19,736 records of traded bulls, 17 farms and 15 attributes related to expected differences on the progenies of sires, place and time of sale. A system with user-driven graphical interface that allows generation and interactive selection of association rules was used. Pareto analysis was applied for the three objective measures (support, trust and lift) that follow each one of the association rules to validate them. It was generated 2,667 association rules, 164 considered to be useful by the user and 107 valid for $lift \geq 1.0505$. The farms participating in the Nelore Brasil program present expertise in bull offering according to maternal ability characteristics, weight gain, fertility, sexual precocity, longevity, yield and termination of carcass. The genetic profiles of the bulls are different for standard and Mocho Nelore varieties. Some Brazilian regions are market niche for sires without registered herd. The analysis of evolution of the market suggests that Mérito Genético Total (MGT), the official index of the Program Nelore Brasil has become an important index for commercialization of bulls. By using association rules, it was possible to discover market forces and to identify combinations of genetic, geographical and temporal attributes that determine commercialization of bulls in Nelore Brasil program.

Key Words: bull trade, data mining, genetic improvement, information technology, market forces

Introdução

O agronegócio representa a principal atividade econômica brasileira e é responsável por cerca de 30 a 40% do produto interno bruto (PIB), das exportações e da geração de empregos. A cadeia produtiva da pecuária de corte oferece matéria-prima para 49 segmentos da indústria (SIC, 2009). Além de produzir avaliações genéticas, programas de melhoramento genético têm a função de melhorar a produtividade da pecuária de forma eficiente e sustentável, por meio do desenvolvimento de tecnologias para gestão dos rebanhos, seleção e acasalamentos.

O programa Nelore Brasil conta com 1,19 milhão de animais avaliados em 413 rebanhos. Com 20 anos de atuação, contabiliza aumento de 1114,6 e 709,8% para animais e rebanhos, respectivamente, promovendo melhoramento genético e aumento da produtividade (Lôbo et al., 2008b). Dada sua grandeza e seu rápido crescimento, necessita técnicas e ferramentas analíticas que possibilitem a gestão dos rebanhos (Nomelini, 2006).

A mineração de dados pode ser definida como um processo de várias etapas, interativo e iterativo, para a identificação de padrões compreensíveis, válidos, novos e potencialmente úteis, embutidos em grandes bases de dados (Fayyad et al., 1996). Como vivemos na era do conhecimento, a sobrevivência de uma instituição ou empresa não depende se ela é proprietária da informação, mas sim, de como trabalha o conhecimento relacionado a essa informação. Partindo dessa premissa, o uso da, em análises do programa Nelore Brasil, é de suma importância para aumento de sua eficiência e competitividade.

Regras de associação, uma tarefa da mineração de dados permite a identificação de padrões intratransações em uma base de dados, mensurando o quanto a presença de um conjunto de atributos nos registros implica na de outro conjunto distinto de atributos. As regras podem evidenciar forças de mercado implícitas nas transações da respectiva base, ou seja, causas e efeitos determinantes da oferta e comercialização de produtos.

O objetivo neste trabalho foi identificar forças de mercado que regem a comercialização de touros com avaliação genética pelo programa Nelore Brasil, com uso de regras de associação.

Material e Métodos

Para análise, foi utilizada a base de dados (disponível em março de 2008) referente à avaliação genética do programa Nelore Brasil (Lôbo et al., 2008a). Nas fazendas participantes do programa, são realizadas quatro pesagens

anuais dos animais (nascimento aos 18 meses de idade) e duas aferições de perímetro escrotal (9 e 8 meses de idade) e, também, pesagens das matrizes e progênes ao parto e à desmama. As diferenças esperadas nas progênes (DEP) foram estimadas pela metodologia dos modelos mistos, com propriedades do melhor preditor linear não-viesado (*BLUP – best linear unbiased predictor*), sob modelo animal (Van Vleck, 1993). As características foram analisadas com o *software* desenvolvido pela equipe técnica do CTAG (CTAG, 2007) e o *software* MTDFREML (Boldman et al., 1995).

A descoberta de regra de associação foi formalmente definida por Agrawal & Srikant (1994). Seja uma base de dados D composta por um conjunto de itens $A = \{a_1, \dots, a_m\}$ ordenados lexicograficamente e por um conjunto de transações $T = \{t_1, \dots, t_m\}$, no qual cada transação $t_i \in T$ é composta por um conjunto de itens, de modo que $t_i \subseteq A$. A regra de associação é uma implicação da forma $LHS \Rightarrow RHS$, em que $LHS \subseteq A$, $RHS \subseteq A$ e $LHS \cap RHS = \phi$. A regra $LHS \Rightarrow RHS$ ocorre no conjunto de transações T com confiança *conf*, se em *conf*% das transações de T , em que ocorre LHS , também ocorre RHS . A regra $LHS \Rightarrow RHS$ tem suporte *sup*, se em *sup*% das transações em T ocorre $LHS \cup RHS$.

