

Impact of discretization with multivariate sequential patterns to do the classification of the survival prediction in Intensive Care Burn Unit

Isidoro J. Casanova¹, Manuel Campos ^{*1}, Jose M. Juarez¹, Antonio Fernandez-Fernandez-Arroyo^{2,3}, and Jose A. Lorente^{2,3,4}

¹ Computer Science Faculty, University of Murcia.

² University Hospital of Getafe.

³ European University of Madrid.

⁴ CIBER Enfermedades Respiratorias.

Abstract. Choosing the method of discretization can have a significant impact on the performance and accuracy of classification algorithms. In this article we compare global equal frequency discretization, one modification of this unsupervised method, in which the equal frequency discretization is made daily and two supervised methods, expert discretization (reference range based) and entropy-based discretization, applied into Intensive Care Burn Units to predict the survival of patients.

1 Introduction

Classification of temporal data, especially of both multivariate and uni-variate time series, is a highly challenging as well as an important task in many domains. It is essential in a multitude of different medical domains, in which correct classification of time-series data has immediate implications for diagnosis, for quality assessment, and for prediction of meaningful outcomes [13].

In general, time-series classification has two main approaches, feature extraction and direct comparison. When time-series are multivariate, sort and are not periodic, those approaches have limitations. Instead, we can generate patterns from this temporal data, and discretization is a fundamental issue.

The method by which continuous features are partitioned into discrete intervals can have a great impact on the accuracy and performance of classification algorithms. Ideally, discretization should result in partitions that (a) reflect the original distribution of the continuous attribute, (b) maintain any patterns in the attribute without adding spurious ones, and (c) are meaningful and interpretable to domain experts [11].

Discretization approaches are normally made by probability distribution or using statistic parameters like the frequency in each class. The discretization can also be made by the experts on the field in a manual way. Though many different discretization algorithms have been devised and evaluated, few studies have examined the discretization of clinical data specifically [8].

* Corresponding author: manuelcampos@um.es

In this article we compare the effect of discretization in the classification of patients in Intensive Care Burn Unit from the daily evolution of the patients using multivariate sequential patterns. We will use six time series of both laboratory and physiologic data, and we will test four different discretization methods: one unsupervised method (equal frequency), one modification of this method, in which the equal frequency discretization is made daily, and two supervised methods, one specific to clinical data with both supervised and unsupervised components (reference range based) and entropy-based method discretization. We discuss about the classification performance, the effect in the number of patterns and the interpretability of the results.

This paper is organized as follows. Section 2 describes the methods for discretization and for survival prediction in Intensive Care Burn Unit. Section 3 describes the case study and the four different discretizations. Section 4 shows the results and discussion. Finally, we provide the conclusions and future works.

2 Literature review

2.1 Discretization

Many studies show that induction tasks can benefit from discretization because discrete features are closer to a knowledge-level representation than continuous ones. Data can also be simplified and reduced through discretization. For both users and experts, discrete features are easier to understand, explain and use [10]. In general, discretization makes learning more accurate and faster, and obtained models (decision trees, induction rules) using discrete features are usually shorter, more compact and more accurate than using continuous ones, hence the results can be more closely examined, compared, used and reused. In addition to the many advantages of having discrete data over continuous one, a suite of classification learning algorithms can only deal with discrete data.

Discretization methods have been developed along different lines due to different needs: supervised vs. unsupervised, dynamic vs. static, global vs. local, splitting (top-down) vs. merging, and direct vs. incremental. A recent survey can be seen in [6].

In the clinical domain there are few articles about discretization of continuous features in clinical datasets. In [5], the authors consider the role of discretization as part of a broader evaluation of classifiers for a trauma surgery dataset. This study looks at decision trees and naive Bayes classifiers, and evaluates their performance with both equal frequency discretization, and an entropy-based supervised method that included oversight by a domain expert. The supervised method shows marginal but statistically significant improvement over the use of quartiles.

