



Predicting Postoperative Nausea and Vomiting Under Patient-Controlled Analgesia Medication: A Study of Machine Learning Approaches

Yuh-Jyh H^{1*}, Jia-Ying S¹ and Tien-Hsiung K²

¹Institute of Biomedical Engineering, National Chiao Tung University, Taiwan

²Department of Anesthesiology, Changhua Christian Hospital, Taiwan

Abstract

In addition to pain, nausea and vomiting persist as the most frequent complaints of patients receiving patient-controlled analgesia (PCA) after surgery. Many patients find postoperative nausea and vomiting (PONV) even more distressing than postoperative pain. Though many studies have evaluated the correlation of patient characteristics with PONV and identified several risk factors, there is little research into constructing models for PONV prediction. In this study, we proposed to analyze patient behaviors and apply machine learning methods to PONV prediction. We evaluated different learning algorithms, and investigated several data preprocessing techniques. We performed a thorough comparative study of machine learning techniques, and the experimental results suggest the application of machine learning to PONV prediction is feasible and promising.

Keywords: PONV (post-operative nausea and vomiting); PCA (patient-controlled analgesia); Machine learning; PCA demand behaviors; Feature selection; Data cleaning

Introduction

With the advance of medical science, people have gradually become aware of the importance of pain management because pain can negatively affect quality of health care and even do more harm than an illness itself when it becomes intolerable. According to the studies, PCA (patient-controlled analgesia) is one of the most effective techniques for postoperative analgesia [1,2].

Despite the fact that IV-PCA (Intravenous PCA) has been widely used in hospitals for its effectiveness and safety as acute postoperative pain management, PCA also usually entails (PONV) post-operative nausea and vomiting that complicates recovery from surgery and decreases patient satisfaction [3,4]. In some studies patients were, on average, willing to pay extra \$56 to avoid PONV; the figure increased to \$73 and \$100 in patients who had experienced postoperative nausea or vomiting, respectively [5,6].

Most previous studies of PONV were focused on identifying the risk factors, using regression techniques or proposing probabilistic models [7-10]. A recent work that applied an artificial neural network to predict postoperative vomiting has been proposed [11]. In this study, we investigated patient PCA demand behaviors, and derived demand pattern attributes by clustering demand profiles for PONV prediction. In addition, we proposed to use a neighborhood-based data cleaning technique to clarify class boundary. Lastly, we conducted a comparison of various machine learning classifiers to identify the best feature set and classifiers for PONV prediction. Our goal is to improve PONV prediction to increase patient satisfaction by applying machine learning methods and analyzing IV-PCA patient demand behaviors.

Materials and Methods

Study subjects

We collected and analyzed IV-PCA usage profiles from bone surgery patient records from 2009 to 2014 at Changhua Christian Hospital. Abbott Pain Management Provider (Abbott Lab, Chicago, IL, USA) was used for IV-PCA treatment. After excluding incomplete IV-PCA log files and patient records with missing values, we obtained 392 patient records. Of these 392 subjects, 121 had PONV and the remaining 271 showed no PONV. Each patient received at least 24 h of

IV-PCA medication without using any antiemetic drugs. Each subject is represented by totally 28 basic attributes divided into 5 categories: (a) demographic, (b) biomedical, (c) operation-related, (d) opioid-related, and (e) PCA-related attributes.

PCA demand behavior pattern attributes

In addition to commonly studied demographic and physiological factors relevant to analgesic consumption, IV-PCA related attributes, such as the number of demands per hour, have been shown to correlate significantly with analgesic consumption prediction [11,12]. These findings suggest that these demand behavior-related attributes are likely to correlate with incidences of postoperative nausea and vomiting. To generate behavior pattern attributes for PONV prediction, we considered two types of pattern attributes based on time domain and frequency domain, respectively.

For time-based behavior pattern attributes, we first characterize different IV-PCA demand behaviors in the course of time. We retrieved the IV-PCA demand data from each patient's IV-PCA treatment log file and derived three types of IV-PCA profiles based on (a) the number of successful IV-PCA demands in each time unit, (b) the number of failed IV-PCA demands in each time unit, and (c) the IV-PCA dose for each time unit. Four different time units were used in this study: 60 min, 45 min, 30 min and 20 min. We show a sample IV-PCA time-based dose profile in Figure 1.