Na implicação $LHS \Rightarrow RHS$, LHS é o lado esquerdo da regra ou antecedente e RHS é o lado direito da regra consequente. Portanto, LHS e RHS indicam causa e efeito, respectivamente, e podem ser interpretados como uma expressão lógica do tipo “se LHS então RHS ”.

Dada a regra $X, Y \Rightarrow Z$, em que $\{X, Y\} \in LHS$ e $\{Z\} \in RHS$ seguem as definições: *item*: um par atributo-valor contido nas transações. X , Y e Z são *itens* que compõem a regra; *itemset*: conjunto de *itens* ordenado lexicograficamente. O conjunto de *itens* X , Y determina um *itemset* e o *item* Z determina outro *itemset*; *K-itemset*: conjunto de k *itens* ordenados lexicograficamente. Os conjuntos X , Y e Z formam um *3-itemset*.

O algoritmo gerador de regras de associação e sua interpretação pelo usuário dependem do uso de medidas objetivas (Agrawal et al., 1993; Brin et al., 1997):

- Suporte: incidência de um *itemset* (Equação 1) ou de uma regra no conjunto de dados (Equação 2). Dado o *itemset* X , $n(X)$ é o número de transações nas quais X ocorre e N o número total de transações na base de dados. O suporte indica a frequência com que X ou $LHS \cup RHS$ ocorre no conjunto de dados.

$$\text{Equação 1: } \text{sup}(X) = \frac{n(X)}{N} \times 100$$

$$\text{Equação 2: } \text{sup}(LHS \Rightarrow RHS) = \text{sup}(LHS \cup RHS) = \frac{n(LHS \cup RHS)}{N} \times 100$$

- **Confiança:** frequência com que *LHS* e *RHS* ocorrem juntos, em relação ao número total de transações em que *LHS* ocorre (Equação 3), sendo $n(LHS)$ o número total de transações nas quais *LHS* ocorre. Indica o poder descritivo da regra (Se $conf = 1$, então todas as vezes em que ocorre *LHS*, também ocorre *RHS*).

$$\text{Equação 3: } conf(LHS \Rightarrow RHS) = \frac{\sup(LHS \cup RHS)}{\sup(LHS)} = \frac{n(LHS \cup RHS)}{n(LHS)} \times 100$$

- **Lift:** proporção de exemplos cobertos simultaneamente por *LHS* e *RHS* (suporte) em relação à *LHS* e *RHS* (Equação 4). Na equação, f significa frequência e P , probabilidade associada. Indica o grau de inesperabilidade e novidade da regra (Se $lift = 1$, então as variáveis são independentes).

$$\text{Equação 4: } lift = \frac{f(RHS | LHS)}{f(RHS)} = \frac{P(LHS \cap RHS)}{P(LHS) \times P(RHS)}$$

O algoritmo, *a priori*, é um modelo computacional clássico, fundamentado no princípio da antimonotonicidade do suporte, capaz de encontrar conjuntos de itens frequentes e gerar regras de associação (Agrawal & Srikant, 1994) utilizando duas etapas:

1. Encontra *k-itemsets* com suporte maior ou igual ao suporte mínimo (sup_min), especificado pelo usuário. Os *itemsets* com sup^3sup_min são definidos *itemsets* frequentes, os demais são *itemsets* não-frequentes;

2. Gera regras de associação para os *k-itemsets* frequentes ($k > 2$). Para cada *itemset* frequente $l \subseteq A$, encontra os subconjuntos \tilde{a} de *itens* de l , não vazios e diferentes de l . Para cada subconjunto, $\tilde{a} \subseteq l$ gera uma regra na forma $\tilde{a} \Rightarrow (l - \tilde{a})$ se a razão de $sup(l)$ por $sup(\tilde{a})$ é maior ou igual à confiança mínima ($conf_min$), definida pelo usuário.

O processo de mineração de dados é dividido em cinco etapas (Rezende et al., 2003): identificação do problema, pré-processamento, extração de padrões, pós-processamento e utilização do conhecimento. As três intermediárias (foco deste trabalho) são iterativas, ou seja, um ciclo repetido inúmeras vezes até a aquisição do modelo do conhecimento desejável. Essas etapas são descritas e já identificadas atividades realizadas com a base de dados do programa Nelore Brasil.

Na fase de identificação do problema, procurou-se identificar forças de mercado que regem a comercialização de touros que podem auxiliar o programa Nelore Brasil na gestão de relação com o cliente (CRM – *customer relationship management*). Segundo Anderson & Kerr (2001), o CRM ajuda as organizações a angariar e fidelizar clientes pela análise de suas necessidades e expectativas, possibilitando a determinação de estratégias de *marketing*.