Clarke and Barton [3] developed a discretization algorithm using clinical data from the National Heart, Lung, and Blood Institute which also used an entropy-based method for deriving partitions of certain clinical attributes, including blood pressure and body mass index. In each of these cases, class labels

were known for each observation, and these were used to minimize the loss of information caused by discretization.

Stacey and McGregor [17] develop a high-frequency patient monitoring system, which will enable detection of multiple temporal patterns across multiple high-frequency patient data streams for multiple patients within the domain of Neonatal Intensive Care, through the use of two automated unsupervised temporal discretization methods, equal width discretization and Symbolic Aggregate approximation (SAX).

Temporal abstraction Temporal abstraction (TA) has become a topic of much interest and research in clinical IDA (interactive data analysis) systems where high dimensionality and large volume of time stamped data is the norm. Its goal is to transform patient data from a low level quantitative form to high-level qualitative descriptions, which are closer to the language of clinicians [17].

The context of analysis changes due to adjusting medications and progression of disease states, meaning data values that were considered normal at one time, or in a particular context, may be dangerously abnormal at another.

The process of TA takes either pre-processed or raw data as input and produces context sensitive and qualitative interval based representations. This is required in order for symbolic reasoning strategies to proceed and can be seen as a first point of entry for knowledge into the data analysis system.

When data is oscillating at high frequencies and/or frequencies are irregular, the process of data abstraction becomes increasingly complex and when compounded with processing constraints imposed by data intensive environments such as intensive care, many standalone systems reviewed would prove insufficient.

Some domains encompass low frequency observations, such as blood glucose levels in the domain of diabetes mellitus, while others entail high frequency characteristics, for instance heart rate fibrillation in the domain of neonatal intensive care. Reviewed literature either abstracted data from a database or from online data streams. In any case, the sample frequency of the data is an important factor to the design of the TA mechanisms and has an impact upon the granularity of the time measure, which determines whether TA will be employed to bridge gaps in data or to summarise more dense data [17].

There are several approaches to the TA task; some exploit context-sensitive knowledge acquired from human experts, a method known as knowledge-based temporal abstraction (KBTA) [15]; others are purely automatic, and rely mostly on a discretization of the raw values and concatenation ([1], [7]).

Temporal Abstraction for time series mining in the form of time intervals was already proposed by [7]. Several common discretization methods, such as equal width discretization, which uniformly divides the ranges of each value, and equal frequency discretization, do not consider the temporal order of the values; other methods, such as SAX [9] (which focuses on a statistical discretization of the values) and Persist [12] (which maximizes the duration of the resulting time intervals), explicitly consider the temporal dimension.

2.2 Survival prediction in Intensive Care Burn Unit

Intensive Care Burn Units (ICBU) are specialized units in which the main pathologies treated are inhalation injuries and severe burn injuries. Early mortality prediction after admission is essential before an aggressive or conservative therapy can be recommended. Severity scores are simple but useful tools for physicians when evaluating the state of the patient [16]. Scoring systems aim to use the most predictive pre-morbid and injury factors to yield an expected likelihood of death for a given patient. Baux and prognostic burn index (PBI) scores provide a mortality rate by summing age and percentage of total burn surface area (%TBSA), while the abbreviated burns severity index (ABSI) also considers the gender and presence of inhalation injury.

Nevertheless, the evolution of other parameters during the resuscitation phase (first 2 days) and during the stabilization phase (3 following days) can also be important. The initial evaluation and resuscitation of patients with large burns that require inpatient care can only be loosely guided by formulas and rules. The inherent inaccuracy of formulas requires continuous re-evaluation and adjustment of infusions based on resuscitation targets. Incomings, diuresis, fluid balance, acid base balance (pH, bicarbonate, base excess) and others help to define objectives and to assess the evolution and treatment response.

In the ICBU, patient's evolution is registered but not considered in scores for mortality prediction. In an previous article [2] we made emergent patterns using a knowledge-based temporal abstraction and then we built classifiers of the survival of the patients with a high sensitivity and specificity.

3 Case study

The database has 480 patients registries recorded between 1992 and 2002. From the database, we have removed only the patients who died during the course of study or those in which could not be estimated the hours of hospital stay.