From a time-based behavior pattern we can observe the change in the number of PCA demands and the amount of analgesic consumption; however, we cannot distinguish the distributions of PCA demands in different frequencies. Therefore, we also applied Fourier transform to time-based profiles to obtain a frequency-based profile. A sample

*Corresponding author: Yuh-Jyh H, Institute of Biomedical Engineering, National Chiao Tung University, Taiwan, Tel: 886 3 571 2121; E-mail: yhu@cs.nctu.edu.tw

Received September 22, 2017; Accepted September 25, 2017; Published September 30, 2017

Citation: Yuh-Jyh H, Jia-Ying S, Tien-Hsiung K (2017) Predicting Postoperative Nausea and Vomiting Under Patient-Controlled Analgesia Medication: A Study of Machine Learning Approaches. Prim Health Care 7: 272. doi: [10.4172/2167-1079.1000272](https://doi.org/10.4172/2167-1079.1000272)

Copyright: © 2017 Yuh-Jyh H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

frequency-based IV-PCA dose profile transformed from Figure 1 is shown in Figure 2.

After the process of various IV-PCA profiles, we applied k-medoid clustering to these profiles to identify significant demand patterns among the study patients. Figure 3 shows the four patterns identified in the time-based IV-PCA dose profiles of the 392 patients in a 12 h time period [13].

The demand profiles grouped into a cluster demonstrated similar demand behaviors, and the medoid of a cluster represented the behavior pattern for that cluster over time. By applying k-medoids to different IV-PCA demand profiles, we generated different demand-pattern attributes. We expected the inclusion of demand patterns of the first few hours of IV-PCA usage to improve PONV prediction.

Feature selection

We used 28 basic patient attributes, classified in 5 categories, to describe each study subject. In addition, we derived a number of different PCA demand pattern attributes from various PCA demand

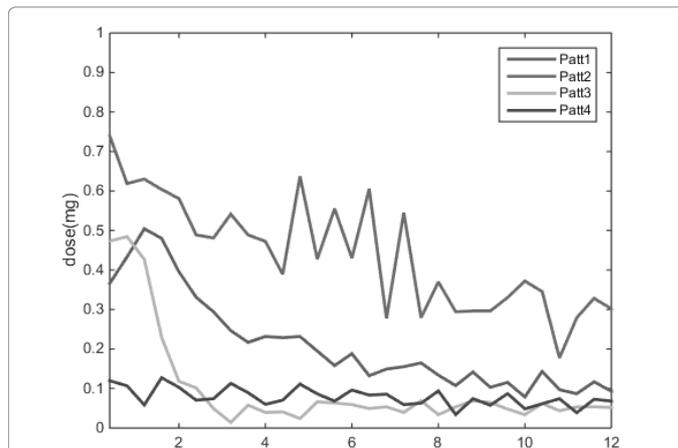


Figure 3: Average IV-PCA demand behavior in each cluster. The X-axis indicates the 12 h time line. The Y-axis represents the PCA dose within a particular 20 min time unit.

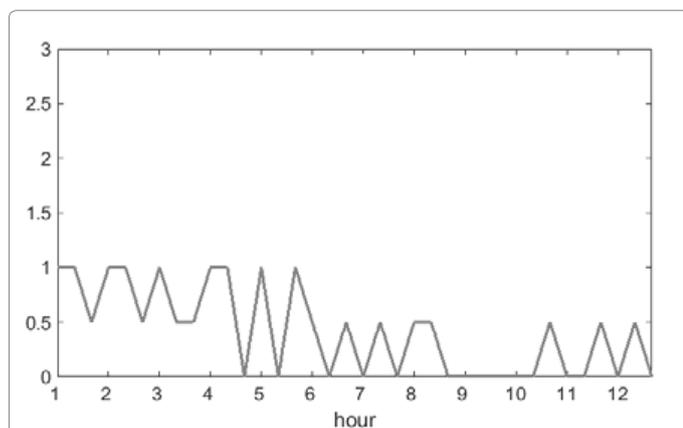


Figure 1: A sample time-based IV-PCA dose profile in a 12 h time period. The X-axis is time and the Y-axis is the total PCA dose in each 20 min interval.

