

# A Novel Contrast Pattern Selection Method for Class Imbalance Problems

Octavio Loyola-González<sup>1,2(✉)</sup>, José Fco. Martínez-Trinidad<sup>1</sup>,  
Jesús Ariel Carrasco-Ochoa<sup>1</sup>, and Milton García-Borroto<sup>3</sup>

<sup>1</sup> Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro No. 1,  
Sta. María Tonanzintla, 72840 Puebla, Mexico  
{octavioloyola, fmartine, ariel}@inaoep.mx

<sup>2</sup> Centro de Bioplasmas, Universidad de Ciego de Ávila.,  
Carretera a Morón Km 9, 69450 Ciego de Ávila, Cuba  
octavioloyola@bioplasmas.cu

<sup>3</sup> Instituto Superior Politécnico José Antonio Echeverría.,  
Calle 114 No. 11901, 19390 Marianao, La Habana, Cuba  
mgarciab@ceis.cujae.edu.cu

**Abstract.** Selecting contrast patterns is an important task for pattern-based classifiers, especially in class imbalance problems. The main reason is that the contrast pattern miners commonly extract several patterns with high support for the majority class and only a few patterns, with low support, for the minority class. This produces a bias of classification results toward the majority class, obtaining a low accuracy for the minority class. In this paper, we introduce a contrast pattern selection method for class imbalance problems. Our proposal selects all the contrast patterns for the minority class and a certain percent of contrast patterns for the majority class. Our experiments performed over several imbalanced databases show that our proposal selects significantly better contrast patterns, obtaining better AUC results, than other approaches reported in the literature.

**Keywords:** Supervised classification · Pattern selection · Contrast patterns · Imbalanced databases

## 1 Introduction

Several classifiers have been proposed for supervised classification, among them, an important family is the contrast pattern-based classifiers. A *pattern* is an expression defined in a language that describes a set of objects. For example, a pattern that describes a set of plants can be the following:  $[Petal.Width \in [0.60, 1.60]] \wedge [Roots \leq 10] \wedge [Stem = "Thick"]$ . A pattern that appears significantly more in a class than in the remaining classes is named as *contrast pattern*. Finally, a classifier which predicts the class of a query object based on a set of contrast patterns is called: *contrast pattern-based classifier*. It is important to highlight that the pattern-based classifiers, as well as their results, can

be understood by the experts in the application domain through the patterns associated to each class. Also, contrast pattern-based classifiers have reported significantly better classification results than other popular classification models, like naive bayes, nearest neighbor, bagging, boosting, and SVM [7, 11].

In some real-world applications, there are problems where the objects are not equally distributed into the classes, like online banking fraud detection, liver and pancreas disorders, forecasting of ozone levels, prediction of protein sequences, and face recognition. In these applications, there exist significantly fewer objects belonging to a class (commonly labeled as *minority* class) regarding the remaining classes. This problem is known as class imbalance problem [18–20].

Some classifiers, which show good classification results in problems with balanced classes do not necessarily achieve good performance in class imbalance problems. The main reason is that they produce a bias toward the *majority* class (the class with more objects). Accordingly, the accuracy of these classifiers for the minority class could be close to zero [19, 20].

On class imbalance problems, some pattern-based classifiers, like CAEP [8], do not achieve good classification results because of contrast patterns from the minority class are fewer and they have low support regarding those contrast patterns from the majority class. Then, some classification strategies, which are based only on the support of the contrast patterns, tend to be biased toward the majority class [2, 18, 19].

A proposal for supervised classification based on contrast patterns in class imbalance problems is selecting just a subset of good contrast patterns. The idea is to select, for each class, a collection of high-quality patterns. Consequently, at the classification stage, those contrast patterns with low support for the minority class do not become overwhelmed by those contrast patterns with high support for the majority class, which are much more [4, 9, 19, 24, 25].

In the literature there are three main approaches for contrast pattern selection: (i) selecting only the best contrast pattern, (ii) selecting the  $k$  best contrast patterns, and (iii) selecting all contrast patterns covering the training dataset [4, 9, 18, 24, 25]. In this paper, we propose a novel contrast pattern selection by class for class imbalance problems; the idea consists in selecting all the contrast patterns for the minority class and only a certain percent of contrast patterns for the majority class. Our proposal allows obtaining better accuracy results, when the selected contrast patterns are used by a contrast pattern-based classifier, than other contrast pattern selection approaches of the state-of-the-art.

The rest of the paper has the following structure. Section 2 contains a brief description of the main contrast pattern selection approaches reported in the state-of-the-art. Section 3 introduces our proposal for selecting contrast patterns in class imbalance problems. Section 4 provides the experimental setup. Section 5 presents the experimental results as well as a discussion of them. Finally, Sect. 6 provides our conclusions and future work.

## 2 Related Work

In pattern-based classification, an important task is to select a collection of high-quality patterns for obtaining good classification results [18, 27]. Additionally, the fewer patterns, the faster the classification stage and easier to understand the results by experts in the application domain.