After this cleansing, 465 patients remain, where 81.29% (378/87) of them survived, 69.68% (324/141) are male, and 43.23% (201/264) of them have inhalation injuries. Table 1 depicts a summary of the static attributes of the database.

Attribute	Min	Max	Media	Std. Dev
Age (years)	9	95	46.42	20.34
Weight (kg)	25	120	71.05	10.77
Length of stay (days)	3	162	25.02	24.24
Total burn surface area (%)	1	90	31.28	20.16
Deep burn surface area (%)	0	90	17.01	17.41
SAPS	6	58	20.67	9.49

Table 1. Attribute summary.

Temporal attributes that allow the monitoring and evaluation of the response to treatment of patients are recorded during five days. All attributes are continuous variables and represent the value accumulated during 24 hours. The

registered variables are: a) total of managed liquids measured in cc; (b) diuresis in DC; (c) balance of fluids in DC; (d) pH; (e) bicarbonate in mmol/L; and (f) excess base in mEq/L.

Note that fluid balance is not the difference between revenues and diuresis, but are considered all the possible eliminations of fluids.

In the evaluation performed in the current study, we have used two unsupervised methods (equal frequency, with four quartiles and one modification of this method, in which the equal frequency discretization is made daily), and two supervised methods (a method specific to clinical data with both supervised and unsupervised components -reference range based- and entropy-based discretization).

3.1 Expert discretization (Reference Range)

The reference ranges discretization was made by the expert, and was determined from a variety of sources. Incomings (INC) were originally given in cc, we decided to make them uniform to the weight of the patient and to the %TBSA according to the Parkland's formula for resuscitation (the most used one). We have used quartiles to make four intervals: [$<$, 2.3), [2.3, 3.66), [3.66, 5.78), [5.78, $>$]. The usual unit with meaning for diuresis (DIU) is cc/kg/h, so we have divided all values between weight and 24 (every value is daily register with cumulated 24 hours). Clinical terms usually used in adults for diuresis are oliguria under 0.5 cc/kg/h, normal diuresis within 0.5 and 1 cc/kg/h, and augmented diuresis over 1 cc/kg/h. Values above 1 mean a normal functioning, but we have used the third quartile to differentiate augmented from high values. Intervals defined are [0, 0.5), [0.5, 1), [1, 1.9), [1.9, $>$]. The fluid balance (BAL) measures the difference between the incomings and the total fluid output (not only diuresis). A desired value would be slightly positive balance. In this case, the therapeutic goal is fulfilled every day, and we have used the median of the consecutive days as a way of improvement. We defined the intervals [$<$, -2), [-2, 10.5), [10.5, 20.4), [20.4, 52.22), [52.22, $>$], being 10.5, 20.4 and 52.22 the medians of the forth, third and second day respectively.

For pH, bicarbonate and base excess, there is no standard criterion for qualitative discretization. A possible abstraction is the distinction of pH values in severe acidosis [$<$,7.20), moderate acidosis [7.20,7.30), mild acidosis [7.30,7.35), normal [7.35,7.45), mild alkalosis [7.45,7.50), moderate alkalosis [7.50,7.60), and severe alkalosis [7.6, $>$).

The qualitative abstraction of base excess (BE) is done with respect pH and maintaining pCO₂ in 40 mmHg. Normal values of BE are between -2 and 2 mEq/L. We have used 4 intervals for different level of acidosis and alkalosis: [$<$, -4), [-4,-2), [-2,2), [2,4), [4, $>$].

Normal levels of bicarbonate (BIC) are within [21,25] mmol/L, and we have defined the following intervals using the interquartile difference Q₃-Q₁ to create a reference around normal values: [$<$, 17), [17,21), [21,25), [25,29), [29, $>$].

3.2 Entropy based discretization

We have used the Weka implementation of Fayyad and Irani's supervised entropy-based discretization with MDL stopping criterion.

The intervals generated are represented in Table 2. In the experiments with this discretization we are not going to use Incoming (INC) and Diuresis (DIU) attributes because they only have one interval.