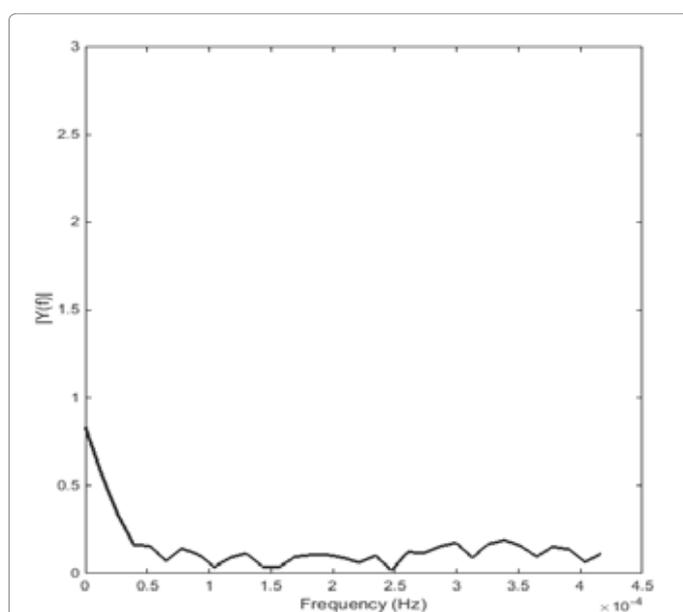


Figure 2: A sample frequency-based IV-PCA dose profile in a 12 h time period.

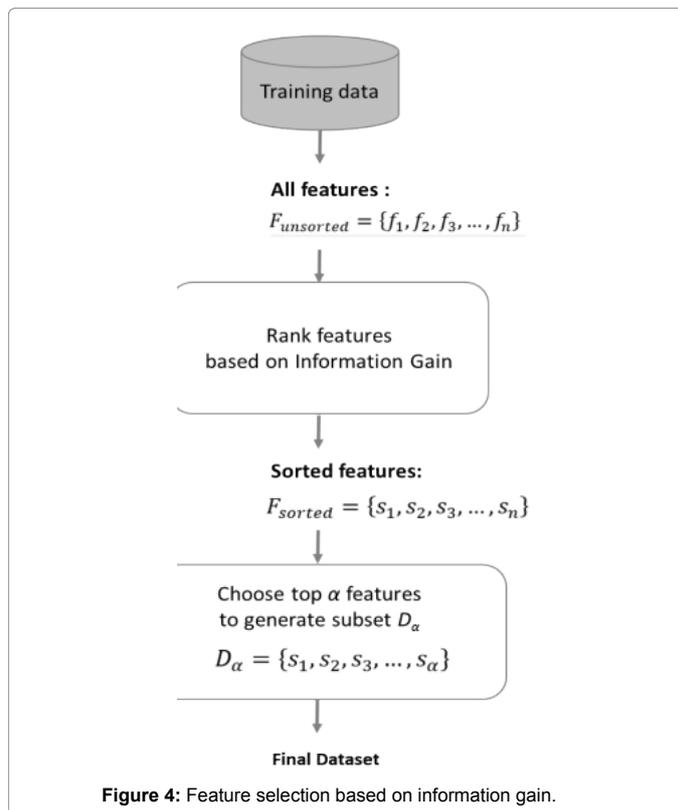


Figure 4: Feature selection based on information gain.

profiles, based on different time units, different demand reference (e.g. dose or successful demand), and various values of k for k-medoids clustering. Though these attributes can characterize patient behaviors, they may also negatively interact with those 28 basic attributes. To avoid negative interaction among the features, we selected important features according to their information gain and used only these selected features to represent each patient. We show the feature selection process in Figure 4.

