

SiPreCE –UM SISTEMA DE PREDIÇÃO DE CÓLICA EM EQUÍNOS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

Adriano D. Pila¹, Andréa C. A. Borim², Marcelo Zani³, Márcia R. F. M. Júlio⁴

¹ Cientista da Computação, Prof. Adjunto, Faculdade Comunitária de Santa Bárbara, Sta Bárbara do Oeste – SP, Fone: (019)34638456, adriano.pila@unianhanguera.edu.br

² Engenheira Eletricista, Profa Titular, Faculdade Comunitária de Campinas, Campinas – SP

³ Cientista da Computação, Prof. Adjunto, Centro Universitário Anhanguera, Leme - SP

⁴ Cientista da Computação, Prof. Adjunto., Faculdade Comunitária de Limeira, Limeira - SP

XXXV Congresso Brasileiro de Engenharia Agrícola
31 de julho a 4 de agosto de 2006 - João Pessoa - PB

RESUMO: Este artigo tem o objetivo de mostrar como os recursos computacionais podem auxiliar no processo de predição e controle de cólica em equínos. A cólica em equínos é um problema de fácil tratamento, desde que diagnosticado a tempo. Entretanto, o diagnóstico precoce nem sempre é factível em tempo hábil, o que leva o a complicações. Em casos extremos e não raros o animal morre ou tem que ser sacrificado. Isso acarreta um dano financeiro considerável aos criadores de equínos que vêm seus investimentos serem frustrados e sem qualquer expectativa de retorno. Este trabalho apresenta um sistema de predição de cólica em equínos denominado *SiPreCE*. Esse sistema utiliza técnicas de Inteligência Artificial, notadamente Aprendizado de Máquina, para a análise de informações coletadas com os criadores e posterior indução de um sistema preditivo de classificação capaz de prever o destino do animal com base nos dados coletados e confrontados com casos anteriores. Dessa forma, indícios iniciais que podem levar o animal à morte podem ser prontamente tratados e solucionados. Assim, neste trabalho é analisada uma base de dados contemplando 368 casos sobre cólicas em equínos, cujo sistema proposto é capaz de apontar situações favoráveis e desfavoráveis à sobrevivência do animal.

PALAVRAS-CHAVE: cólica em equínos, aprendizado de máquina, inteligência artificial.

ABSTRACT: The purpose of this paper is to show how computer system can help the prediction and control processes in the case of horse colic. Horse colic is an easy treatment problem if diagnosed just in time. However, early diagnostic is not always easy to get just in time, what takes the animal into a serious pain situation. In most cases, the exams to detect the colic are invasive and provoke a lot of pain in the animal. In extreme and not rare cases the animal dies or must be sacrificed. This cause a considerable financial injury to the horse creators that see their investments being frustrated and without any expectative to get the money back. This work presents a horse colic prediction system called *SiPreCE*, which in portuguese stands for Horse Colic Prediction System. This system uses Artificial Intelligence techniques, actually Machine Learning, to analyze the information collected with the creators help and later induction of a classification prediction system capable of predict the animal destiny based on the data collected and compared with the early real cases. In this way initial symptoms that may lead the animal to death can be treated and solved quickly. Thus, in this work it is analyzed a dataset with 368 cases about horse colic which proposed system is capable to point out situations where the animal may live or die.

KEYWORDS: horse colic, machine learning, artificial intelligence.

INTRODUÇÃO: A crescente coleta de informações das mais variadas áreas abriu margem ao questionamento da importância e, principalmente, da necessidade e utilidade das informações armazenadas. Nesse cenário a computação aplicada com o uso de técnicas de Inteligência Artificial, mas precisamente a área de Aprendizado de Máquina, busca a análise inteligente das informações de forma a construir modelos preditivos que possam ser úteis na tomada de decisão futura ou para dar suporte a uma decisão já tomada. Os modelos preditivos são construídos com base de informações de casos reais, de forma que o comportamento de novos casos possa ser previsto com o uso desses modelos. O presente projeto tem como objetivo a construção de um sistema preditivo de apoio à decisão aqui nomeado *Sistema de Predição sobre Cólica em Equínos – SiPreCE*. Esse sistema será capaz de dada algumas informações sobre o estado atual do animal se a atitude a ser tomada é uma intervenção cirúrgica, além de

prever se é iminente a entrada do animal em quadro de cólica para que sejam tomadas atitudes preventivas. O SiPreCE poderá ser usado como ferramenta de auxílio ao estudo desta síndrome, podendo ser usado desde por estudantes de veterinária para a simulação de diagnóstica e tratamento como para análise estatística.

