IUI Mining: A Variation of Information Theoretic Network Approach

Suwimon Kooptiwoot, Muhammad Abdus Salam

School of Information Technologies, University of Sydney
The University of Sydney, NSW, 2006, Australia
{suwimon, mslam}@it.usyd.edu.au
TEL: (61 2) 9351 3523 , (61 2) 9351 4276 , FAX: (61 2) 9351 3838

Abstract: The IUI (Intra uterine insemination) is an assisted reproductive technique, enabling infertile couples to achieve successful in pregnancy. We propose the variation of information theoretic network approach to mine knowledge to suggest possible modifications to IUI treatment plan in order to improve overall success rates. The information theoretic network algorithm employs the statistic significant to construct the network. We propose a new algorithm by adding up the medical significant criteria to construct the information theoretic network. We found that this new algorithm give us the reasonable result. And this shows that the medical knowledge cannot be ignored in data mining in medical domain.

1 Problem

Fertility treatment is still obscured. Success rate is still less. As we know that Data mining is the method in the discovery process which can help us to uncover the knowledge [1-4]. Data Mining is widely used in many applications. So we think that we should be able to use the data mining technique to reveal the knowledge in the fertility treatment to help infertile couple get success in pregnancy.

As seen in [5] there is an effort to uncover the knowledge about the fertility treatment. Dr. Susan Adams, a world leader in the field of uterine receptivity and Professor Chris Murphy, Bosch Professor of Histology and Embryology, who heads the Cell and Reproductive Biology Laboratory, the University of Sydney's Institute for Biomedical Research, have been working on IVF (In Vitro Fertilization) technique to combat infertility for some time. The key to success is based on customized hormone treatment and the receptivity of the uterus of each patient. Dr Adams and Professor Murphy were approached by a number of couples who were totally desperate, had tried IVF unsuccessfully and had been told to forget it. They have had a more than 50 percent success rate with these patients.

While personalized medicine is the key of success of the treatment, it is also dependent on Dr. Adams's expertise with the scanning electron microscopes of the Australian Key

Centre for Microscopy and Microanalysis, based at Sydney University. Pregnancy is the successful union of healthy eggs and sperm and also the receptivity of the uterine wall. Dr. Adams analysed uterine samples of the patients individually in minute detail. She said that hormonal priming of the uterus is critical and the standard IVF treatment gives the standard dose, but it does not produce an optimal result in every woman. They took a biopsy and looked at it and decided whether it was sufficiently stimulated with oestrogen or progesterone, and checked that it was in phase with the cycle. Then they suggested a treatment and tested again to see whether it had been effective, before suggesting that embryo transfer should proceed. Professor Murphy said that this is the personalized medicine in the most direct sense.

2 IUI Data

We got the data collected in the IUI treatment. We tried to use the data mining method to find the knowledge to improve the success rate in IUI treatment. IUI (Intra Uterine Insemination) is an assisted reproductive technique [6, 7]. This data is AIH data. AIH (Artificial Insemination with the husband's semen) is used in the management of infertility due to impotence or anatomical abnormality in the male, especially hypospadias which prevents the normal ejaculation of sperm into the upper vagina [7]. The semen which obtained by masturbation is injected onto the surface of the cervix at the ovulation time. Injecting the semen directly into the uterus through the cervix has been advocated in cases postcoital tests have shown failure of sperm penetration. The success rate of AIH is not greater than that normal intercourse when this can take place. The IUI treatment is using both AIH and hormonal treatment.

This data set consists of 104 attributes, 1597 records. The attributes are about the history of pregnancy of the couple, the present situation of the couple, the normal situation of the couple, the results from laboratory examinations, and the treatment details, for example, the type of hormone used in the treatment, the dose of the hormone used, and the pregnancy result from the treatment.