Data cleaning

Nausea and vomiting are most common adverse effects of IV-

PCA with reported incidence of 3.1 to 34% [14,15]. From the point of machine learning, prediction of nausea and vomiting is a classification problem in an imbalanced class domain. Conventional machine-learning algorithms are typically biased toward the majority class, and produce poor predictive accuracy for the minority class. In addition to unequal class distribution, instances sparsely scattered in the data space make the prediction of a minority class even more difficult. We applied a neighborhood-based data cleaning approach to remove spurious data points of the majority class. It first identifies the k -nearest neighbors of each instance of the minority class and considers any majority class neighbor as “dirty.” After examining each instance in the minority class and its neighbors, the proposed approach removes those “dirty” instances. The rationale behind this process is that the nearest majority class neighbors of a minority class member are likely to mislead learning algorithms. Without them, learning algorithms can more easily recognize the minority class boundary. We illustrate the concept in Figure 5. Figure 5a shows an imbalanced data set before removing “dirty” instances. The rectangles in this figure represent the decision regions of the minority class, and several majority class examples are also included. The proposed approach first locates the k -nearest neighbors (e.g. $k=3$) for each minority class example and then presents the neighbors as linked to each minority class example (Figure 5b) and crosses out the “dirty” majority class neighbors (Figure 5c). Removing the “dirty” examples produces the “clean” decision regions of the minority class (Figure 5d).

performance measures

We evaluated prediction performances by using several measures: percentage accuracy, F-score and MCC. Table 1 lists the definitions of these measures.

For the problem of PONV prediction, high true positive rate is more desirable compared with other measures, e.g. accuracy, because nausea and vomiting, in addition to pain, are the most frequent negative effects of patient satisfaction and the number of patients showing PONV is significantly smaller than those showing no PONV (3.1~34% PONV). Therefore, our goal is to apply machine learning techniques to obtain the highest F-score rather than the overall accuracy. We show the other performance measures for reference.

Machine learning classifiers in comparison

We tested 9 classifiers on PONV prediction. These classifiers can

Performance Measure	Definition
Recall	$TP/(TP+FN)$
Precision	$TP/(TP+FP)$
F-score	$2 \times \text{Recall} \times \text{Precision}/(\text{Recall}+\text{Precision})$
MCC	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$

Table 1: Performance measures.

