Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC), Washington, DC, November 5–9, 1994, 730–734

Improving Prediction of Preterm Birth Using a New Classification Scheme and Rule Induction

Jerzy W. Grzymala-Busse
Department of Electrical Engineering and Computer Science
University of Kansas
Lawrence, KS 66045
E-mail: jerzy@cs.ukans.edu

Linda K. Woolery
School of Nursing
University of Missouri
Columbia, MO 65203
E-mail: nurswool@mizzou1.missouri.edu

Abstract. *Prediction of preterm birth is a poorly understood domain. The existing manual methods of assessment of preterm birth are 17% – 38% accurate. The machine learning system LERS was used for three different datasets about pregnant women. Rules induced by LERS were used in conjunction with a classification scheme of LERS, based on "bucket brigade algorithm" of genetic algorithms and enhanced by partial matching. The resulting prediction of preterm birth in new, unseen cases is much more accurate (68% – 90%).*

1. INTRODUCTION

Many healthcare providers collect data on pregnant women for assessment of preterm birth risk. Current technology makes possible collection of a plethora of data, yet a perinatal healthcare provider has no access to a general, reliable and valid method of preterm birth assessment [1]. Preterm birth is defined ambiguously in the literature on the subject. In our work we will assume that preterm delivery is before the 36th week of gestation and fullterm starts from the 36th week.

In the United States the rate of preterm birth has been between 8 and 12% for the last two decades [9]. Preterm birth is the most common cause of low birthweight and perinatal mortality and it causes almost 70% of all neonatal deaths [16]. At the same time, preterm infants are 40 times more likely to die than infants born at term. Moreover, surviving preterm infants are at increased risk of lifelong handicaps including cerebral palsy, respiratory diseases, blindness and deafness [9]. Accurate assessment of preterm birth permits intervention with appropriate educational programs, bedrest, and early symptom identification. Tocolyting drugs may be used to suppress preterm labor.

Most existing methods to assess preterm birth are based on risk scoring, done manually. These methods are between 17% and 38% predictive in determining preterm birth [9]. This range of accuracy is obviously not satisfactory. Some authors conclude that—in general—manual risk screening tools are not sufficient to be used in the prediction of preterm labor [3, 8]. Our research shows that performance of computer-based methods for prediction of preterm birth is significantly better than performance of manual methods.

2. MACHINE LEARNING

The exact causes for initiation of uterine contractions leading to delivery are mostly unknown. Since this domain is poorly understood, our hypothesis was that AI methods, which usually work well under these circumstances, should be applied. The task is to identify regularities hidden in large datasets characterized by many attributes containing information about pregnant women. In this work the chosen methodology was of machine learning from examples. Classification rules for prediction of preterm birth were induced using machine learning program LERS (Learning from Examples based on Rough Sets), developed at the University of Kansas. System LERS may work with imperfect data, e.g., with missing attribute values, continuous attributes, or inconsistencies in input data. The system handles inconsistencies using rough set theory [4–6, 12, 13, 18]. Other machine learning systems that are use rough set theory include Datalogic/R [19], and RoughDAS and RoughClass [17].

A machine learning system looks for regularities in a data set. In most of these systems, such regularities are expressed as rules in the following format:

if (attribute_1, value_1) **and** ···
and (attribute_n, value_n) **then** (decision, value).

Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC), Washington, DC, November 5–9, 1994, 730–734

Table 1. Decision Table—Training Examples

Example	Attributes			Decision
	Risk_factor	Infant_sex	Age	Delivery
1	smoking	male	31–45	preterm
2	smoking	female	31–45	fullterm
3	none	female	<20	fullterm
4	none	male	20–30	fullterm
5	smoking	male	20–30	preterm
6	none	female	31–45	fullterm
7	none	male	<20	preterm

Input data for system LERS may be presented in the form of a decision table, where *examples* (e.g., patients) are characterized by *attributes* (e.g., *Risk_factor*, *Infant_sex*, *Age*, etc.) and a *decision* (e.g., *Delivery*). An example of a very simple decision table is presented in Table 1. This table is presented here only for illustration. The actual decision tables will have many more variables and examples. The current version of system LERS can handle more than one hundred attributes and many thousands of examples. The decision table from Table 1 has six examples, named 1, 2,..., 6. In machine learning from examples, a *concept* is understood as the set of all examples having the same value for a decision. Patients 1, 2, and 3 all belong to the same concept of *Delivery* having the value *preterm.*

In this work algorithm LEM2 of LERS has been used [6]. This algorithm induces rules in their simplest form. The first criterion in looking for attribute-value pairs, candidates for the left-hand side of a rule, is relevancy of the attribute-value pair to the concept described by the right-hand side of the rule.