Entropy	INC	DIU	BAL	BIC	pH	BE
First interval	All	All	(-inf, 4.087]	(-inf, 19.55]	(-inf, -7.295]	(-inf, -6.05]
Second interval			(4.087, inf)	(19.55, 25.25]	(-7.295, inf)	(-6.05, 0.35]
Third interval				(25.25, inf)		(0.35, inf)

Table 2. Entropy discretization intervals of every attribute.

3.3 Equal frequency discretization (quartiles)

We have divided all the data recorded in five days of every attribute into four groups which contains approximately the same number of values. The minimum, maximum and the three quartiles of every attribute are in Table 3.

Quartile	INC	DIU	BAL	BIC	pH	BE
Minimum	0.0025	0.0378	-6.5833	10	6.98	-22
First quartile (Q1)	0.1057	1.0786	-0.0694	22	7.36	-2.15
Second quartile (Q2)	0.1766	1.4713	1.0838	24	7.41	0
Third quartile (Q3)	0.2971	2.0238	3.0750	26	7.44	2.1
Maximum	18.6667	56.5067	166.08	42	7.67	11.3

Table 3. Quartiles of every attribute.

3.4 Daily equal frequency discretization (daily quartiles)

We have calculated for every attribute and day their quartiles. This has been possible because the sample frequency of the data is daily and the granularity of time measure is the same, so we only have the amount of data acquire at the end of every day, in 5 days, for every attribute.

As example, we show in Table 4 the discretization of Incomings (INC), where we can observe that one of the objectives of the resuscitation phase is to restore the fluids, and so it is higher in the first days.

4 Experiments and discussion

We have followed the 4-step Knowledge discovery process described in our previous article [2] to do the experiments. In "Step 0: Discretization of temporal attributes" we have used the four different discretization processes described above.

Quartile (INC)	Day 1	Day 2	Day 3	Day 4	Day 5
Minimum	0.0417	0.0366	0.0025	0.0209	0.0207
First quartile (Q1)	0.2218	0.1279	0.0992	0.0865	0.0818
Second quartile (Q2)	0.3112	0.1959	0.1518	0.1331	0.1222
Third quartile (Q3)	0.4848	0.2998	0.232	0.2151	0.1975
Maximum	18.6667	9.369	4.2345	4.3769	2.3922

Table 4. Quartiles of Incoming attribute.

Since we want to extract rules on both survivors and non survivors, we extract patterns from the subset of survivors and from the subset of non-survivors in order to remove common behaviour or patient’s evolution that is not discriminative. In this manner we get the contrast patterns that we will use in the classification step to generate interpretable models.

To do the last step "Classification algorithms with interpretable models", we are going to use a decision tree and a rule learning algorithm.

On the one hand we chose a J48 decision tree, that is the Weka implementation of the C4.5 top-down decision tree learner proposed by Quinlan [14]. The algorithm uses the greedy technique and it is a variant of ID3, which determines at each step the most predictive attribute, and splits a node based on this attribute.

On the other hand, we chose a sequential covering algorithm, such a Repeated Incremental Pruning to Produce Error Reduction, RIPPER [4]. In this case, rules are learned one at a time, and each time a rule is learned, the tuples covered by the rule are removed. This process is repeated until there are no more training examples or if the quality of a rule obtained is below a user-specified threshold.

We have chosen two different supports (10% and 8%) to generate the patterns, using expert, quartiles and daily quartiles discretization. With a support of 10% we obtain a small number of patterns (hundreds), and with a support of 8% the number of patters increases considerably (thousands).

With entropy discretization we have had problems choosing the support. As we are going to use less attributes, the number of patterns increase exponentially. With a support of 10% we obtain many contrast patterns (158296 patterns), so we are going to increase the support to reduce the number of patterns.

In all cases, we configured the classifiers with the same minimum number of elements in each leaf or rule to 2% of the instances. The accuracy, sensitivity, specificity and AUC are calculated with a 10-fold cross validation.