MATERIAL E MÉTODOS: O acúmulo de informações nos repositórios de dados cresce vertiginosamente e meios efetivos de análise desses dados são necessários. Aprendizado de Máquina é uma área da Inteligência Artificial que explora bases de dados (exemplos) com o propósito de extrair conhecimento desses exemplos.

O SiPreCE é capaz de dada algumas informações sobre o estado atual do animal se a atitude a ser tomada é uma intervenção cirúrgica, além de prever se é iminente a entrada do animal em quadro de cólica para que sejam tomadas atitudes preventivas.

O sistema foi construído seguindo o modelo de prototipação (Somerville, 2000; Pressman, 2004). Primeiramente foi desenvolvido um esboço do sistema de apoio à decisão de forma que os possíveis usuários possam interagir com o sistema e opinar sobre a usabilidade do mesmo (Nielsen, 1994).

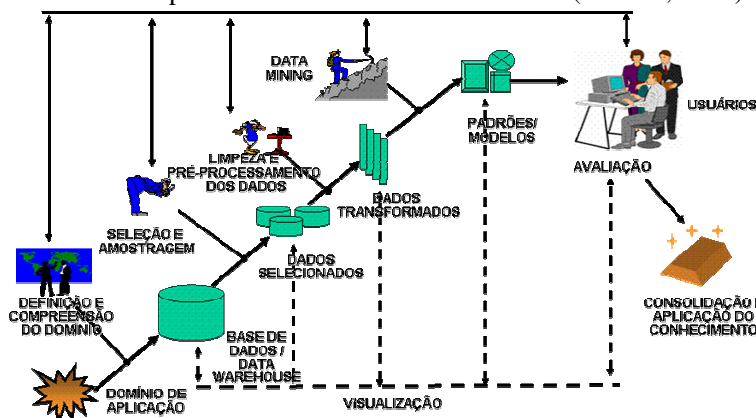


Figura 1: Processo metodológico para o desenvolvimento de um sistema preditivo.

Quanto ao sistema preditivo propriamente dito foram seguidos os passos representados pela Figura 1 que ilustra o processo de desenvolvimento de um sistema preditivo. O processo formado de 6 passos: entendimento do domínio, entendimento do sistema de informação, preparação dos dados, a extração do modelo preditivo, entendimento do conhecimento e aplicação do conhecimento.

Na etapa de entendimento do domínio um especialista que entenda da área cujas informações foram colhidas auxilia o profissional que trabalhará sobre os dados. Assim um médico veterinário fez a análise dos dados coletados para 368 casos sobre cólicas em equínos.

A próxima etapa trata do entendimento das informações e se elas foram coletadas de forma adequada.

Estima-se que aproximadamente 80% desse processo estejam concentrados na terceira etapa, que é a preparação das informações. As informações nem sempre são coletadas de forma confiável, principalmente porque as pessoas que alimentam os sistemas de informação são pouco treinadas para esse fim, ou algum dado não foi coletado. A ausência de um dado pode as vezes ter algum significado semântico.

