Horse-Expert: An aided expert system for diagnosing horse diseases

H. Qin, J. Xiao, X. Gao, H. Wang

Department of Veterinary Surgery, Northeast Agricultural University, Harbin, Heilongjiang 150030, PR China

Abstract

In contrast to the rapid development of the horse husbandry in China, the ability of horse veterinarians to diagnose diseases has not been improved and only a few domain experts have considerable expertise. At present, many expert systems have been developed for diseases diagnosis, but few for horse diseases diagnosis have been studied in depth. This paper presents the design and development of a computer-aided expert system for diagnosing horse diseases. We suggest an approach for diagnosis of horse diseases based on the analysis of diagnostic characteristics and the experiential knowledge of domain experts. It is based on using evidence-weighted uncertainty reasoning theory, which is a combination of evidence theory and an uncertainty pass algorithm of confidence factors. It enables drawing of inferences with atypical clinical signs and the uncertainty of the user's subjective understanding. It reduces the influence of subjective factors on diagnostic accuracy. The system utilizes a user friendly interface for users and requests a confidence factor from users when feedback is given to the system. Horse-Expert combines the confidence factors with weight factors assigned to clinical signs by experts during the knowledge acquisition process to make diagnostic conclusions. The system can diagnose 91 common horse diseases, and provides suggestions for appropriate treatment options. In addition, users can check the medical record through statistical charts. The system has been tested in seven demonstration areas of Xinjiang province in northwestern China. By constantly maintaining and updating the knowledge base, the system has potential application in veterinary practice.

Key words: horse diseases, aided diagnosis, uncertainty reasoning, evidence weighted, confidence factor

Introduction

The horse husbandry in China began developing about 5,000 years ago. Horses in ancient times were instrumental to agricultural production, transportation and military use (Creel 1965, Deng et al. 2014), however, the role of horses in modern day society is dwindling due to the development of technology and power machinery.

According to official government statistics (www.stats.gov.cn/), China’s horse breeding stock population was 6,043,000 in 2014. China is also one
of the world’s largest consumers of horses. Horse health status is associated with industrial development and economic loss for horse breeders (Padalino 2015). Prevention and control of horse disease is a serious challenge that is probably as old as horse breeding itself in China. Horse husbandry is developing rapidly, but experts with extensive experience are scarce. It is difficult to solve this problem in a short time frame simply by relying on cultivating personnel who can undertake clinical diagnosis and treatment. This is a serious hindrance to the healthy and sustainable development of the horse husbandry. Many expert systems have been developed for disease diagnosis and have proven to be effective (Li et al. 2002, Duan et al. 2003, Zetian et al. 2005, Devraj et al. 2011, Cui et al. 2012, Munirah et al. 2016), however, for diagnosing horse diseases have not yet been studied in-depth. Nevertheless most existing expert systems fully account for the uncertainty of conclusions, and do not consider the uncertainty of the evidence (Kolhe et al. 2011). In this paper, we designed an evidence-weighted uncertainty reasoning method to address the widespread existence of uncertainty in the process of diagnosing horse diseases, and reduced the influence of the subjective factor on the diagnosis accuracy. It provides online support for horse farmers and extension workers. Hence, the aim of the study was to design a computer-aided diagnostic tool for horse veterinarians and to improve the present situation of horse diseases diagnosis in China.

Horse disease diagnosis characteristic analysis

A common problem in diagnosing disease is that animals cannot express their feelings, and diagnosing disease is based on the expression of symptoms. For example, if a horse lies on its back with its legs tucked, it may be indicated that the horse is suffering from colic. Moreover, in most cases, there are timeliness and regional factors with obvious influences on the development of horse disease (Zeldis et al. 2000). As large animals, horses generally do not show obvious clinical symptoms when suffering from mild disease, as a result, horse diseases are commonly misdiagnosed. Despite improvements in horse disease diagnosis in recent years, most veterinarians have not extensive experiences, resulting in serious losses from delayed diseases control. Improving the quality of veterinary education and clinical diagnosis accuracy, and reducing losses caused by diseases are the most critical problems at present.
Table 1. An example showing the structure of rule for identifying Vesicular Stomatitis.