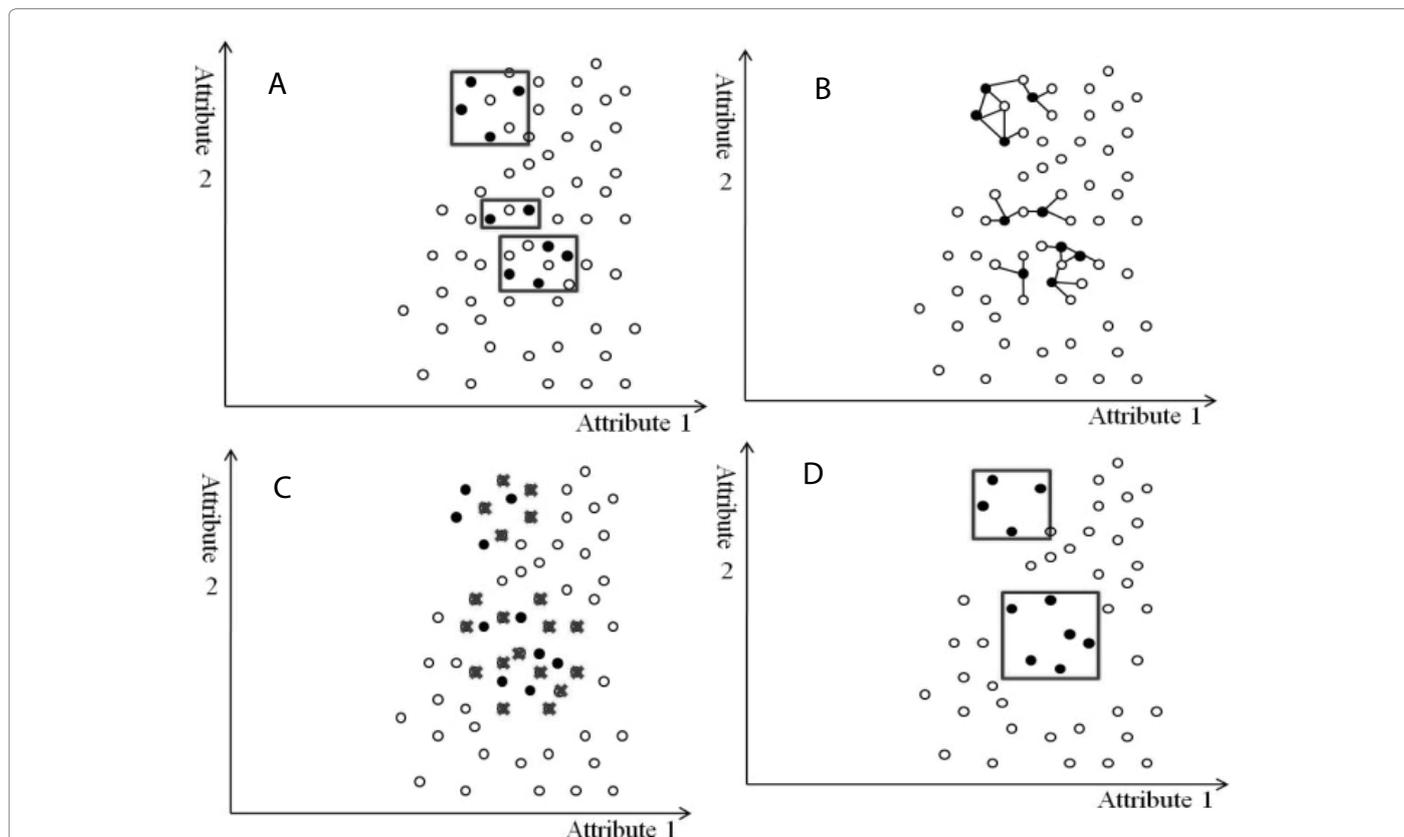


Figure 5: The X- and Y-axes represent two attributes in the feature space. The minority class examples are denoted by black circles and the majority class examples are denoted by white circles. Black rectangles indicate the axis-parallel decision regions of the minority class. (a) We show an imbalanced data set with sparse minority class examples. The decision regions of the minority class contain the majority class examples. (b) To identify the “dirty” examples that may mislead learning, the proposed method locates k -nearest (where k is 3 in this example) neighbors for each minority class example. The 3-nearest neighbors of a minority class example are indicated by links. (c) A black cross marks each “dirty” example. (e) After the “dirty” examples are removed, the decision regions are “clean” (i.e., they contain only the minority class examples). Using these clean decision regions, learning algorithms can more easily recognize the correct boundary between classes.

be characterized into six categories: (a) decision-based, (b) instance-based (c) probabilistic, (d) neural network, (e) feature-based, and (f) ensemble. We list the classifiers in Table 2. These classifiers have different design philosophies and applicability. There is little research into applications of machine learning to postoperative nausea and vomiting prediction. Through a comparative study, we intended to identify the superior classifiers and the appropriate patient features for PONV prediction.

Results

The goal of this study is twofold: (1) to compare the effects of different types of patient features as PONV risk factors, and (2) to evaluate the performance of different machine learning techniques for predicting PONV. To conduct a comparative study of risk factors, we divided patient features into 3 groups: (a) basic patient features, including demographic, biomedical, operation-related, and analgesics-related attributes, (b) in addition to basic features, PCA-related attributes are included, and (c) the complete feature set with behavior pattern attributes added. We performed experiments to evaluate the feasibility of various machine learning techniques, namely feature selection, data cleaning, and classification, and verified the synergy of the combination of these techniques. The experiments were conducted by performing stratified 10-fold cross-validation of 392 study subjects.

Experiment of classifiers using different groups of patient features

We tested the classifiers listed in Table 2, using different groups of patient features. We present the results in Table 3. The results for each classifier are presented in the order of groups (a), (b), (c) based on time domain, and (c) based on frequency domain, separately. For each classifier, we also performed a paired *t*-test between using group (a) and using the other feature groups, individually. A significant difference (p -val<0.05) is indicated by a star symbol.

Table 3 shows that the addition of more features (PCA-related and behavior pattern) had little effect on most of the classifiers in study. On the other hand, we observed that these extra features could adversely hinder the learning of particular types of classifiers such as probabilistic learners, and feature-based learners. It suggests that the interactions incurred by more features significantly affect some classifiers. This finding reconfirmed that these learning algorithms have their own distinct characteristics and different applicability.