Dentro do contexto de estudo, para análise do CRM, podem ser definidos: ponto de venda (fazenda), produto (touro), tipos do produto (categoria, variedade, idade), qualidade do produto (DEP e MGT¹ - perfil genético). Seguem hipóteses e soluções propostas para análise de CRM utilizando regras de associação:

- **Predileção de produtos comercializados por ponto de venda (PDV):** regras com *2-itemsets*, fazenda (antecedente) e DEP ou MGT (consequente), indicam o perfil genético dos touros comercializados;

- **Força de mercado do PDV:** regras com *2-itemsets*, DEP ou MGT (antecedente) e fazenda (consequente), indicam predomínio da fazenda no mercado, segundo o perfil genético dos touros comercializados;

- **Nichos de mercado para produtos específicos:** alguns produtos possuem menores escalas de venda, como certas categorias de touros. Regras com *2-itemsets*, categoria (antecedente) e fazenda, região, DEP ou MGT (consequente), indicam nichos de mercado para estes touros;

- **Cesta de produtos:** regras com *3-itemsets*, fazenda (antecedente) e duas DEP ou MGT (consequente), verificam associação entre as principais características dos touros comercializados;

- **Análise de exceções:** diferentes tipos de produtos podem ser comercializados de forma distinta, como variedades de touros. Regras com *3-itemsets*, DEP ou MGT e variedade (antecedente) e fazenda (consequente), determinam se as fazendas comercializam touros com diferentes perfis genéticos, dependendo da variedade;

- **Evolução do mercado:** regras com *2-itemsets*, ano (antecedente) e DEP ou MGT (consequente), são úteis para analisar mudanças de exigências dos clientes pelo perfil genético dos touros. Há algoritmos específicos para regras de associação em padrões sequenciais (Agrawal & Srikant, 1995) e regras de associação periódicas (Özden et al., 1998), que não foram objetivos neste trabalho. As análises temporais nesta pesquisa visam, apenas, verificar perfis genéticos comercializados em determinados anos e meses com maiores movimentações comerciais das fazendas, não caracterizando tendências temporais.

Na fase de pré-processamento, foram utilizadas funções relacionadas a captação, organização, tratamento e preparação dos dados necessários para atender aos objetivos da análise e execução do algoritmo (Goldschmidt & Passos, 2005), foi utilizado o SAS Interprise Guide (SAS, 2002).

Para análise de CRM, só podem ser consideradas fazendas que declaram corretamente a venda do animal,

¹Mérito genético total (MGT): índice do programa Nelore Brasil (Lôbo et al., 2008a).

representada pelos atributos última situação do animal (USA) e data da USA (Dt_USA). Foram selecionados fazendas ativas, que ingressaram no programa Nelore Brasil até 2002, e touros comercializados com menos de 20 anos de idade. Houve considerável redução entre o arquivo oriundo da avaliação genética e o enviado para análise (Tabela 1), adequando-se às capacitações do sistema utilizado para extração de padrões.

Os atributos analisados foram fazenda (NFA), região (REG), classe de idade do animal à USA (USA.C), ano de atualização da USA (USA.A), mês de atualização da USA (USA.M), categoria (CAT), variedade (VAR), mérito genético total (MGT), DEP para efeito materno no peso aos 120 dias de idade (MP120), efeito direto no peso aos 365 dias de idade (DP365), perímetro escrotal (DPE365), idade ao primeiro parto (DIPP), longevidade (DSTAY), área de olho-de-lombo (DAOL), espessura de gordura na costela (DEG).

A USA.C foi discretizada em bezerro ($idade \leq 240$ dias), garrote ($240 < idade \leq 550$ dias), touro jovem ($550 < idade \leq 1100$ dias) e touro erado ($idade > 11000$ dias). As diferenças esperadas na progênie e o mérito genético total foram discretizados pelo percentil (Lôbo et al., 2008a) em top_25 (25% melhores), top_50 (50% melhores) e bottom_50 (50% piores).

Na fase de extração de padrões, utilizou-se o Sistema I₂E (Yamamoto et al., 2007), que é formado pelo Módulo I₂E e pelo Módulo RulEx. O Módulo I₂E possui uma interface gráfica usuário-dirigida para seleção de *itemsets*, segundo objetivo da análise, aumentando a interatividade do usuário com o processo de mineração de dados, enquanto o RulEx possui uma interface para exploração de regras. Esse tipo de sistema é adequado para o problema identificado, em que usuários conhecem o domínio da aplicação (programa Nelore Brasil) e utilizam seus conhecimentos prévios (Geneticistas) para dirigir o processo de mineração de dados (Yamamoto et al., 2008).

Após testes preliminares, dependendo das soluções propostas para análise de CRM, optou-se por utilizar três valores distintos para suporte mínimo, gerando conjuntos de regras que satisfaziam à necessidade de atributos e classes presentes: sup_min= 1974 (10,00%), sup_min= 998 (5,01%) e sup_min= 768 (3,89%). Foram selecionadas regras com 2-*itemsets* e 3-*itemsets*. Para cada solução

proposta, foi salvo um arquivo contendo o conjunto de regras selecionadas.