A etapa tida como principal é a extração do modelo preditivo propriamente dito. Nessa etapa são aplicados alguns programas computacionais capazes de, através dos dados, descobrirem relações implicitamente descritas nas informações. É possível ainda descobrir relações de causa efeito ou até mesmo agrupar casos em que ocorra semelhança semântica das informações. Essa etapa resulta um modelo que descreve com certa precisão o conhecimento presente naqueles dados. Várias são as teorias para análise das informações e variados são os modelos. As próximas etapas incluem a avaliação do conhecimento extraído com o auxílio de um especialista do domínio que possa apontar a verdadeira relevância do conhecimento extraído e a aplicação do conhecimento extraído propriamente dito. Para validação do modelo a técnica usada foi a *10-fold-cross-validation*, a qual também é utilizada neste trabalho (Weiss e Kulikowski, 1991; Kohavi, 1995). Nessa técnica os exemplos iniciais são divididos em 10 partições disjuntas, cada qual contendo o mesmo número de exemplos escolhidos aleatoriamente. Então, um classificador é induzido utilizando-se 9 partições de exemplos e suas matrizes de contingência e confusão são calculadas sobre os exemplos contidos na única partição não utilizada. Esse processo é repetido 10 vezes, de forma que todas as 10 partições são utilizadas em separado para calcular as matrizes de contingência e confusão. Ao final do processo, tem-se 10 matrizes que relatam a precisão do classificador e as taxas das regras. A matriz de confusão final é a média aritmética

das 10 matrizes de confusão obtidas nesse processo.

RESULTADOS E DISCUSSÕES: Para o desenvolvimento do sistema preditivo foram coletados dados reais de forma que existem 368 casos armazenados no banco de dados. As informações presentes para cada caso estão distribuídas em 23 características apresentadas a seguir a título de exemplo:

1. Cirurgia: 1 = Sim, foi tratado com cirurgia. 2 = Foi tratado sem cirurgia.
2. Idade: 1 = Cavalo adulto. 2 = Cavalo jovem (< 6 meses).
3. Identificação do cavalo pelo hospital ou veterinário.
4. Temperatura retal: linear e em graus celsius. Temperaturas elevadas podem ocorrer por causa de infecção. A temperatura pode ser reduzida quando o animal esta em choque. A temperatura normal é de 37.8C, mas pode mudar de acordo com alguma lesão ou estado de choque.
5. Pulso: linear e medido em batimentos por minuto. É um reflexo das condições do coração. 30-40 batimentos por minuto é normal num animal adulto. Mesmo em animais atletas o batimento é nessa faixa, raros são os casos entre 20-25. Animais com lesão ou choque circulatório têm alta taxa de batimento cardíaco.
6. Taxa respiratória: linear, sendo a faixa normal entre 8 a 10. A utilidade é duvidosa dada a grande flutuação da informação.
7. Temperatura das extremidades: é uma indicação subjetiva da circulação periférica. Os valores normais são: 1 = Normal; 2 = Quente; 3 = Frio; 4 = Gelado. Extremidades entre frio e gelado indicam possível choque. Temperaturas altas podem estar correlacionadas à uma elevada temperatura retal.

O sistema desenvolvido procura prever com base nas características do animal qual o quadro futuro mais provável. As possibilidades são de que o animal sobreviva ou morra, sendo que a segunda opção pode ser natural ou induzida. Nessa situação existem três possíveis classes a serem preditas com base nas informações colhidas do animal: 1 = viveu, 2 = morreu e 3 = praticada eutanásia.

Os resultados mostram que o sistema proposto ainda possui uma taxa média de erro em torno de 36%. Essa taxa de erro é bastante elevada para considerar o sistema como um classificador a ponto de prever com exatidão qual o quadro futuro para um determinado animal. Entretanto, o conhecimento extraído pelo SiPreCE é inteligível ao usuário, o que possibilita sua análise e uso para a tomada de decisão, senão em todos os casos, pelo menos para a maior parte deles. Ainda, o conhecimento extraído pode ser utilizado para suporte à decisão do especialista.

O uso de conhecimento extraído para suporte à decisão é importante quando o especialista do domínio possui algumas experiências prévias que indicam que em determinadas situações certas decisões devem ser tomadas. Porém, essas decisões baseadas em experiências nem sempre possuem o respaldo acadêmico ou técnico. Sistemas de apoio à decisão ajudam a suprir essa deficiência de suporte técnico.