In Tables 5 and 6 we can see the results of the experiments with RIPPER and J48 algorithms. In general the best result is obtained with entropy discretization. After that, with daily quartiles we obtain good results, even with a small number of patterns. There is little difference in the results if we use a decision tree or a rule learning algorithm.

In this discussion, we explore two aspects: classification performance and the number of patterns used.

Regarding performance, the worst results are produced by the equal frequency discretization (quartiles) using a J48 decision tree algorithm. Expert dis-

cretization produces good results, although it is necessary to have a big number of patterns.

Discretization	Support	Patterns	Sensitivity	Specificity	Accuracy	AUC
Expert	10%	391	100.00%	40.23%	88.62%	0.704
	8%	4931	100.00%	58.62%	92.26%	0.777
Entropy	12%	18752	100.00%	40.23%	88.82%	0.711
	10%	158296	100.00%	72.41%	94.84%	0.859
Quartiles	10%	430	100.00%	47.13%	90.11%	0.733
	8%	8978	100.00%	55.17%	91.61%	0.767
Daily quartiles	10%	462	100.00%	59.77%	92.47%	0.789
	8%	5298	100.00%	57.47%	92.04%	0.776

Table 5. Results of the experiments with RIPPER rule learning algorithm.

Discretization	Support	Patterns	Sensitivity	Specificity	Accuracy	AUC
Expert	10%	391	100.00%	43.68%	89.46%	0.709
	8%	4931	100.00%	56.32%	91.83%	0.782
Entropy	12%	18752	100.00%	51.72%	90.97%	0.749
	10%	158296	100.00%	71.26%	94.62%	0.859
Quartiles	10%	430	100.00%	48.28%	90.32%	0.728
	8%	8978	100.00%	49.43%	90.54%	0.749
Daily quartiles	10%	462	100.00%	60.92%	92.69%	0.789
	8%	5298	100.00%	56.32%	91.83%	0.801

Table 6. Results of the experiments with J48 decision tree algorithm.

The expert discretization introduces knowledge, but also some bias. This can produce worst classification results, but greatly facilitates the interpretation. The automatic discretization (quartiles and entropy) produces arbitrary cut-offs that usually do not correspond with clinical knowledge and complicates understanding and interpretation. At least, the cut-offs on the quartiles have a mathematical interpretation meaningful to the clinicians, which information based discretization does not have.

Entropy discretization produces the best results in our experiments. Formally, it is characterized by finding the split with the maximal information gain. A negative aspect of this discretization is that the selected cut points do not have any meaning for the clinician. Besides, as shown in Table 2, two of the variables, INC and DIU, have been discarded since the discretization produced one single interval. Nevertheless, the third variable, BAL, associated to liquids infusion remains.

Regarding the number of patterns obtained, all experiments, except the one using entropy, have generated few patterns with a 10% support. And we have even obtained a good result with this small number of patterns using daily quartiles discretization. With 8% support, global quartiles generates almost twice

patterns than with daily quartiles or expert discretizations. When we use entropy, the number of generated patterns had to be very large in order to obtain acceptable results. This is due to the fact that we use two attributes less.

All discretizations that we have made are not affected by outliers. It had been different if, for example, we would have chosen a equal width discretization.

5 Conclusions and future work

In our study for the prediction of mortality in an Intensive Care Burn Unit, we have 6 non periodic sort time-series of physiological and laboratory data. We opted for the generation of multivariate sequential patterns from the time series data that represent patients' evolution, and then we used them as predictor for building classifiers. In order to allow the interpretability of the results, discretization is fundamental. To perform the discretization we have used interval-based abstract concepts generated through the use of either knowledge-based temporal abstraction or automated temporal discretization using global equal-frequency discretization and information based techniques. In addition, we have also compared the effect of the equal-frequency discretization for every day.

We have examined the impact of the discretization techniques on the accuracy of two different classification algorithms and the number of patterns needed for classification. Since we needed a model easy to interpret by the physician, we have chosen a decision tree and a rule learning algorithm as classification algorithms. To the best of our knowledge this is the first work that measures the effect of discretization in sequential patterns for classification.