3 IFN Algorithm

First we tried to find the data mining algorithm to reveal the relationships among the data attributes which related to success in pregnancy. We see that the association rules algorithms give us the very huge number of rules as seen in [8, 9]. If we use the association rules algorithms, we have to face the difficulty in finding the useful relationships which is the main problem of using the association rules algorithms as seen in [10-12]. Then we looked at the classification algorithms which give us the less number of the rules. We see that classification algorithms still have many flaws from avoiding over-fit model as seen in [13-18]. We decided to use IFN (Information Theoretic Network) algorithm which is the part of Info-fuzzy Network methodology as proposed in [19-22]. The advantages of this algorithm are automatically reducing the dimension of the data, the less number of the rules given, and the ability to tell us the appropriate rules. The details of this original algorithm is in [21]

4 Data Mining Process

We followed the data mining process as proposed in [23] which consists of

1. Understanding the problem domain
2. Obtaining and understanding the data
3. Preparation of the data
4. Construction of the knowledge model from the data
5. Evaluation of the model
6. Using the model (Interpretation and Post processing)

We are interested in finding the factors in the intrauterine insemination (IUI) treatment which effect on getting success in pregnancy. Then we studied about the intrauterine insemination treatment and the success in pregnancy. We have got the IUI data. Then we prepared the data, cleaned the data and checked the reliable of the data. If we find that the data is absolutely wrong, then we treat it as missing value. As everyone knows that the data quality consists of data reliability, accuracy, relevance, completeness, consistency, precision, etc, so we have to clean the data and check the reliable of the data as much as we can before processing. For the missing values, we delete the attributes which have large number of missing values out of our consideration because they don't have information for us to process, for example, one attribute has only two records which have the data, the rest of all of other records are missing values, so we discard this attribute out of our consideration.

We don't estimate the value for any missing value because our data are very sensitive and the missing values can be anything beyond any method to be able to make guessing to interpolate them. So if there are missing values, we still process them as missing values. We don't use any interpolation method to guess their values. Then we transform the data to make them be able to process.

We discrete continuous data by using equal width and equal frequency techniques.

Then we construct the knowledge model from the data by using the information theoretic network method. This method mainly employs the idea of the information sharing. The root node which is the first node of the network is the starting point and this is the first layer of the network. For the second layer or the first hidden layer of the network is the first attribute selected by its lowest conditional entropy with the target. And the number of nodes in this layer is the number of the distinct discrete values of the attribute in this layer.

Then we decide which node in this layer should be split by using the likelihood ratio. The node will be split if its likelihood ratio is higher than the significant level we set. In this experiment, we set the likelihood ratio is zero. If we find that there is (are) node(s) to be split, then we select the next attribute to be the next hidden layer of the network. And we build up the network by connecting all of the distinct discrete values of the selected attribute in Cartesian product with all of the split nodes. The next attribute is selected by the highest increasing the conditional mutual information with the target attribute. And we keep constructing the network on the same way until no node should be split or no attribute should be selected to be the next layer.

After finishing the network construction, we get the network model. Each connection of unsplit node to the target values represents the relation between unsplit node (the attributes' value in a branch) and the target value in the form of the rule *IF, THEN*. Then we have to interpret the result from the model constructed. These relations are not so classification rules. They are not so association rules. They tell us the information content by using the weight of the rule. The weight of the rule can be interpreted into three types of the relations which are 1. *IF node, THEN target value*, 2. *IF node, THEN not target value*, and 3. *The node is independent of the target value*. If the weight of the rule is positive, then the relation is *IF node THEN the target value*. If the weight of the rule is zero, then the relation is *the node is independent of the target value*. If the weight of the rule is negative, then the relation is *IF node THEN not the target value*.

In the part of the evaluation of the model, the model from the information theoretic network is in the form of the network. It can be evaluated by looking at the dimensionality reduction which is measured by the proportion of the total selected attributes with the total number of the candidate input attributes or with the total number of the unselected candidate attributes. And we can also evaluate the model by looking at the network size, the amount of the reduced uncertainty of the target attribute. We don't evaluate the model because this algorithm is already proved that its advantages are reducing the dimensions in the data set automatically and the deeper layer construction, the more reduction of the uncertainty of the target attribute. Our main work is using the model (interpretation and post-processing). We worked with the IUI data set because we want to get the knowledge about the factors in the IUI treatment which effect on the success in getting pregnancy of the couple.