Utilizando o SiPreCE como um sistema de suporte à decisão, pode-se destacar as seguintes regras (de um total de 114 regras), que expressão conhecimento novo presente nos dados e que são úteis para o especialista do domínio.

```
R00032 SE total_proteinas <= 9
      E lesão_cirurgica = 1
      E abdome = 3
  ENTÃO CLASSE = 3
```

A regra R00032 está relacionada a um quadro futuro no qual o animal morre por eutanásia. Com base nessa regra é possível verificar a associação de fatores que levam o animal ao óbito. Nesse caso, quando as proteínas totais estão abaixo do fator 9 (a faixa de normalidade é 6 – 7.5), o animal sofreu cirurgia e o abdome do animal possui fezes firmes no intestino grosso o quadro sugere que o animal vira a óbito por eutanásia. Vale a pena ressaltar que essa regra é importante também para determinar quando um animal ainda esta fora de risco mas pode vir a entrar num quadro de características que sugiram seu óbito. Por exemplo, se o animal possui total de proteínas menor que 9 e lesão cirúrgica. Daí, conclui-se que se o animal passar a ter o fator 3 para a característica abdome, a regra anterior passa a ser válida e então passar a ser eminente o óbito do animal.

```
R00070 SE total_proteinas > 9
      E dor = 1
      E mucosas = 1
```

ENTÃO CLASSE = 1

A regra R00070 está relacionada a um quadro futuro no qual o animal sobrevive. Nesse caso, embora o total de proteínas seja maior que 9, sugerindo um quadro de desidratação, como não há dor e as mucosas também apresentam aspecto normal, o quadro geral futuro sugere a sobrevivência do animal.

```
R00114 SE total_proteínas > 9
      E dor = 5
      E pulso > 92
  ENTÃO CLASSE = 2
```

E, finalmente, a regra R00114 está relacionada a um quadro no qual o animal vem a óbito naturalmente. Nesse caso, os animais com alta desidratação, dor severa contínua e pulsação elevada, indicando lesão ou choque circulatório, tem tendência a virem a óbito. Mais uma vez, a conjunção de dois desses fatores já deve colocar o criador atendo, pois a soma do outro fator ainda ausente colocará o animal na rota indicativa de não sobrevivência.

CONCLUSÕES: Sistemas preditivos são a nova promessa da área da computação para o auxílio na detecção e prevenção de cenários futuros nas mais diversas áreas do conhecimento. A área escolhida para esse trabalho esta relacionada à predição de sobrevivência ou óbito de eqüinos frente à quadros de cólica.

O problema de cólica em eqüinos remonta a fatores emocionais, mas fundamentalmente financeiros. Os criadores são bastante penalizados quando os animais são levados à óbito pela falta de detecção prévia do quadro que leva àquele fato.

Neste trabalho é apresentado o SiPreCE, que é um sistema preditivo capaz de predizer, com base nas características fisiológicas do animal, qual o possível quadro que o animal assumirá num futuro próximo. Embora o sistema seja capaz de predizer a possibilidade de sobrevivência ou óbito, o erro ainda é bastante grande para ser admissível pela área veterinária.

Entretanto, resultados isolados como os apresentados na seção anterior sugerem que em alguns casos é possível prever o quadro futuro do animal. Ainda, com base nas características atuais do animal é possível monitorá-lo para que outro fator conjunto seja instalado a ponto de levá-lo à óbito, como é o caso do animal com dor severa contínua (dor = 5) e alta desidratação (total_proteínas > 9). As regras utilizadas na análise dos resultados são apenas um subconjunto de um total de 114 regras que expressão o conhecimento extraído dos exemplos de casos reais.

Embora o não seja possível utilizar o SiPreCE como um sistema preditivo, os resultados isolados destinados ao suporte à decisão foram satisfatórios e uma investigação mais aprofundada poderá levar a resultados mais consistentes.

REFERÊNCIAS

- Baranauskas, J. A. & Monard, M. C. (2000a). Reviewing some machine learning concepts and methods. Technical Report 102, ICMC-USP. ftp://ftp.icmc.sc.usp.br/pub/BIBLIOTECA/rel_tec/RT_102.ps.zip.
- Bernardini, F. C. (2002). Combinação de classificadores simbólicos para melhorar o poder preditivo e descritivo de Ensembles. Dissertação de Mestrado, ICMC-USP.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AIMagazine*, Fall:37–54.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34.
- Freitas, A. A. (1998a). A multi-criteria approach for the evaluation of rule interestingness. In *Proceedings of the International Conference on DataMinig*, pages 7–20, Rio de Janeiro, RJ.