Rules, induced by LERS, are more general than information contained in the original decision table

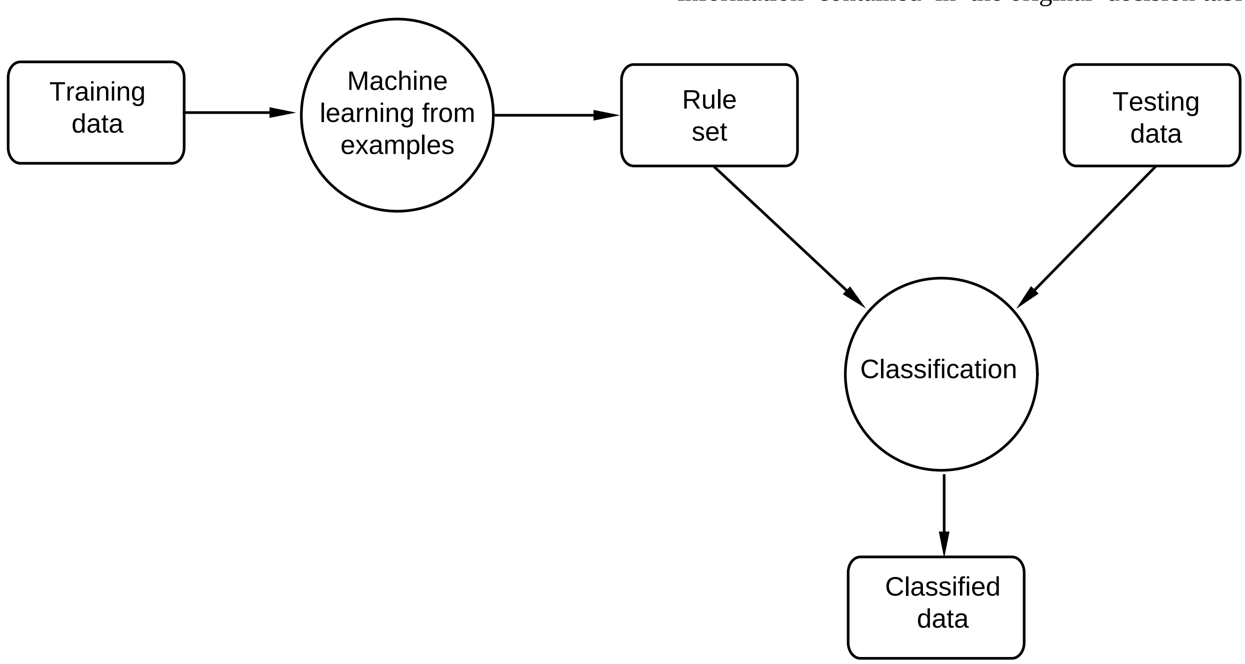

Fig.1 Machine learning from examples and classification

Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC), Washington, DC, November 5–9, 1994, 730–734

Table 2. Decision Table—Unseen Examples

Examples	Attributes			Decision
	Risk_factor	Infant_sex	Age	Delivery
8	smoking	female	20–30	preterm
9	none	female	20–30	fullterm
10	none	male	31–45	preterm

representing input data, since—in general—more new examples may be correctly classified by rules than may be matched with examples from the original decision table.

All rules, induced by LEM2 from Table 1, are:

(Infant_sex, male) & (Risk_factor, smoking) ->
(Delivery, preterm),

(Age, <20) & (Infant_sex, male) ->
(Delivery, preterm),

(Infant_sex, female) -> (Delivery, fullterm),

(Age, 20-30) & (Risk_factor, none) ->
(Delivery, fullterm).

3. CLASSIFICATION

In practice, rules induced from training examples are used to classify new examples never before seen by a machine learning system. Obviously, during classification of these unseen examples, some of them will not be classified properly because matched rules will indicate a wrong concept and some of them will not be classified at all because no one single rule will match the example.