Experiment of feature selection

We hypothesized that the addition of extra features did not show improvement for PONV prediction in the first experiment was mainly due to adverse feature interactions. To verify our hypothesis, we first selected important features based on their information gain and then

Classifier	Category
PART [16]	Decision-based
LADTree [17]	Decision-based
K* [18]	Instance-based
K-NN [19]	Instance-based
Logistic Regression [20]	Probabilistic
Bayes Net [21]	Probabilistic
ANN [22]	Neural network
VFI [23]	Feature-based
Random Forest [24]	Ensemble

Table 2: List of classifiers.

Classifier	F-score	MCC	Accuracy
PART	0.368	0.090	0.611
	0.369	0.104	0.628
	0.342	0.066	0.607
	0.373	0.138	0.648
LADTree	0.294	0.072	0.641
	0.391	0.147	0.646
	0.316	0.072	0.628
	0.336	0.077	0.615
K*	0.333	0.023	0.571
	0.333	0.063	0.605
	0.321	0.019	0.581
	0.371	0.072	0.592
K-NN	0.341	0.078	0.619
	0.344	0.093	0.630
	0.256	0.021	0.620
	0.315	0.039	0.602
Logistic Regression	0.369	0.168	0.671
	0.385	0.127	0.633*
	0.366	0.019	0.538*
	0.356	0.019	0.546*
Bayes Net	0.516	0.291	0.689
	0.501	0.268	0.676*
	0.395*	0.032*	0.513*
	0.435	0.098	0.546*
ANN	0.440	0.185	0.648
	0.363	0.097	0.617
	0.353	0.078	0.605
	0.383	0.122	0.628
VFI	0.513	0.251	0.638
	0.517	0.254	0.641
	0.402*	0.019*	0.495*
	0.433	0.095	0.543*
Random Forest	0.260	0.076	0.656
	0.270	0.130	0.681
	0.166	-0.044	0.620
	0.255	0.069	0.645

*Indicates a significant difference (p -val<0.05) from group (a)

Table 3: Results of classifiers using different features.

re-ran the experiment, using the selected features. We show the results after feature selection in Table 4. Like in Table 3, we present the results for each classifier in the order of groups (a), (b), (c) based on time domain, and (c) based on frequency domain, separately. Compared with those in Table 3, the numbers are presented in italics to indicate no performance improvement or performance decrease after feature selection.

According to Table 4, we clearly verify the merits of feature selection. The F-score and MCC have been substantially increased for all classifiers after feature selection, which indicates that feature selection resolves the feature interaction problem. As for the comparison between feature group (a) and the others, there is no significantly lower performance for groups (b) and (c) than group (a). On the contrary, we identified several significant positive results after feature selection. For example, the addition of PCA-related features and behavior-derived patterns increased F-score and MCC significantly for VFI and Random Forest.

We list the top-6 features in Table 5. The top-3 features are the demographic attributes, among which sex has been reported to be one

Classifier	F-score	MCC	Accuracy
PART	0.427	0.240	0.696
	0.478	0.263	0.689
	0.425	0.218	0.684
	0.415	0.183	0.656
LADTree	0.403	0.148	0.636
	0.403	0.148	0.636
	0.394	0.188	0.666
	0.352	0.105	0.628
K*	0.448	0.206	0.648
	0.422	0.208	0.669
	0.464	0.242	0.679
	0.458	0.248	0.687
K-NN	0.426	0.192	0.663
	0.399	0.162	0.655
	0.434	0.222	0.679
	0.418	0.193	0.663
Logistic Regression	0.518	0.313	0.699
	0.518	0.313	0.699
	0.498	0.316	0.707
	0.455	0.272	0.707
Bayes Net	0.523	0.306	0.697
	0.523	0.303	0.694
	0.519	0.310	0.702
	0.521	0.306	0.700
ANN	0.537	0.304	0.682
	0.531	0.311	0.694
	0.470	0.221	0.659
	0.472	0.258	0.686
VFI	0.553	0.322	0.682
	0.553	0.322	0.682
	0.555	0.316	0.684
	0.611*	0.404*	0.699
Random Forest	0.401	0.157	0.648
	0.424	0.184	0.656
	0.486*	0.267*	0.689
	0.448	0.231	0.679

* Indicates a significant difference (p -val<0.05) from group (a)

Table 4: Results of applying feature selection.