The information-based discretization obtained the best performance classification. It was expected, because it finds the best split so that the bins are as pure as possible with regards to the output. But, in our application, we have needed a high number of patterns, and the selected cut points not have any sense for the clinician and even the method itself is not easy to understand. A non desirable side effect is that two of the variables were removed because their discretization only produced one interval.

Global equal frequency discretization (Quartiles) gets the worst classification results when a low support is used. The equal-frequency discretization made daily produces very good results even with a small number of patterns. Besides, it allows the use of the same labels for the patterns (e.g. normal, high, etc.), regardless of the day. This method and its results are easily interpretable.

The expert discretization (reference range) reaches a comparable classification performance. It has the advantage the it generate less patterns for the same support and would allow.

For further research, we will include another step for variable selection in order to reduce the number of patterns, and to evaluate the effect of daily entropy and daily expert discretization.

Acknowledgement This work was partially funded by the Spanish Ministry of Economy and Competitiveness under project TIN2013-45491-R, and by Eu-

ropean Fund for Regional Development (EFRD), and Instituto de Salud Carlos III (Ref: FIS PI 12/2898).

References

1. R. Azulay et al. Discretization of medical time series - A comparative study. In *Procs of the IDAMAP 2007*, Amsterdam, The Netherlands, 2007.
2. I.J. Casanova, M. Campos, J.M. Juaraz, A. Fernandez-Fernandez-Arroyo and J.A. Lorente. Using multivariate sequential patterns to improve survival prediction in Intensive Care Burn Unit. In *Procs of the 15th Conf. on Artificial Intelligence in Medicine*, AIMIE 2015, pages 277-286, Pavia, Italy, 2015.
3. E.J. Clarke, B.A. Barton. Entropy and MDL discretization of continuous variables for Bayesian belief networks. *International Journal of Intelligent Systems*, Vol. 15, Pages 61-92, 2000
4. W. W. Cohen. Fast effective rule induction. In *Procs of the 20th Int. Conf on Machine Learning*, pages 115-123. Morgan Kaufmann, 1995.
5. J. Densar, B. Zupan, N. Aoki, et al. Feature mining and predictive model construction from severe trauma patient's data. *International Journal of Medical Informatics*, Vol. 63, Pages 41-50, 2012.
6. S. Garcia, J. Luengo, J.A. Saez, V. Lopez and F. Herrera A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, Volume 25(4), pages 734-750, 2013.
7. F. Hppner. Time series abstraction methods - A Survey In *Workshop on Knowledge Discovery in Databases*, Dortmund, 2002
8. M.D.C. Lima et al. Heuristic discretization method for bayesian networks *Journal of Computer Science*, Volume 10(5), pages 869-878, 2014
9. J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A Symbolic Representation of Time Series with Implications for Streaming Algorithms. In *Procs of the 8th ACM SIGMOD DMKD workshop*, 2003.
10. H. Liu, F. Hussain, C.L. Tan, and M. Dash. Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery*, Volume 6(4), pages 393-423, 2002
11. D.M. Maslove, T. Podchiyska and H.J. Lowe Discretization of continuous features in clinical datasets. *Journal of the American Medical Informatics Association*, Volume 20(3), pages 544-553, 2013
12. F. Mrchen, and A. Ultsch. Optimizing Time Series Discretization for Knowledge Discovery. In *Procs of the KDD05*, 2005.
13. R. Moskovitch and Y. Shahar Classification-driven temporal discretization of multivariate time series *Data Mining and Knowledge Discovery*, Volume 29(4), pages 871-913, Springer Link, 2015.
14. J.R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81-106, 1986.
15. Y. Shahar. A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90(1-2), 1997.
16. N.N. Sheppard, S. Hemington-Gorse, O.P. Shelley, B. Philp, and P. Dzielwski. Prognostic scoring systems in burns: A review. *Burns*, 37(8):1288 - 1295, 2011.
17. M. Stacey and C. McGregor. Temporal abstraction in intelligent clinical data analysis: A survey *Artificial Intelligence in Medicine*, Vol. 39, Pages 1-24, 2007