The naive approach to classification of unseen examples, where an attempt is made to classify an example using all possible rules, usually produces poor results. Some more sophisticated mechanisms of classification have been developed. One of them is a *decision list* [15], used in machine learning systems CN2 [11] and C4.5 [14], where rules are ordered and classification of a new example starts from the first rule. The process is terminated when the first matching rule is identified. The last rule is a default rule.

Another mechanism of classification is used in AQ15 [10]. In AQ15, first the *complete matching* is examined, in which all attribute-value pairs of a rule must match all values of the corresponding attributes for the example. If complete matching is impossible, a partial matching is done, where some attribute-value pairs of a rule match the values of corresponding attributes. The choice of the best rule is made on the basis of estimates of probabilities.

In LERS yet another approach was used, similar to the "bucket brigade algorithm" of *genetic algorithms* [2, 7]. Every rule is equipped with a number of correctly classified examples during training, called *strength*. In our example, this number is equal to two for the first rule:

(Infant_sex, male) & (Risk_factor, smoking) ->
(Delivery, preterm),

because this rule correctly classifies two examples: 1 and 5. Similarly, strengths of the second and the fourth rules are both equal to one, and the strength of the third rule is equal to three.

For every example LERS first attempts complete matching. The example is classified as belonging to concept C with the largest value of support, defined by the following formula:

$$\sum_{\text{matching rules } R \text{ describing } C} \text{Strength}(R) * \text{Specificity}(R),$$

where *Specificity* (R) is the total number of attribute-value pairs of the rule R.

In the bucket brigade algorithm partial matching is not considered to be worth the trouble. On the other hand, LERS uses partial matching because it is a successful addition to complete matching. When complete matching is impossible partially matching rules are considered, with some attribute-value pairs of a rule matching the values of corresponding attributes for the example. For a partially matched rule R, the additional factor is computed, called *Matching_factor* (R), the ratio of the number of

Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC), Washington, DC, November 5–9, 1994, 730–734

Table 3. Experimental Results

	Dataset 1	Dataset 2	Dataset 3
Number of Training Examples	1654	1218	6608
Number of Unseen Examples	1593	1218	6608
Number of Attributes	13	73	67
Number of Rules	178	170	1133
Prediction Rate: Naive Classification Scheme	72.69%	44.66%	35.43%
Prediction Rate: New Classification Scheme	89.96%	72.00%	67.99%

matched attribute-value pairs of the rule *R* to the total number of attribute-value pairs of the rule *R*. In partial matching, the example is classified as belonging to the concept *C* with the largest value of support defined by

$$\sum_{\text{partially matching rule } R \text{ describing } C} \text{Matching\_factor}(R) * \text{Strength}(R) * \text{Specificity}(R).$$

The classification mechanism of LERS will be illustrated using examples from Table 2.

Example 8 is completely matched by the rule

(Infant_sex, female) -> (Delivery, fullterm),

with the strength equal to 3. This is the only rule that completely matches the example, hence the decision is (Delivery, fullterm). That is inconsistent with the value of *Delivery* for example 8 from Table 2, so this example is incorrectly classified. Example 9 is completely matched by two rules

(Infant_sex, female) -> (Delivery, fullterm),

(Age, 20-30) & (Risk_factor, none) -> (Delivery, fullterm),

with strength equal to 3 and 1, respectively. Both rules support the same value *fullterm* of the decision. In this case the value *fullterm* is consistent with the value of decision for example 9 from Table 2, so this example is correctly classified. Example 10 cannot be matched completely by any rule. However, it is partially matched by the following three rules

(Infant_sex, male) & (Risk_factor, smoking) -> (Delivery, preterm),

(Age, <20) & (Infant_sex, male) -> (Delivery, preterm),

(Age, 20-30) & (Risk_factor, none) -> (Delivery, fullterm).

All three rules have the same *Matching_factor* equal to 0.5. Pair (*Delivery*, *preterm*) has support $0.5 * 2 * 2 + 0.5 * 1 * 2 = 3$, while pair (*Delivery*, *fullterm*) has support $0.5 * 1 * 2 = 1$, so the final decision is (*Delivery*, *preterm*), consistent with the value of *Delivery* for example 10 from Table 2. Thus, this example is correctly classified.