<table>
<thead>
<tr>
<th>IF</th>
<th>THEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever (0.1)</td>
<td>Vesicular Stomatitis(1, 0.8)</td>
</tr>
<tr>
<td>AND blister-like lesions on the tongue, mouth lining, nose and lips (0.8)</td>
<td></td>
</tr>
<tr>
<td>AND lesions also on the coronary bands (0.25)</td>
<td></td>
</tr>
<tr>
<td>AND drooling or frothing at the mouth (0.25)</td>
<td></td>
</tr>
<tr>
<td>AND excessive salivation (0.25)</td>
<td></td>
</tr>
<tr>
<td>AND weight loss (0.1)</td>
<td></td>
</tr>
<tr>
<td>THEN</td>
<td></td>
</tr>
</tbody>
</table>

**Overall design of Horse-Expert**

**Architecture of Horse-Expert**

Horse-Expert was developed according to the structured systems development method (Sutcliffe et al. 1991). It is comprised of a knowledge base, knowledge acquisition subsystem, disease diagnosis subsystem, administration subsystem and a user interface (Fig. 1). The core of system is the knowledge base and the diagnostic inference machine. System uses N-tier web architecture design (Finkelstein et al. 2004).

**Knowledge acquisition and representation**

As an expert system, it must contain the knowledge acquired from multiple experts in domain field. Knowledge acquisition and representation is the crucial step and the bottleneck in the construction of expert systems (Potter et al. 2000). The basic task is to establish a sound, efficient, accurate and comprehensive knowledge base to meet the expert system’s requirements for problem solving (Li et al. 2015, Walsh 2015). Many techniques have been developed for knowledge acquisition from domain experts (Boose 1985, Leung et al. 2006, Marcus 2013, Tudorache et al. 2013). In this study, we analyzed and summarized the body of knowledge by conducting literature reviews and interviewing experts by a questionnaire to analyze diseases (Fig. 2A). It consisted of questions about the consistency and accuracy of knowledge (Armoni 1995). Experts modified the questionnaire based on their experiences and the requirements for developing an expert system. Meanwhile, experts gave each rule an activated threshold weight factor ($\omega$) based on their own experiences. Lower thresholds indicate that the diagnostic rule can be fired even if the user gives a relatively low confidence value, while high thresholds indicate that a rule can only be fired when the user is certain of the value.

Using the above mentioned method, we could gather high quality knowledge from experts. Nevertheless, relying solely on this method has some shortcomings. When experts are working with many lists, it is likely that some aspects of the problem will be ignored and the final results become one-sided. To compensate for this deficiency, a web-based knowledge elicitation subsystem was designed to gather/update the facts, generate rules and modify the weight factors (Fig. 2B). Domain experts were authorized to access this system. Continuous feedback was maintained in the knowledge acquisition process to ensure the field is properly represented. Production rules are used for knowledge representation (Table 1). Each rule is in the form of “IF symptoms pattern A, THEN the disease B”. The structure of a rule for identifying a particular disease makes use of textual symptoms in the form of conditions. These conditions are connected through logical operators (and/or).

**Database**

Database stores domain knowledge required for solving problems including the age, breed of the horse, symptoms, pictures and other pertinent information. The database is designed mostly with MS SQL Server 2005. Open Database Connectivity (ODBC) is used to access the data from the database (Kolhe et al. 2011). Object-oriented programming was used to build the database. The database stores facts and data which are required for knowledge. A horse disease diagnosis system knowledge base is established using a relational database.

Horse-Expert’s knowledge base contains up-to-date information for 91 types of major diseases and 286 symptoms of diseases occurring in the horses (Hunt et al. 2010, Khan et al. 2012, Yamanaka et al. 2012, Wise et al. 2013, Balasuriya 2014). The disease diagnosis and treatment knowledge can be divided into three parts: textual knowledge, multimedia knowledge and prevention knowledge.
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Type</td>
<td>Disease Name</td>
<td>Symptom</td>
<td>Initial Weight</td>
<td>Modified Weight</td>
<td>Experts opinions about the symptoms</td>
<td>Comprehensive opinions</td>
</tr>
<tr>
<td>Digestive system</td>
<td>Stomatitis</td>
<td>fever</td>
<td>Factor (≤1)</td>
<td>Factor (≤1)</td>
<td>description (a: Without modification; b: Need modification; c: Delete)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>0.75</td>
<td>blister-like lesions on the tongue, mouth lining, nose and lips</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lesions also on the coronary bands</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>drooling or frothing at the mouth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>excessive salivation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>weight loss</td>
<td></td>
</tr>
</tbody>
</table>

Please provide your opinions here.

Fig. 2 A, B
knowledge is stored by production rules in the rule base (Kolhe et al. 2011). The dynamic knowledge base generates a union set of questions-answers and automatically calculates the total confidence estimate for each question (Kolhe et al. 2011).