Na fase de pós-processamento, os arquivos contendo o conjunto de *items* selecionados foram analisados com o Módulo RulEx, que permite visualização e seleção interativa das regras geradas com os respectivos *itemsets*. Foram selecionadas regras que satisfaziam às soluções propostas para análise de CRM, independente dos valores de suporte, confiança e *lift*.

O princípio de Pareto, aplicado ao âmbito da avaliação de regras de associação, diz que apenas 20% das regras são responsáveis por 80% do valor acumulado de uma medida objetiva (Melanda, 2005). Para o conjunto contendo todas as regras induzidas pelo I₂E, com 2-*itemsets* e 3-*itemsets*, foram gerados três gráficos (Figuras 1, 2 e 3) referentes ao valor da medida objetiva (suporte, confiança e *lift*) em função da regra (em ordem decrescente para o valor da medida objetiva). A distribuição de Pareto é representada por um gráfico com formato característico de decaimento inicial acentuado. Segundo diagnóstico visual dos gráficos, elegeu-se qual destas três medidas objetivas representa melhor a distribuição de Pareto (Evans et al., 1993).

Utilizou-se a medida objetiva eleita na validação de regras, ou seja, valor mínimo para as mesmas serem consideradas válidas. Para esta medida, foi montado um gráfico (Figura 4) da soma cumulativa dos valores da medida para cada regra. O ponto de corte foi determinado, conforme princípio de Pareto, em 80% do valor acumulado. São consideradas válidas, todas as regras com medida objetiva eleita maior ou igual ao ponto de corte. Os gráficos foram plotados e analisados com o uso do programa SAS Interprise Guide (SAS, 2002).

Resultados e Discussão

Como resultado do processamento com o Sistema I₂E (Tabela 2), foram induzidas 2.594 regras. Utilizando-se da seleção de *itemsets* usuário-dirigido (Módulo I₂E) e posterior seleção de regras (Módulo RulEx), obtiveram-se 96 regras válidas, que satisfazem às soluções propostas pelas análises de CRM, redução de 96,6%.

Os gráficos das medidas objetivas em função da regra sugerem formatos característicos de decaimento inicial acentuado para suporte e *lift* (Figuras 1 e 3) e decaimento linear, para confiança (Figura 2). Apenas, o primeiro formato é indicativo da distribuição de Pareto (Evans et al., 1993) e a confiança é descartada de uso como filtro para regras válidas. Como o suporte já é utilizado como pré-filtro durante a fase do processamento, o *lift* foi eleito como medida para filtro do conjunto de regras selecionadas.

Tabela 1 - Pré-pocessamento

Número	Arquivo de avaliação genética	Arquivo analisado	Redução
Registros	1.190.045	19.736	98,3%
Atributos	103	15	85,4%
Fazendas	413	17	95,9%

Tabela 2 - Número de regras induzidas, selecionadas (redução em relação às induzidas) e válidas (redução em relação às induzidas e selecionadas)

Análise da gestão de relação com o cliente (CRM)	Regra		
	Induzidas	Selecionadas ¹	Válidas ²
Predileção de produtos comercializados por PDV ³	306	6 (98,0%)	5 (98,4%;15,7%)
Força de mercado do PDV	306	6 (98,0%)	5 (98,4%;15,7%)
Nichos de mercado de touros para categoria	563	43 (92,4%)	18 (96,8%;58,1%)
Cesta de produtos por PDV	1184	24 (98,0%)	23 (98,1%;4,2%)
Análise de exceções para variedade	206	27 (86,9%)	20 (98,3%;25,9%)
Evolução do mercado	29	29 (0,0%)	18 (37,9%;37,9%)
Total	2594	135 (94,8%)	89 (96,6%;34,1%)

¹ Consideradas úteis pelo usuário; ² lift = 1,0505; ³ Ponto de venda.

O gráfico, da soma cumulativa dos valores do lift para cada regra (Figura 4), possui inclinação inicial muito acentuada, modificando a relação clássica de Pareto (80/20). Segundo Melanda (2005), neste caso deve-se assumir o valor cumulativo de 80% da distribuição de Pareto. Foi considerada a relação aproximada 80/30, em que, 80,0% da representatividade do lift são selecionadas 30,2% das regras. O valor de corte foi lift = 1,0505. Reduziu-se 96,6% de regras induzidas e 34,1% de regras selecionadas, sendo consideradas 89 regras válidas (Tabela 2).

A redução do número de regras (Tabela 2) sugere que as metodologias empregadas nas fases de extração de padrões (Yamamoto et al., 2007) e pós-processamento (Melanda, 2005) estão adequadas ao propósito de facilitar a identificação de regras válidas e potencialmente úteis, pelo usuário.