Significant Features
Surgery Size
Sex
Patient Height
Frequency-based Behavior Patterns, using 20 min time units
Frequency-based Behavior Patterns, using 60 min time units
Time-based Behavior Patterns, using 60 min time units

Table 5: Significant patient features as risk factors.

strong factor for PONV in several studies, and our study reconfirmed this finding. In addition, we also identified patient height to be an important factor, which agrees with a similar finding has been reported in a survival analysis [25-27]. The remaining are pattern features that characterize PCA patient demand behaviors.

Experimnt of data cleaning

While nausea and vomiting are most common adverse effects of IV-PCA, the incidence of PONV is relatively low, which makes the machine learning task an imbalanced classification problem. In this

Classifier	F-score	MCC	Accuracy
PART	0.526	0.249	0.597
	0.560	0.305	0.633
	0.521	0.244	0.585
	0.553	0.294	0.574
LADTree	0.544	0.273	0.602
	0.544	0.273	0.602
	0.549	0.294	0.638
	0.546	0.273	0.551*
K*	0.519	0.219	0.508
	0.535	0.275	0.631*
	0.535	0.268	0.618*
	0.576*	0.344*	0.630*
K-NN	0.536	0.267	0.525
	0.552	0.296	0.597*
	0.531	0.286	0.657*
	0.566	0.320	0.628
Logistic Regression	0.561	0.317	0.635
	0.561	0.317	0.635
	0.577	0.352	0.671
	0.539	0.272	0.526
Bayes Net	0.550	0.290	0.554*
	0.544	0.284	0.564*
	0.591	0.366	0.671*
	0.563	0.322	0.649
ANN	0.558	0.313	0.644
	0.558	0.301	0.564*
	0.548	0.279	0.584*
	0.557	0.307	0.618
VFI	0.557	0.307	0.618
	0.557	0.307	0.618
	0.613*	0.406*	0.697
	0.554	0.298	0.594
Random Forest	0.516*	0.230*	0.582
	0.549	0.291	0.605
	0.561	0.317	0.602

* Indicates a significant difference (p -val<0.05) from group (a)

Table 6: Results of data cleaning.

study, we proposed to apply a neighborhood-based data cleaning method to better balance the classes and reveal a clearer class boundary by removing redundant data points of the major class.

Discussion

After feature selection, we performed data cleaning. We compared the effects of data cleaning for feature groups (a), (b) and (c). We show the results in Table 6. Compared with those in Table 4, the numbers are presented in italics to indicate no performance improvement or performance decrease after data cleaning. From Table 6 we notice that data cleaning improved F-score and MCC for most of the classifiers except that the MCCs of Bayes Net and VFI decreased. Nevertheless, it is worth notice that when frequency-based behavior pattern features were used, data cleaning increased both F-score and MCC for all classifiers, and VFI produced the highest performance for F-score and MCC. In contrast to F-score and MCC, accuracy of all the classifiers decreased variably after data cleaning in exchange for higher F-score and MCC. For an imbalanced class prediction problem such as PONV, F-score and MCC are more appropriate measures than accuracy, and we have verified that our data cleaning method can warrant better performance.

Conclusion

Despite advancements in postoperative pain management, postoperative patient satisfaction remains inadequate in a large fraction of hospitalized patients. In addition to pain, nausea and vomiting have been the most distressing side effects of IV-PCA. Significant efforts have been focused on identifying and analyzing risk factors for PONV [8-10] whereas few previous works ever tested the identified factors for evaluating their predictive strengths. Unlike most previous research that mainly adapted statistical approaches, we not only applied machine learning methods for PONV prediction, but also made a thorough comparison of their performances. In addition, we proposed to consider patient PCA demand behaviors to improve PONV prediction. We conducted stratified 10-fold cross-validation, and the results confirmed the feasibility of the application of machine learning to pain management.

Acknowledgement

This work is partially supported by Ministry of Science and Technology of Taiwan (MOST 106-2221-E-009-184). The authors thank the department of anesthesiology at Changhua Christian Hospital for providing the IV-PCA patient data, and participating in this study.