5 Results from the Original Algorithm

The network model constructed by the original algorithm is shown in Figure 1

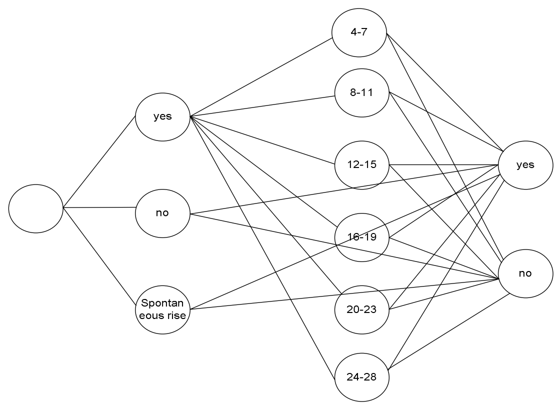

Figure 1. The information theoretic network model 1 from the original approach

The results from using the original information theoretic algorithm are

If we don't give HCG or we find that HCG is spontaneous rise, then we can expect that the couple will not get pregnant.

If we give gonadotrophin before LH peak 4-7 days or 12-15 days or 24-28 days, then we can expect that the couple will not get pregnant.

If we give HCG, and give gonadotrophin before LH peak 8-11 days or 16-23 days, then we can expect that the couple will get pregnant.

From these results we know that *if we want to treat the couple to get success in pregnancy, we should give HCG and give gonadotrophin before LH peak 8-11 days or 16-23 days.*

6 The Variation Approach

We showed the results from the original algorithm to the expert. The expert's comment is the results shown are not quite reasonable because they lack of the other factors which should effect on the achievement in pregnancy. Then we decided to employ the expert knowledge in the network construction. We selected the first attribute to be the first hidden layer of the network by using the medical significance instead of using the statistical significance. Then we kept going with the original algorithm from the second hidden layer of the network until finish.

Algorithm Procedure

Step 1: define the minimum significant level of likelihood ratio for splitting a network node (default = 0.1)

Step 2: Repeat every target attribute i

Step 2.1 Calculate unconditional probability (apriori) of each value of the target attribute by

$$P(V_{ij}) = O_{ij} / n$$

where

O_{ij} – number of occurrences of the value j of a target attribute i in the relation

n – number of complete tuples in the relation

Step 2.2 Calculate unconditional entropy of the target attribute by

$$H(A_i) = - \sum P(V_{ij}) * \log P(V_{ij})$$

Step 2.3 Initialize the information-theoretic network (one hidden layer including the root node associated with all tuples, no input attributes, one target layer for the values of the target attribute).

Step 2.4 Select the first attribute to be the first hidden layer in the network by using clinical significance.

Step 2.5 Repeat for the maximum number of hidden layers (default = number of candidate input attributes).

Step 2.5.1 Repeat for every candidate input attribute i' which is still not an input attribute

Step 2.5.1.1 Initialize zero to conditional mutual information, and the likelihood ratio statistic of the candidate input attribute and the target attribute given the final hidden layer of nodes.

Step 2.5.1.2 Repeat for every node of the final hidden layer

Step 2.5.1.2.1 Calculate the conditional mutual information of the candidate input attribute and the target attribute, given the node, by

$$MI\,(A_{i'} ; A_i \mid z) = \sum_{j=0}^{Mi-1} \sum_{j'=0}^{Mi'-1} P(V_{ij} ; V_{i'j'} ; z) * \log\left(P(V^{ij}{}_{i'j'} \mid z) / (P(V_{i'j'} \mid z) * \mathrm{P}(V_{ij} \mid z)) \right)$$

where

$P(V_{ij}; V_{i'j'} ; z)$ – joint probability of a value j of the target attribute i , a value j' of the candidate input attribute i' and the node z

$P(V^{ij}{}_{i'j'} \mid z)$ – conditional probability of a value j' of the candidate input attribute i' and a value j of the target attribute i given node z
$P(V_{i'j'} \mid z)$ – conditional probability of a value j' of the candidate input attribute i' given node z
$P(V_{ij} \mid z)$ - conditional probability of a value j of the target attribute i given node z

Step 2.5.1.2.2 – Calculate the likelihood ratio statistic of the candidate input attribute and the target attribute, given the node z by

$$G^2 (A_{i'}; A_i \mid z) = 2 * (ln2) * E(z) * MI (A_{i'}; A_i \mid z)$$

where
$E(z)$ is the number of tuples in the node z

Step 2.5.1.2.3 – If the likelihood ratio statistic is significant, mark the node as "split" and increment the conditional mutual information of the candidate input attribute and the target attribute, given the final hidden layer of nodes; else mark the node as "unsplit"

Step 2.5.1.2.4 – Go to the next node

Step 2.5.1.3 Go to the next candidate input attribute

Sep 2.5.2 – Find the candidate input attribute maximizing the conditional mutual information ("the best candidate attribute")

Step 2.5.3 – If the maximum conditional mutual information is greater than zero:

- Make the best candidate attribute an input attribute
- Define a new layer of hidden nodes for a Cartesian product of split hidden nodes of the previous layer and values of the best candidate attribute.

