Predicting Caesarean Section by Applying Nearest Neighbor Analysis

SUNANTHA SODSEE

KING MONGKUT'S UNIVERSITY OF TECHNOLOGY NORTH BANGKOK, THAILAND

Content

- **Introduction**
  - **Caesarean Section Diagnosis**
- **Overview of Idea**
- **Detail of Work**
- **Simulation Results**
- **Conclusion**

Introduction

- Maternal mortality and childbirth complications are major problem of delivery in rural area of many developing countries.
  - Caesarean section is the first major operation to prevent childbirth complications from high-risk pregnancy.
  - Reports from the United States and around the world have marked a steadily rising Caesarean section rate.
  - In Thailand, at Bhumibol Adulyadej Hospital, 30% of delivery is Caesarean section and indication for Caesarean section from Cephalopelvic disproportion is 5-7% and 8.8% at Lamphun Hospital.
- It would be beneficial
  - if the risk of delivery from uncertainty information could be informed or recommended to patients at earlier sign.
- Physicians could draw approximate decision before it occurred.

Introduction

- Due to the medical care, the patients' care histories are necessary to be analysed for diagnosing disease, as well as the nearest medical patterns of patients could be, however, considered too.
- This paper proposes
  - a modified nearest neighbor analysis, which is called CPD-NN algorithm to approximate risks about Caesarean sections due to Cephalopelvic disproportion (CPD).
- CPD-NN algorithm consists of three phases:
  - initial phase
  - distance measure phase
  - predicting phase

Overview of Idea

- The algorithm CPD-NN to be developed is to create a model in form of a cluster set from medical history data, with each cluster representing a group of similar medical patterns.
- The nearest medical patterns of patients are analysed to diagnose.
- To determine a group of similar medical patterns,
  - the threshold distance D_{min} is set to identify the closest medical patterns.
  - D_{max} is defined to identify the farthest medical patterns.

Detail of Work

- Medical patterns of pregnant women are acquired.
- Some features of patterns are selected by experts, who are physicians,
  - and then the modified nearest neighbour analysis is applied for classifying the medical patterns in order to predict the risk to CPD.
- Finally, the probability analysis is deployed to estimate the Caesarean section.

Detail of Work

- **Step0**: Initial phase
  - Input a set of training medical pattern
  - Estimate the number of training medical pattern (n)
  - Set the threshold distance ($0<D_{min}< 1$), to identify the closest medical patterns (the closest neighbors).
  - Set the radius of cluster, which is defined by the distance ($0<D_{max}<1$) in order to identify the farthest medical patterns (the farthest neighbors).Hence, neighbors are located within the cluster.

- **Step1**: Distance measure phase ($0<D_i<1$)
  - Input a set of test medical patterns
  - Calculate the distance (D_i) between a test medical pattern (i) and all training medical patterns
  - Determine the nearest neighbor, and the farthest neighbors, who are considered based on D_{min} and D_{max}.
  - The '*k*' nearest neighbours are located within the cluster between D_{min} and D_{max}, then '*k*' is dynamic.

Detail of Work

- **Step2**: Predicting phase
  - Collect the targets of the '*k*' nearest neighbors, which are located within the specified area of cluster
  - Estimate the target of test medical patterns by using probability analysis among collected targets of '*k*' nearest neighbors
  - Assign a target to the test medical patterns (*i*) by using majority of closest neighbors estimated target probability

Simulation Results

- The medical dataset used comprises 802 cases of pregnant women who delivered in a public hospital in Thailand between 1 July 2005 and 31 May 2007.
- It contains two equally sized groups, namely 401 women who delivered by Caesarean section due to cephalopelvic disproportion, and 401 women delivered normally during the same period.
- Attributes in the dataset relevant for consideration by the proposed algorithm were selected by experts

Simulation Results

- *Applying ordinary k-nearest neighbors algorithm for diagnosing CPD:*
  - The highest accuracy is around 75% when applying 400 training cases, 402 test cases with the nearest rule, Manhattan distance (city block) and the size of '*k*' is 25% of training cases.

# Training Cases	Distance Measure	Rule	Size of '*k*'(%)	Accuracy
100	City block	Nearest, Random	5	66.09
400	City block	Nearest	25	74.62

Simulation Results

- *Applying the CPD-NN algorithm for diagnosing CPD:*
  - D_{min} is defined to 0.1, D_{max} is varied from 0.3 to 0.9.