Inference process

In the real world, uncertainty plays an important role in the continuous change of knowledge over time (Zadeh 1983). In this section, we presented
a method for diagnosing horse diseases based on an evidence-weighted uncertainty reasoning model, which combined evidence theory and uncertainty pass algorithm of confidence factor (CF) to overcome the complexity and uncertainty of the characteristic conditions of horse diseases. We used a forward chaining method to guide the process of rule and outcome selection (Mansingh et al. 2007, Xu et al. 2010). System and inference strategy do not make assumptions, and the reasoning is entirely dependent on the evidence provided by the users instead of the confidence factor by experts. Knowledge is expressed in production rules as follows.

\[
\text{IF } E_1 (\omega_1) \text{ and } E_2 (\omega_2) \text{ and } ... \text{ and } E_n (\omega_n) \text{ then } H (\text{CF}(H, E), \lambda) \text{ CF } \in (0,1)
\]

where:
\( E_1, E_2, ..., E_n \) are the preconditions for knowledge, called symptoms here. It could be a simple symptom, and could also be the combination of multiple symptoms connected with AND & (\&) OR. \( H \) is the diagnosis conclusion and also refers to the disease. \( \omega_i \) (\( i = 1, 2, 3 ..., n \)) is the weight factor of the symptom, whose initial value is assigned by the domain experts. \( \lambda \) is the threshold value, which provides a limit to the availability of the corresponding rules. The value of \( \text{CF}(H, E) \) and \( \lambda \) should be set by experts of horse disease domain. \( \text{CF}(E_i) \) represents the confidence factor for the evidence. Its initial value is assigned by the users.

The uncertainty pass algorithm. For the precondition,

\[
E = E_1 (\omega_1) \text{ AND } E_2 (\omega_2) \text{ ANG } ... \text{ AND } E_n (\omega_n)
\]

the algorithm (assuming that \( \sum \omega_i = 1 \)) for the confidence factor of the corresponding combination evidence is as follows:

\[
\text{CF}(E) = \sum_{i=1}^{n} \omega_i \times \text{CF}(E_i) \tag{1}
\]

If \( \omega_i \) (\( i = 1, 2, 3..., n \)) does not meet the normalization condition,

\[
\sum_{i=1}^{n} \omega_i \neq 1
\]

Calculation of \( \text{CF}(E) \) is made as shown below:

\[
\text{CF}(E) = \frac{1}{n} \times \sum_{i=1}^{n} (\omega_i \times \text{CF}(E_i)) \tag{2}
\]

When \( \text{CF}(E) \) of the conclusion of a rule exceeds the threshold (\( \lambda \)),

\[
\text{CF}(E) \geq \lambda
\]

The rule can be used for inference of the corresponding conclusion confidence. The confidence factor of the conclusion of \( H \) can be calculated as follows:

\[
\text{CF}(H) = \text{CF}(H,E) \times \text{CF}(E) \tag{3}
\]

**User uncertainty**

Uncertain evidence reasoning fully considers the uncertainty and unpredictability of the objective world, as well as the inaccuracies and incomplete understanding of the world (Palacharla et al. 1994). It also reduces the influence of subjective factors on judgment and the conscious reasoning process.

Horse-Expert’s aim is to simulate the process of problem solving in the field. The process of diagnosing disease by domain experts includes the quantitative conversion of their individual clinical experiences, and some reasoning and judgment about those experiences. Many times experts are not sure of the exact relationship between symptoms and disease (Tocatlidou et al. 2002). In Horse-Expert, knowledge is actually a set of questions along with multiple answers, each question having a confidence factor (\( \text{CF}(E_i) \)) estimate that includes definite (100% certain), almost certain (75-100% certain), probably (50-75% certain), slight evidence (25-50% certain), and ignored (0-25% certain), which is internally calculated into a numerical value (1, 0.75, 0.5, 0.25, 0.05) by the system itself (Devraj et al. 2011). Users assign these confidence factors when users feed their information to the system. System then combines confidence factors with weight factors in the diagnostic process, to reduce the influence of uncertainty on the diagnosis.

**System interface**

Horse-Expert provides three kinds of intelligent diagnostic modes: symptoms-based diagnosis, diseases-based diagnosis and cases-based diagnosis. Users select the appropriate stall and horse before diagnosis so that the system can save separate medical records for each horse. An interface for horse inspecting and diagnosing results are illustrated in Figure 2C. Combined the type of users’ experience three diagnostic mode have different strategies.
Table 2. Results of successful diagnosis of system testing.