References

1. Dolin SJ, Cashman JN, Bland JM (2002) Effectiveness of acute postoperative pain management: Evidence from published data. *Br J Anaesth* 89: 409-423.
2. Walder B, Schafer M, Henzi I, Tramer MR (2001) Efficacy and safety of patient-controlled opioid analgesia for acute postoperative pain. A quantitative systematic review. *Acta Anaesthesiol Scand* 45: 795-804.
3. Gan T (2002) Postoperative nausea and vomiting – Can it be eliminated?. *JAMA* 287: 1233-1236.
4. Koivuranta M, Läärä E, Snåre L, Alahuhta S (1997) A survey of postoperative nausea and vomiting. *Anaesthesia* 52: 443-449.
5. Gan T, Sloan F, Dear Gde L, El-Moalem HE, Lubarsky DA (2001) How much are patients willing to pay to avoid postoperative nausea and vomiting. *Anesth Analg* 92: 393-400.
6. Scuderi PE, Conlay LA (2003) Postoperative nausea and vomiting and outcome. *Int Anesthesiol Clin* 41: 165-174.
7. Palazzo M, Evans R (1993) Logistic regression analysis of fixed patient factors for postoperative sickness: A model for risk assessment. *Br J Anaesth* 70: 135-140.
8. Stadler M, Bardiau F, Seidel L, Albert A, Boogaerts JG (2003) Difference in risk factors for postoperative nausea and vomiting. *Anesthesiology* 98: 46-52.
9. Gan T (2006) Risk factors for postoperative nausea and vomiting. *Anesth Analg* 102: 1884-1898.
10. Apfel C, Greim C, Haubitz I, Goepfert C, Usadel J, et al. (1998) A risk score to predict the probability of post-operative vomiting in adults. *Acta Anaesthesiol Scand* 42: 495-501.
11. Hu YJ, Ku TH (2012) Pattern discovery from patient controlled analgesia demand behavior. *Comp biol med* 42: 1005-1011.
12. Hu YJ, Ku TH, Yang JY (2017) Prediction of patient-controlled analgesic consumption: A multimodel regression tree approach. *IEEE J Biomed Health Inform.*
13. Reynolds AP, Richards G, Rayward-Smith VJ (2004) The application of k-medoids and pam to the clustering of rules. *Intelligent Data Engineering and Automated Learning*, pp: 173-178.
14. Dolin SJ, Cashman JN (2005) Tolerability of acute postoperative pain management: Nausea, vomiting, sedation, pruritis and urinary retention. Evidence from published data. *Br J Anaesth* 95: 584-591.
15. Werner WU, Soholm L, Rotboll-Nielsen P, Kehlet H (2002) Does an acute pain service improve post-operative outcome. *Anesth Analg* 95: 1361-1372.
16. Frank E, Witten IH (1998) Generating accurate rule sets without global optimization.
17. Holmes G, Pfahringer B, Kirkby R, Frank E, Hall M (2002) Multiclass alternating decision trees. *European Conference on Machine Learning*, pp: 161-172.
18. Cleary JG, Trigg LE (1995) An instance-based learner using an entropic distance measure. *Proceedings of the 12th International Conference on Machine learning*, pp: 108-114.
19. Duda RO, Hart PE, Stork DG (2001) *Pattern Classification 2nd ed.* New York, NY, USA: Wiley, pp: 182-188.
20. Cox DR (1992) *Regression models and life-tables in breakthroughs in statistics*, ed: Springer, pp: 527-541.
21. Pearl J (2000) *Causality: Models, reasoning and inference.* Cambridge University Press.
22. Bishop CM (1995) *Neural networks for pattern recognition.* Oxford university press.
23. Demiröz G, Güvenir HA (1997) Classification by voting feature intervals. *European Conference on Machine Learning*, pp: 85-92.
24. Breiman L (2001) Random forests. *Machine Learning* 45: 5-32.
25. Apfel CC, Kranke P, Greim CA, Roewer N (2001) What can be expected from risk scores for predicting post-operative nausea and vomiting. *Br J Anaesth* 86: 822-827.
26. Gong CS, Yu L, Ting CK, Tsou MY, Chang KY, et al. (2014) Predicting postoperative vomiting for orthopedic patients receiving patient-controlled epidural analgesia with the application of an artificial neural network. *Biomed Res Int.*
27. Lee SY, Hung CJ, Chen CC, Wu CC (2014) Survival analysis of post-operative nausea and vomiting in patients receiving patient-controlled epidural analgesia. *J Chin Med Assoc* 77: 589-593.