Padrões de comportamento mercadológico de touros com avaliação genética, descritos na identificação do problema, podem ser identificados no conjunto de regras válidas (Tabelas 3 a 6), de acordo com os valores de confiança e lift.

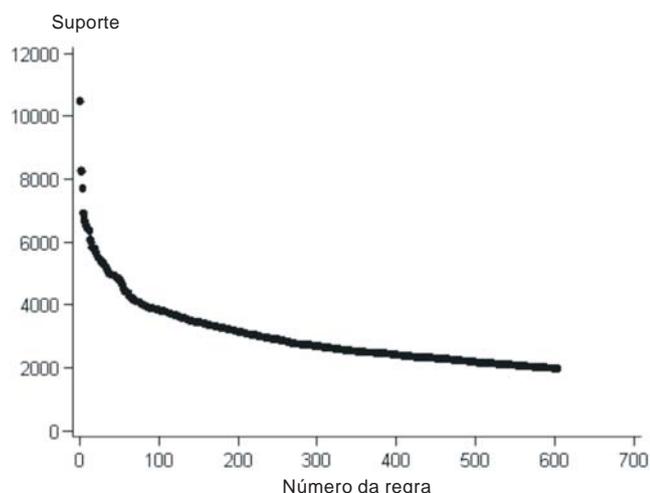


Figura 1 - Valor do suporte em função da regra (ordem decrescente).

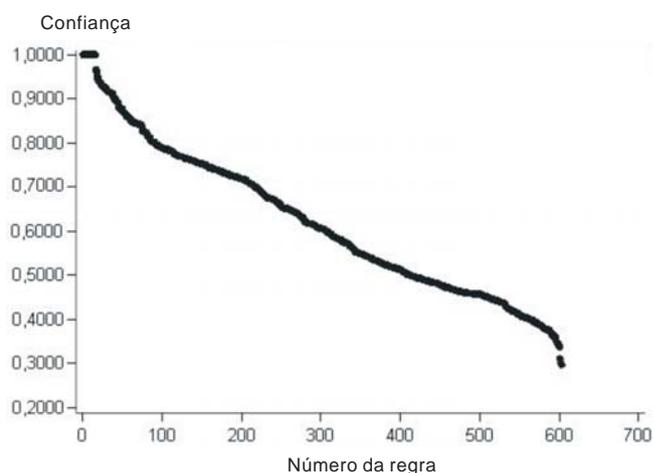


Figura 2 - Valor da confiança em função da regra (ordem decrescente).

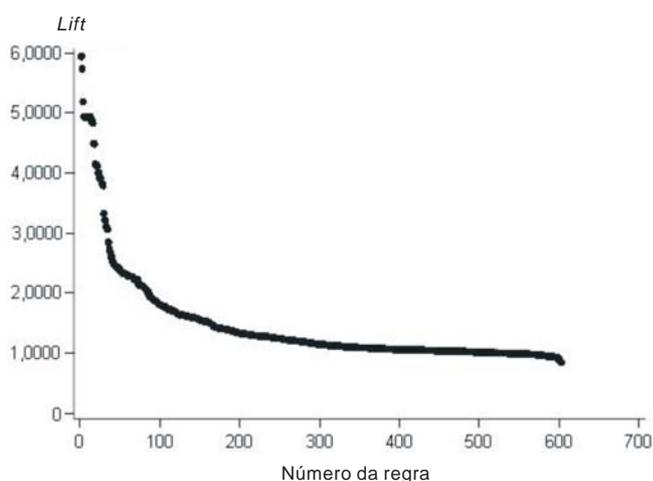


Figura 3 - Valor do lift em função da regra (ordem decrescente).

Os valores de confiança para *Regra 2* e *Regra 3* (Tabela 3) significam que 61,68% dos touros comercializados pela fazenda E são *top_25* para MGT (harmonicamente balanceados para habilidade materna, ganho de peso, fertilidade e precocidade sexual) e 52,40%

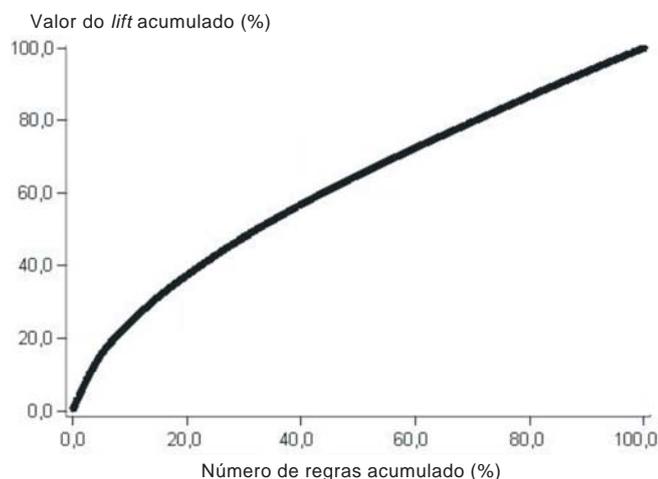


Figura 4 - Soma cumulativa dos valores do lift para cada regra.

são *top_25* para DAOL (rendimento de carcaça). A confiança indica relações que ocorrem dentro da própria fazenda, ou seja, embora ela seja especializada no comércio de touros bons para ambas as características, há uma maior especialização para MGT. A *Regra 1* sinaliza que a fazenda B é especializada no comércio de touros *top_25* para DSTAY.