Or else stop search (go to step 2.5)

Step 2.5.4 – Go the next hidden layer

Step 2.6 repeat for every split node z (including the nodes of the final layer)

Step 2.6.1 – Calculate the connection weights linking the unsplit nodes and the nodes of the final layer to the target nodes by

$$w_z{}^{ij} = P(V_{ij} ; z) * \log (P(V_{ij} \mid z) / P(V_{ij}))$$

where
$P(V_{ij} ; z)$ – joint probability of the value V_{ij} and the node z
$P(V_{ij} \mid z)$ – conditional probability of the value V_{ij} , given the node z
$P(V_{ij})$ – unconditional probability of the value V_{ij}

Step 2.6.2 – Select a value j maximizing the conditional probability of the target attribute i at the node z ($P(V_{ij} \mid z)$) and make it the predicted value of the target attribute i at the node z

Step 2.6.3 – Go to the next unsplit node

Step 2.7 Go to the next target attribute

Step 3 – End

7 Results from the Variation of the Original Algorithm

The network models constructed by the variation of the original algorithm are shown in Figure 2 and Figure 3.

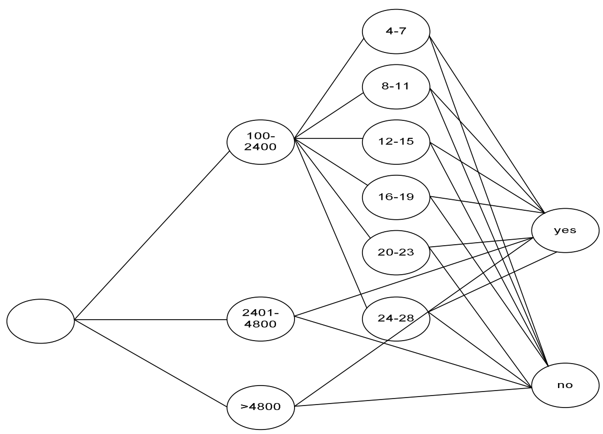

Figure 2. The information theoretic network model 1 from the variation approach

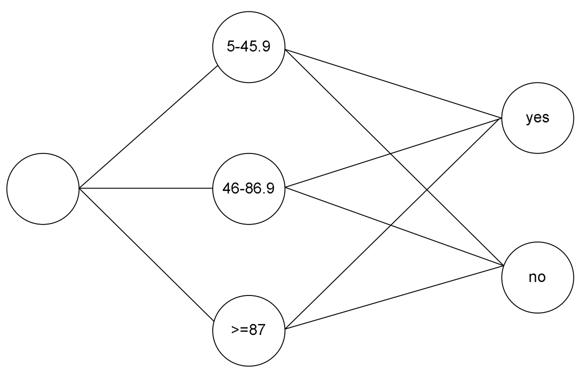

Figure 3. The information theoretic network model 2 from the variation approach

The results from the variation of the original algorithm are

We can predict that the couple should get pregnant if we find that the level of Luteal progesterone is between 46-86.9 nmol/L.

If we want the couple get pregnant, then we should give the gonadotrophin during 8-11 days or 16-23 days before LH peak day, when we find that the maximum level of oestradiol is in the range 10-2400 pg/ml.

The expert sees that the results from the variation approach are better than the results from the original algorithm.

8 Conclusion

In this paper, we presented the using of the information theoretic network approach and the variation of this technique. The results show that the variation approach by using the clinical significance instead of using the statistical significance to construct the first hidden layer of the network gives us better reasonable result than the original algorithm. So we cannot ignore the medical knowledge from medical domain expert in medical data mining.