4. EXPERIMENTAL RESULTS

Experiments were done on three large datasets. Each of the datasets was split in half; half of the data were used for machine learning using system LERS, while the other half of the data were used for validation of the rule sets using two approaches: a naive approach for complete matching of every example with all possible rules and the classification scheme of LERS. As follows from Table 3, the prediction rate (or accuracy) for the classification scheme of LERS for all three datasets was much higher than for manual methods.

Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC), Washington, DC, November 5–9, 1994, 730–734

5. CONCLUSIONS

The experimental results show that the prediction rate of rule sets with appropriate classification scheme is much higher (68% – 90%) than the traditional manual methods (17% – 38%). All of these rule sets were induced from raw data. Only the examples with the most obvious errors such as maternal weight equal to 10 or 700 pounds or systolic pressure of 14,000 were removed. There was an attempt to validate rules through inspection by experts in the area; however, diagnosticians were not prepared to interpret rules, in spite of the fact that the rules were written almost in plain English. Further research is necessary to include additional attributes, such as stress, sexual activity, or nutritional status that were not taken into account.

Acknowledgment. The authors are grateful for datasets provided by: St. Luke's Regional Perinatal Center (Kansas City, MO), Healthdyne Perinatal Services (Marietta, GA), and Tokos Corporation (Santa Anna, CA).

References

[1] G. Alexander, J. Weiss, T. Hulsey, E. Papiernik. Preterm birth prevention: an evaluation of programs in the United States. *Birth* 18, 1991, 160–169.

[2] L. B. Booker, D. E. Goldberg, J. F. Holland. Classifier systems and genetic algorithms. In *Machine Learning. Paradigms and Methods*. J. G. Carbonell (ed.), The MIT Press, 1990, 235–282.

[3] R. Creasy. Preventing preterm birth. *New England J. of Medicine* 325, 1991, 727.

[4] J. W. Grzymala-Busse. Knowledge acquisition under uncertainty—a rough set approach. *Journal of Intelligent & Robotic Systems* 1, 1988, 3–16.

[5] J. W. Grzymala-Busse. *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, 1991.

[6] J. W. Grzymala-Busse. LERS—A system for learning from examples based on rough sets. In *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory.* R. Slowinski (ed.), Kluwer Academic Publishers, 1992, 3–18.

[7] J. H. Holland, K. J. Holyoak, R. E. Nisbett. *Induction. Processes of Inference, Learning, and Discovery*. The MIT Press, 1986.

[8] J. McGregor, J. French, R. Richter, A. Franco-Buff, A. Johnson, S. Hillier, F. Judson, J. Todd. Antenatal microbiologic and maternal risk factors associated with prematurity. *Am. J. Obstet. Gynecol*. 163, 1990, 1465–1473.

[9] M. McLean, W. Walters, R. Smith. Prediction and early diagnosis of preterm labor: a critical review. *Obstet. Gynecol. Surv*. 48, 1993, 209-225.

[10] R. S. Michalski, I. Mozetic, J. Hong, N. Lavrac. The multi-purpose incremental learning system AQ 15 and its testing application to three medical domains. *Proc. of the 5th Nat. Conf. on AI*, 1986, 1041–1045.

[11] T. Niblett, I. Bratko. Learning decision rules in noisy domains. *Proc. of the Expert Systems '86, the 6th Annual Technical Conference of the British Computer Society*, 1986, 25–34.

[12] Z. Pawlak. Rough sets. *Int. J. Computer and Information Sci.* 11, 1982, 341–356.

[13] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, 1991.

[14] R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[15] R. L. Rivest. Learning decision lists. *Machine Learning* 2, 1987, 229–246.

[16] M. Rosen, I. Merkatz, J. Hill. Caring for our future: a report of the expert panel on the content of prenatal care. *Obstet. Gynecol*. 77, 1991, 782–787.

[17] R. Slowinski, J. Stefanowski. 'RoughDAS' and ''RoughClass' software implementations of the rough set approach. In *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. R. Slowinski (ed.), Kluwer Academic Publishers, 1992, 445–456.

[18] L. Woolery, J. W. Grzymala-Busse, S. Summer, A. Budihardjo. On the use of LERS-LB knowledge acquisition for expert system development in nursing. *Comput. Nurs*. 9, 1991, 227–234.

[19] W. P. Ziarko. Acquisition of control algorithms from operation data. In *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. R. Slowinski (ed.), Kluwer Academic Publishers, 1992, 61–75.