O lift da *Regra 3* (Tabela 3) indica relação que ocorre entre todas as fazendas. Dentre os touros comercializados, *top_25* para DAOL, há 2,1196 chances a mais de serem da fazenda E, do que não serem da respectiva fazenda. O lift da *Regra 3* maior que da *Regra 2* sugere maior novidade, o comércio de touros bons para rendimento de carcaça que touros bons para MGT. Por ser uma situação inesperada para o consumidor, a fazenda E poderia utilizar a respectiva informação em campanhas de marketing, para angariar novos

clientes. A interpretação dos valores de confiança e lift é idêntica para as outras regras.

As Regras 4, 5 e 6 (Tabela 3) sugerem que a fazenda B domina o mercado de touros *top_25* para DSTAY (longevidade) e a fazenda E, touros bons para MGT e rendimento de carcaça. A *Regra 6* chama atenção, pois 42,94% dos touros *top_25* para DAOL são provenientes da fazenda E, o que sugere enorme força de mercado e capacidade de angariar clientes que necessitam investir em touros capazes de transmitir bons rendimentos de carcaça.

Como o programa Nelore Brasil é dedicado a animais registrados, o mercado de touros é dominado pela categoria PO (Puros de origem - registro genealógico com toda a ascendência conhecida), observado em 14170 (71,8%) transações da base de dados analisada, portanto faz-se necessário analisar nichos de mercado para outras categorias. Na categoria PI ou POI (Puros de origem importado - animais registrados com toda a ascendência importada), a $conf = 1,0000$ (Tabela 4 - *Regra 7*) sugere que apenas a Região Sudeste do Brasil é propensa à comercialização de touros para respectiva categoria.

A Região Norte (Tabela 4 - *Regra 11*) possui grande propensão ao comércio de touros da categoria LA (livro aberto - animais registrados que não possuem toda a genealogia conhecida). Os valores idênticos de confiança e lift nas Regras 10 e 11 esclarecem a força de mercado da fazenda E para touros LA. Há uma parcela considerável de touros comercializados para a categoria LA bons para rendimento de carcaça (*Regra 13*), apresentando $conf = 0,3672$. Uma novidade é a participação de *touros erados* para LA em relação ao total de touros comercializados (*Regra 12*), demonstrados pelo $lift = 1,3624$.

Tabela 3 - Predileção de produtos comercializados por ponto de venda (PDV) e sua força de mercado

Regra	Antecedente	Conseqüente	Suporte	Confiança	Lift
1	NFA_b	DSTAY_top25	2159	0,7865	1,9347
2	NFA_e	MGT_top25	2466	0,6168	1,6159
3	NFA_e	DAOL_top25	2095	0,5240	2,1196
4	DSTAY_top25	NFA_b	2159	0,2691	1,9347
5	MGT_top25	NFA_e	2466	0,3274	1,6159
6	DAOL_top25	NFA_e	2095	0,4294	2,1196

Tabela 4 - Nichos de mercado de touros para categoria

Regra	Antecedente	Conseqüente	Suporte	Confiança	Lift
7	CAT_pi	REG_se	1118	1,0000	2,8810
8	CAT_cl	NFA_a	1242	1,0000	8,4700
9	CAT_cl	REG_co	1242	1,0000	2,2207
10	CAT_la	NFA_e	1842	0,5747	2,8370
11	CAT_la	REG_n	1842	0,5747	2,8370
12	CAT_la	USA.C_touroerado	1382	0,4312	1,3624
13	CAT_la	DAOL_top25	1177	0,3672	1,4854

Por outro lado, a Região Centro-Oeste é mais propensa ao comércio de touros da categoria CL (Cara Limpa - animais sem registro genealógico), apresentando 100% dos animais comercializados (Tabela 4 - *Regra 9*). O *lift* da *Regra 8* maior que da *Regra 9* indica uma novidade observada, a fazenda A domina o comércio para respectiva categoria, em comparação com todo o conjunto de fazendas analisadas.

A Região Sudeste possui terras mais caras e, conseqüentemente, pecuária mais tecnificada que as Regiões Centro-Oeste e Norte (FNP, 2007). As regras confirmam o grau de tecnificação, demonstrando que não há nicho de mercado para touros das categorias CL e LA na Região Sudeste. Nas outras regiões, que concentram a pecuária de corte brasileira, touros com bom perfil genético são propensos à comercialização, independentemente da categoria.