References

1. Kantardzic, M., *DATA MINING : Concepts, Models, Mthods, and Algorithms*, ed. E.i.C. Stamatios V Kartalopoulos. 2003, USA: IEEE press.
2. Berry, M.J.A. and G.S. Linoff, *Data Mining Techniques and Algorithms*, in *Mastering Data Mining*, R.M. Elliott, Editor. 2000, John Wiley & Sons, Inc.: USA.
3. Groth, R., *Data Mining : Building Competitive Advantage*. 2000, New Jersey, USA: Prentice-Hall.
4. Han, J. and M. Kamber, *Data Mining : Concepts and Techniques*. 2001: Morgan Kaufmann Publishers.
5. Handmer, A., *Fertility treatment produces tiny miracles*, in *The University of Sydney News*. 2003: Sydney. p. 8.
6. Formosa, M. and M.P. Brincat, *Infertility*, in *Understanding Common Disorders in Reproductive Endocrinology*, M.M. Dooley and M.P. Brincat, Editors. 1994, John wiley & sons Ltd: West Sussex, England. p. 67-86.
7. Brudenell, M., et al., *Gynaecology By Ten Teachers*. fifteenth edition ed, ed. T.L. Cbe and G. Chamberlain. 1989: Edward Arnold.
8. Agrawal, r., T. Imielinski, and A. Swami. *Mining Association Rules between sets of items in large databases*. in *In Proc. of the ACM SIGMOD Conference on Management of Data*. 1993. Washington, D.C.
9. Agrawal, R. and R. Srikant. *Fast Algorithms for Mining Association Rules*. in *In Prpceddings of the 20th International Conference Very Large Data Bases*. 1994.
10. Liu, B., et al., *Analyzing Subjective Interestingness of Association Rules.* IEEE Intelligent Systems, 2000. **15**(5): p. 47-55.
11. Tseng, S.-M., *Mining Association Rules with Interestingness Constraints in Large Databases.* International Journal of Fuzzy Systems, 2001. **3**(2): p. 415-421.

12. Silberschatz, A. and A. Tuzhilin. *On Subjective Measures of Interestingness in Knowledge Discovey*. in *in Proceedings of the First International Conference on Knowledge Discovery and Data Mining*. 1995. Montreal, Canada.
13. Wu, X., et al., *Learning Bayesian-network Topologies in Realistic Medical Domains*, in *ISMDA*. 2001. p. 302-308.
14. Taylor, M., *Discovering Multi-Level Classification Rules in Platelet Transfusion Databases*. 1996, University of Maryland, college Park, MD 20742.
15. Salzberg, S., *On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach*, in *Data Mining and Knowledge Discovery*. 1997, Kluwer Academic Publishers: Boston. p. 312-327.
16. Quinlan, J.R., *Data Mining Tools See5 and C5.0.* 2001, RuleQuest.
17. Mehta, M., R. Agrawal, and J. Rissanen. *SLIQ: A Fast Scalable Classifier for Data Mining*. in *In Proceedings of the Fifth International Conference on Extending Database Technology*. 1996. Avignon, France.
18. Cheeseman, P. and J. Stutz, *Bayesian Classification(AutoClass):Theory and Results*, in *Advances in Knowledge Discovery and Data Mining*, R.S. Uthurasamy, Editor. 1996, AAAI Press.
19. Maimon, O., A. Kandel, and M. Last, *Information Theoretic Fuzzy Approach to Knowledge Discovery in Databases*, in *Advances in Soft Computing-- Engineering Design and Manufacturing*, R. Roy, T. Furuhashi, and P.K. Chawdhry, Editors. 1999, Springer-Verlag: London, UK. p. 315-326.
20. Maimon, O. and M. Last, *Knowledge Discovery and Data Mining, the Info-Fuzzy Network (IFN) methodology*. 2000, Norwell, MA: Kluwer.
21. Maimon, O. and M. Last, *Knowledge Discovery and Data Mining : the Info-fuzzy Network (IFN) Methodology*. 2001, Netherlands: Kluwer Academic Publishers. 168.
22. Last, M., Y. Klein, and A. Kandel, *Knowledge Discovery in Time Series Databases.* IEEE Transactions on Systems, Man, and Cybernetics, 2001. **31**(1): p. 160-169.
23. Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, *Advances in Knowledge Discovery and Data Mining*. 1996: AAAI/MIT Press.