<table>
<thead>
<tr>
<th>Case type</th>
<th>Number of cases</th>
<th>Confirmed cases</th>
<th>Coincidence rate (%)</th>
<th>Miss rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infectious disease</td>
<td>154</td>
<td>141</td>
<td>91.56</td>
<td>8.44</td>
</tr>
<tr>
<td>Parasitic disease</td>
<td>126</td>
<td>120</td>
<td>95.23</td>
<td>4.76</td>
</tr>
<tr>
<td>Digestive system</td>
<td>185</td>
<td>174</td>
<td>94.05</td>
<td>5.95</td>
</tr>
<tr>
<td>Respiratory system</td>
<td>104</td>
<td>95</td>
<td>91.35</td>
<td>8.65</td>
</tr>
<tr>
<td>Poisoning disease</td>
<td>24</td>
<td>18</td>
<td>75.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Nutritional metabolic disease</td>
<td>25</td>
<td>20</td>
<td>80.00</td>
<td>20.00</td>
</tr>
<tr>
<td>Urinary system</td>
<td>17</td>
<td>13</td>
<td>76.47</td>
<td>23.53</td>
</tr>
<tr>
<td>Nervous system</td>
<td>18</td>
<td>12</td>
<td>66.67</td>
<td>33.33</td>
</tr>
<tr>
<td>Surgical disease</td>
<td>72</td>
<td>64</td>
<td>88.89</td>
<td>11.11</td>
</tr>
<tr>
<td>Eye disease</td>
<td>54</td>
<td>45</td>
<td>83.33</td>
<td>16.67</td>
</tr>
<tr>
<td>Genital system</td>
<td>25</td>
<td>19</td>
<td>76.00</td>
<td>24.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>804</strong></td>
<td><strong>721</strong></td>
<td><strong>89.68</strong></td>
<td><strong>10.32</strong></td>
</tr>
</tbody>
</table>

Notes: Confirmed cases are the number of cases had been correctly diagnosed by the system. Coincidence rate (%) = confirmed cases / total cases. Miss rate (%) = (total cases – confirmed cases) / total cases.

– Symptoms-based diagnosis is suitable for users who have little experience for diagnosing or with moderate experience but cannot fully diagnose. Users select the observed symptoms and assign a confidence factor, and system makes a diagnosis. Meanwhile, users can confirm their choices based on the pictorial disease symptoms.

– Diseases-based diagnosis is suitable for users who have extensive experiences. These users make an educated guess at the diseases and system searches for all of the symptoms associated with the diseases to verify the correctness of the suspected diagnosis.

– Case-based diagnosis is based mainly on the results of the other two diagnostic modes, and can provide the results of previous diagnoses for users. Users can access the last onset of symptoms, diagnostic results, etc.

In addition, a case-statistical function was designed for users to more intuitively grasp the occurrence of horse disease. The function is based on the system’s collection of a number of cases, uses multiform statistical methods, and includes the breed, sex, diagnosis results, stall and onset time of disease (Fig. 2D).

**System evaluation**

The evaluation process was carried out in user interface evaluation and system testing. The goal of the evaluation was to guarantee that there are no dead-end lines of reasoning, and that all possible errors and bugs were discovered (Kolhe et al. 2011). The user interface evaluation involved verifying that the system was built correctly, and ensuring the knowledge in the system is consistent, complete and correct to satisfy the users’ demands. The system program was run multiple times using different combinations of all possible inputs. In accordance with the requirements of the project, we conducted a demonstration in seven demonstration areas of Xinjiang province in China.

804 sample disease cases were selected to evaluate the coincidence rate (%) of system. These test cases include: (i) 150 cases accumulated from 1984-2008 at the animal hospital of Northeast Agriculture University, (ii) 351 cases collected from domestic and foreign literature related to horse disease, such as the Equine Veterinary Journal, Journal of Equine Veterinary science, etc, (iii) 303 cases provided by cooperation unit from Xinjiang province.