Distance Measure	D_{max}	Accuracy	Average '*k*'	Sensitivity	Specificity
Euclidean	0.3	92.29	6.99	1	0.83
	0.5	99.50	18.22	1	0.98
	0.7	100.00	19.75	1	1
	0.9	100.00	19.92	1	1
City block	0.3	59.20	0.126	1	0.10
	0.5	77.61	1.42	1	0.50
	0.7	91.29	6.22	1	0.80
	0.9	97.76	12.69	1	0.95
Cosine	0.3	100.00	20	1	1
	0.5	100.00	20	1	1
	0.7	100.00	20	1	1
	0.9	100.00	20	1	1
Correlation	0.3	100.00	19.64	1	1
	0.5	100.00	19.95	1	1
	0.7	100.00	20	1	1
	0.9	100.00	20	1	1

Simulation Results

- *Applying the CPD-NN algorithm for diagnosing CPD:*
  - The prediction accuracy when applying 400 training cases and 402 test cases. The result shows that prediction accuracy is 100 % when cosine similarity and correlation are applied with '*k*' ≈ 5.

Distance Measure	D_{max}	Accuracy	Average '*k*'	Sensitivity	Specificity
Euclidean	0.3	63.96	1.62	1	0.62
	0.5	98.43	4.53	1	0.96
	0.7	99.85	4.92	1	1
	0.9	100	4.98	1	1
City block	0.3	51.28	0.03	1	0.02
	0.5	61.82	0.34	1	0.23
	0.7	80.34	1.48	1	0.60
	0.9	92.00	3.14	1	0.84
Cosine	0.3	100	5	1	1
	0.5	100	5	1	1
	0.7	100	5	1	1
	0.9	100	5	1	1
Correlation	0.3	100	4.90	1	1
	0.5	100	4.99	1	1
	0.7	100	5	1	1
	0.9	100	5	1	1

Conclusion

- We proposed a modified *k*-nearest neighbors algorithm, which is called CPD-NN algorithm, in order to diagnose the Caesarean sections due to Cephalopelvic disproportion (CPD).
- The CPD-NN algorithm employs the threshold distance D_{min} and D_{max} to identify the specified area for "*k*" nearest neighbors.
- It means that "*k*" value, here, is dynamical relating to D_{min} and D_{max}.
- the CPD-NN algorithm shows high diagnosis performances, not only prediction accuracy but also sensitivity and specificity.
- To increase the effectiveness of diagnosis, the evidence-based medical diagnosis will be considered in the future work.

References

- *Khunpradit S*, Patumanond J, Tawichasri C.Risk Indicators for Cesarean Section due to Cephalopelvic Disproportion in Lamphun Hospital.*Journal of The Medical Association of Thailand*.2005;88:63–68.
- Wianwiset W. Risk Factors of Cesarean Delivery due to Cephalopelvic Disproportion in Nulliparous Women at Sisaket Hospital.*Thai Journal of Obstetrics and Gynaecology*. 2011;19:158–164.
- Moryadee S, Smanchat B, Rueangchainikhom W, Phommart S. Risk Score for prediction of Caesarean Delivery due to Cephalopelvic Disproportion in BhumibolAdulyadej Hospital.*Royal Thai Air Force Medical Gazette*.2010; 56:1.
- Komkhao A. Delivery due to Cephalopelvic Disproportion in BhumibolAdulyadej Hospital.*Thailand, Royal Thai Air Force Medical Gazette*. 2008;54:70.
- Spörri S. MR Imaging Pelvimetry: A Useful Adjunct in the Treatment of Women at Risk for Dystocia? *American Journal of Roentgenology*. 2002;179:1:137–144.
- Han J, Kamber M,Pai J.*Data Mining: Concepts and Techniques*.3rd ed. Waltham: MA: Morgan Kaufmann Publishers; 2012.
- Tsipouras MG, Exarchos TP, Fotiadis DI, Kotsia AP, Vakalis KV, Naka KK, Michalis LK.Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling.*IEEE Transactions on Information Technology in Biomedicine*. 2008;12:4:447–458.
- Karaolis MA, Moutiris JA, Hadjipanayi D, Pattichis CS. Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining with Decision Trees.*IEEE Transactions on Information Technology in Biomedicine*. 2010;14:3:559–566.
- Barakat NH, Bradley AP, Barakat MNH.Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus.*IEEE Transactions on Information Technology in Biomedicine*. 2010;14:4:1114–1120.
- Isola R, Carvalho R, Tripathy A. Knowledge Discovery in Medical Systems Using Differential Diagnosis LAMSTAR, and k-NN.*IEEE Transactions on Information Technology in Biomedicine*. 2012;16:6:1287–1295.
- Jain AK, Murthy MN, Flynn PJ. Data Clustering: A Review. *ACM Computing Surveys*. 1999;31:3:264–323.
- Jain AK, Nandakumar K, Ross A. Score Normalization in Multimodal Biometric Systems.*Pattern Recognition*. 2005;38:12:2270–2285.