A análise de cesta de produtos (Tabela 5) identificou perfis genéticos agregando duas características (DEP/MGT), para quatro fazendas, com casos favoráveis (*top_25*) e desfavoráveis (*bot_50*) em relação ao perfil. Mais da metade dos animais comercializados pela fazenda A (*Regra 15*), $conf = 0,5086$, é concomitantemente ruim para rendimento e terminação de carcaça, típico caso em que o programa de melhoramento genético deve alertar ao criador para modificar critérios de seleção e ofertar bons produtos a seus clientes. A fazenda E, com $lift = 2,5485$, destaca-se no comércio de animais concomitantemente bons para MGT e rendimento de carcaça (*Regra 16*), denotando sucessos de seleção e comercial, uma vez que o MGT não contempla características quantitativas de carcaça, o PDV agregou mais um benefício aos seus produtos.

A análise de exceções para variedade (Tabela 5), destaca a fazenda B na variedade M, ofertando bons animais para longevidade e a fazenda E na variedade P (Padrão - com

chifres), com bons touros para rendimento de carcaça, dados os valores de *lift* iguais a 2,3670 e 2,1979, respectivamente.

Para análise de evolução de mercado (Tabela 6), foram induzidas regras referentes aos anos de 2003 a 2007. Em 2003, o programa Nelore Brasil lançou DEP para características quantitativas de carcaça (Lôbo et al., 2008a), nesta época a maior parte dos animais comercializados, $conf = 0,5617$, era ruim para acabamento de carcaça (*Regra 19*), porém não havia conscientização do mercado sobre a importância dessas características. Regra contendo DAOL aparece apenas em 2007 (*Regra 21*), com $lift = 2,1301$, sugerindo que o mercado ofertou uma grande quantidade de touros bons para rendimento de carcaça neste ano. O maior valor de confiança, $conf = 0,6982$, também sugere que o MGT tornou-se um importante índice para a comercialização de touros (*Regra 20*).

Esses dois resultados de análise de evolução de mercado são bons, pois o grupo de fazendas analisadas atingiu o objetivo do programa de melhoramento genético, que tem a função de melhorar a produtividade da pecuária, de forma eficiente e sustentável, por meio do desenvolvimento de tecnologias para gestão dos rebanhos, seleção e acasalamentos (Faria et al., 2008).

O resultado da avaliação genética é um rico repositório de dados, que deve ser utilizado para gestão de relação com o cliente, com foco no aumento da competitividade comercial das fazendas, em fornecer touros com potencial genético que a pecuária necessita para tornar-se mais eficiente. Deve haver estímulos do programa Nelore Brasil, para que as fazendas exportem para sua base de dados, atributos referentes à sua atividade comercial, como última situação do animal e data da última situação do animal, possibilitando análises de gestão de relação com o cliente.

Tabela 5 - Cesta de produtos por PDV e análise de exceções para variedade

Regra	Antecedente	Conseqüente	Suporte	Confiança	<i>Lift</i>
15	NFA_a	DAOL_bot50, DEG_bot50	1185	0,5086	1,4500
16	NFA_e	MGT_top25, DAOL_top25	1443	0,3609	2,5485
17	VAR_m,NFA_b	DSTAY_top25	1020	0,9623	2,3670
18	VAR_p,NFA_e	DAOL_top25	1660	0,5434	2,1979

Tabela 6 - Evolução do mercado

Regra	Antecedente	Conseqüente	Suporte	Confiança	<i>Lift</i>
19	USA.A_2003	DEG_bot50	1347	0,5617	1,1205
20	USA.A_2007	MGT_top25	2034	0,6982	1,8293
21	USA.A_2007	DAOL_top25	1534	0,5266	2,1301

Conclusões

Regras de associação são potencialmente úteis para identificar forças de mercado que regem a comercialização de touros. A análise de Pareto contribui para identificar combinações de atributos genéticos, geográficos e temporais que determinam segmentações do mercado quanto à seleção dos rebanhos e ao comércio de touros com avaliação genética. O Sistema I₂E proporciona facilidade e agilidade na obtenção e seleção de regras úteis por usuários que possuem conhecimento do domínio (melhoramento genético), aliviando-o de complexidades computacionais. Enfim, a mineração de dados demonstra potencial para aumentar a eficiência e competitividade do programa Nelore Brasil.

Agradecimentos

Às instituições de fomento à pesquisa CNPq/RHAE, FINEP/BID, PRONEX, FAPESP, CAPES, FAEPA, ANCP, pelo apoio financeiro. Às instituições de ensino e pesquisa FMRP-USP e ICMC-USP, pela excelência em programas de pós-graduação. Aos laboratórios de informática GEMAC, LABIC e LCAD, pelo suporte de recursos humanos e tecnológicos.