All of the horse disease cases were divided into 11 categories according to their characteristics. Successful and unsuccessful diagnoses were noted for each case. Successful diagnosis is defined as the symptoms users input that result in proper diagnosis, and unsuccessful diagnosis is defined as the symptoms that resulted in either an incorrect or inconclusive diagnosis. The results were verified by domain experts at Northeast Agriculture University, Harbin, China (Table 2). It shows that Horse-Expert made successful diagnoses in 721 cases and had a coincidence rate of 89.7%.
Conclusions and future work

An expert system for diagnosing horse disease (Horse-Expert) which is based on evidence weighted uncertainty reasoning approach has been developed to provide online support to veterinarians with minimal experience in China. The development of Horse-Expert addresses the present situation of China's horse husbandry by aiding the ability of veterinarians to diagnose diseases. Symptoms of disease are quite variable because horse diseases are often caused by a variety of factors such as bacteria, or viruses. To further complicate matters, new differences appear frequently. The language which veterinary experts use to discuss clinical cases also varies from person to person. So far there is no way to fully express the veterinary experts' clinical experiences. That is to say, despite an 89.7% diagnosis coincidence rate, Horse-Expert still does not completely replace clinical veterinarians. It can only play an important advisory role and serve as a reference function to veterinarians. It is clear that we cannot rely on system for optimal diagnostic results alone, hence bacteriological and virological examination is the necessary means of diagnosis for some diseases. It is reasonable to utilize the Horse-Expert as an aid. This positioning is not only conforming to the actual farming, but also greatly simplifying the system knowledge base data capacity and over-fine reasoning process.

The representation and processing of uncertainty is one of the main tasks of the expert system development process. In terms of diagnosing animal disease, evidence uncertainty is widespread. This uncertainty is mainly manifested in the objective uncertainty of symptoms, and the subjective uncertainty of people in analyzing and describing the symptoms. However, users often experience difficulty determining the clinical manifestations in the animal. Treatment, vaccination or other environmental factors lead to atypical symptoms of clinical manifestations of disease. Therefore, even if the user is in a state of hesitation, there are only two options, affirmative or negative. This results in error of reasoning and increases the influence of subjective factors on the reasoning conclusion. Therefore, evidence-weighted uncertainty reasoning fully considers the objective complexity of the world, as well as the imprecision and incompleteness of human understanding of objective reality. At the same time, it reduces the influence of subjective factors on the judgment and reasoning process. In this theory, the weight factor is introduced to carry out inference under the condition that evidence is uncertain and incomplete. When assigning the weight factor, we need to pay attention to the question that symptoms should be assigned a higher weight factor for established diagnoses. If an independent symptom is depended upon by other symptoms, it should have a greater weight factor. Therefore, it is important to consider the importance and independence of symptoms. Generally, the range of weight factors is from 0 to 1 for all the symptoms of a disease, and the sum of their weight factors is to be 1.

Horse-Expert's knowledge base covers 604 rules and 400 images and videos for different types of symptoms and diseases. The system can diagnose 91 kinds of common horse diseases. It provides an aid to diagnosis and a learning aid for veterinary staff and students. The system can also address the issue of a scarcity of horse veterinarians in China. Its inference machine uses forward chaining methods and three kinds of user interface to meet the needs of users of different levels, who have different diagnostic capabilities. Users assign confidence factors when input their response to the system, after which the system uses a weight factor initially assigned by the domain experts to draw a conclusion. The "User Feedback" fully considers the uncertainty associated with the object uncertainty of symptoms and the subject uncertainty in the analysis and description of symptoms and diseases. If the system cannot make a diagnosis, it will report high weight factor symptoms according to the feedback from users. Thus, users can make a diagnosis after further observation of symptoms. The responses from evaluators (farmers from Xinjiang province, veterinarians, veterinary students, domain experts) showed that the system is user-friendly and convenient to operate. It can simulate the thoughts and processes of diagnosing and treating done by domain experts and also can disseminate expert knowledge to non-experts. The system test showed that the diagnostic accuracy of the aided system was 89.7%.

The expert system developed for the horse is not yet complete. It is necessary to further study the knowledge representation and inference mechanism according to the characteristics of horse diseases in the respective areas worldwide. We also plan to develop the current system in Kazakh, Mongolian, Uyghur and other languages, so that all horse keepers in China can access and use the program. In addition, some new diseases have arisen with the development of the horse husbandry, so it is necessary to further expand the knowledge base.

Acknowledgments

The research was supported by the National Science and Technology Support Program of China (Project No.: 2012BAD46B04). We thank all domain
experts from Northeast Agriculture University (NEAU), South China Agriculture University (SCAU), Huazhong Agriculture University (HZAU), Yang Zhou University (YZU), Shandong Agriculture University (SDAU) and Xinjiang Agriculture University (XJAU) for their supports.

Reference