Referências

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A.N.; Mining association rules between sets of items in large databases. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 7., 1993, Washington. **Anais...** New York: ACM SIGMOD, 1993. p.207-216.
- AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago de Chile. **Proceedings...** San Francisco: Morgan Kaufmann, 1994. p.487-499.
- AGRAWAL, R.; SRIKANT, R. Mining Sequential Patterns. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 9., 1995, Taipei. **Anais...** Washington: IEEE Computer Society, 1995. p.3-14.
- ANDERSON, K.; KERR, C. **Customer relationship management**. New York: McGraw-Hill, 2001. 164p.
- BOLDMAN, K.G.; VAN VLECK, L.D.; KACHMAN, S. et al. **MTDFREML – User's guide**. 3.ed. Washington: USDA-ARS, 1995. 115p.
- BRIN, S.; MOTWANI, R.; ULLMAN, J.D. et al. Dynamic itemset counting and implication rules for market basket data. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 8., 1997, Tucson. **Anais...** New York: ACM SIGMOD, 1997. p.255-264.
- CENTRO TÉCNICO DE AVALIAÇÃO GENÉTICA - CTAG [2007]. **Software CTAG**. Ribeirão Preto. Disponível em: <<http://www.ctag.com.br>>. Acesso em: 07/09/2008.
- EVANS, M.; HASTINGS, N.; PEACOCK, B. **Statistical distributions**. New York: J. Wiley, 1993. p.151-154.
- FARIA, C.; MAGNABOSCO, C.; LÔBO, R.B. et al. Impactos dos acasalamentos otimizados na produtividade dos rebanhos bovinos de corte. In: LÔBO, R.B.; FARIA, C. (Eds.) **Princípios e resultados de pesquisas científicas no Programa Nelore Brasil**. Ribeirão Preto: ANCP, 2008. p.15-26.
- FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH. From data mining to knowledge discovery. **AI Magazine**, Menlo Park, v.17, 1996. p.37-54.
- FNP CONSULTORIA E COMÉRCIO - FNP. **Anuário da pecuária brasileira**. 14.ed. São Paulo: FNP, 2007. 368p.
- GOLDSCHMIDT, R.; PASSOS, E. **Data Mining: um guia prático**. Rio de Janeiro: Campus, 2005. 261p.
- LÔBO, R.B.; BEZERRA, L.A.F.; OLIVEIRA, H.N. et al. **Avaliação genética de touros e matrizes da raça Nelore**: Sumário 2008. 14.ed. Ribeirão Preto: ANCP, 2008a. 124p.
- LÔBO, R.B.; NOMEINI, J.; OLIVEIRA, H.N. et al. Programa Nelore Brasil em números. In: LÔBO, R.B. ; FARIA, C. (Eds.) **Princípios e resultados de pesquisas científicas no Programa Nelore Brasil**. Ribeirão Preto: ANCP, 2008b. p.15-26.
- MELANDA, E.A. **Pós-processamento em regras de associação**. 2005. 225f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, São Carlos.
- NOMEINI, J. **Aplicação de metodologias de extração de conhecimentos em pesquisas e gerenciamento de programa de melhoramento genético em bovinos da raça Nelore**. 2006. 113f. Dissertação (Mestrado em Ciências Biológicas) - Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto.
- ÖZDEN, B.; RAMASWANY, S.; SILBERSCHATZ. Cyclic association rules. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 14., 1998, Orlando. **Anais...** Washington: IEEE Computer Society, 1998. p.412-421.
- REZENDE, S.O.; PUGLIESI, J.B.; PAULA, M.F. et al. Mineração de dados. In: REZENDE, S.O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Manole, 2003. p.307-335.
- SERVIÇO DE INFORMAÇÃO DA CARNE - SIC. **Para que serve um boi?** [2009]. São Paulo: SIC. Disponível em: <<http://www.sic.org.br/praqserve.asp>>. Acesso em: 16/11/2009.
- STATISTICAL ANALYSIS SYSTEM - SAS. **SAS Enterprise guide**. Versão 2.0. Cary: SAS Institute Inc., 2002. (CD-ROM).
- VAN VLECK, L.D. **Selection index and introduction to mixed model methods for genetic improvement of animals: the green book**. Florida: CRC Press, 1993. p.255-266.
- YAMAMOTO, C.H.; OLIVEIRA, M.C.F.; REZENDE, S.O. et al. An itemset-driven cluster-oriented approach to extract compact and meaningful sets of association rules. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS, 6., 2007, Cincinnati. **Anais...** Los Alamitos: IEEE Computer Society, 2007. p.87-92.
- YAMAMOTO, C.H.; REZENDE, S.O.; NOMEINI, J. et al. Including the user in the knowledge discovery loop: Interactive itemset-driven rule extraction. In: SYMPOSIUM ON APPLIED COMPUTING, 23., 2008, Fortaleza. **Anais...** New York: ACM, 2008, p.1212-1217.