Three main approaches have been proposed in the literature for selecting contrast patterns [4, 17, 18, 22, 24, 25]. These approaches use a quality measure for contrast patterns with the aim of creating a ranking of contrast patterns. A quality measure is a function  $q(P, C, \bar{C}) \rightarrow R$ , which assigns a higher value to a pattern  $P$  when it better discriminates the objects in a class  $C$  from the objects in the remaining problem classes  $\bar{C}$  [17, 18]. Usually, the measure for ranking the contrast patterns depends on the contrast pattern-based classifier to be used (e.g., confidence for association rules,  $X^2$  for decision trees, growth rate for emerging patterns). The three main approaches for selecting contrast patterns are:

**Best contrast pattern (Best CP):** Select the best contrast pattern, according to the ranking, covering the query object. This approach is used by several rule-based classifiers, like CBA [16], for classifying query objects. A drawback of this approach, in class imbalance problems, is that commonly the best contrast pattern according to the ranking is from the majority class. Consequently, the accuracy of the classifier for the minority class is bad.

**Best  $k$  contrast patterns (Best  $k$ ):** Select the best  $k$  contrast patterns from the ranking which cover the query object. Usually, this approach is used by rule-based classifiers like CPAR [26] and some emerging pattern selection methods, as the one proposed in [17]. Recently in [18] the authors proposed to use a fixed percent of patterns instead of a fixed number of patterns. A disadvantage of this approach in class imbalance problems is that the patterns for the minority class are too few and they have low support, then selecting a few patterns of the minority class could degrade the accuracy of the classifier for the minority class.

**Covering the training dataset (Covering):** Select the best contrast patterns, according to the ranking, covering all the objects of the training dataset. This approach is used by some rule based-classifier, like ACN [14] and CMAR [15], and some emerging pattern selection methods [11, 17, 18], showing good accuracy results.

The second and third approaches were studied in [18] for selecting contract patterns in class imbalance problems. The authors tested several quality measures for contrast patterns and they concluded that the best quality measure for ranking contrast patterns in class imbalance problems is *Jaccard* [23]. Based on this conclusion, we will propose a novel contrast pattern selection method, which uses the quality measure *Jaccard* for ranking the contrast patterns.

### 3 New Contrast Pattern Selection Method

In this section, we introduce a contrast pattern selection method for class imbalance problems.

Usually, in class imbalance problems, contrast pattern mining algorithms extract several patterns with high support for the majority class and only a few patterns, with low support, for the minority class [2, 18, 19]. This produces that some contrast pattern-based classifiers, like CAEP [8], become biased toward the majority class. Some strategies find a solution by selecting a collection of high-quality patterns, but their main drawback is that some patterns of the minority class, which could help at the classification stage, are discarded [15, 16, 26]. For solving this problem, we propose to select the patterns by class; selecting all the contrast patterns of the minority class and only a few contrast patterns of the majority class. The main idea is not to discard useful patterns of the minority class and avoiding the selection of many patterns of the majority class, which could overwhelm the patterns of the minority class at the classification stage.

Our pattern selection method can be described by the following steps:

1. Select all the contrast patterns of the minority class to avoid reducing the number of patterns of this class.
2. Rank the contrast patterns of the majority class by using a quality measure for contrast patterns.
3. Select the best  $k$  contrast patterns of the majority class. The  $k$  value is a percent of the total number of patterns, which is provided by the user.

Commonly, the number of patterns extracted from a database depends on different factors such as the nature of training dataset, the contrast pattern mining algorithm, the a-priori global discretization, among others. Hence, instead of selecting a fixed number of contrast patterns, in our contrast pattern selection method, we propose to select just a percent of patterns. The main reason is that the number of patterns to select could be too high, regarding the amount of mined patterns, which means that the selected patterns could be almost all; or too small, which would lead to reduce more than necessary the number of patterns.

Finally, it is important to highlight that in step 2, we suggest to use the quality measure *Jaccard* [23] because this measure has shown good results for ranking contrast patterns in class imbalance problems [18].

### 4 Experimental Setup

In order to evaluate the performance of the proposed selection method, we will perform a comparison of our proposal against the three main approaches for selecting contrast patterns reported in the literature. To do this, first, we will extract the patterns by using a contrast pattern miner. After that, we will create a pattern ranking by applying the quality measure *Jaccard* [23] over the collection of patterns previously extracted. Next, we will select the patterns by using

the three main pattern selection approaches shown in Sect. 2 and our proposal. Finally, each subset of patterns will be used to build a contrast pattern-based classifier. By doing this, we can detect which selection method attains better classification results. As the contrast pattern miner and the classifier are the same and the only difference are the contrast patterns selected by means of the selection method, then a good or bad performance in the classification results can be attributed to the selection method employed.

Table 1 shows the 95 databases used in our experiments, which were taken from the KEEL dataset repository<sup>1</sup> [1]. For avoiding problems due to data distribution in class imbalance problems, for each database, we performed a distribution optimally balanced stratified five cross-validation, as suggested in [20].

For assessing the performance of our classification results, we used the AUC measure [13] because it is the most used measure for class imbalance problems [18–20]. All our AUC results were averaged over the 5-fold cross validation.

As contrast pattern-based classifier, we selected PBC4cip [19], since, it has reported better AUC results than other state-of-the-art classifiers for class imbalance problems [19].

For mining contrast patterns, we selected the Random Forest miner (RFm) [10] using the Hellinger distance [3] as node splitting measure, as suggested in [19]. The main reason is that RFm has been used jointly with the PBC4cip classifier, obtaining higher accuracies than other state-of-the-art contrast pattern mining algorithms [19].

For selecting contrast patterns, we used the three main approaches reported in the literature (see Sect. 2). For selecting the best  $k$  contrast patterns by class, we used the values 10%, 50%, and 80% which have been used in previous studies for class imbalance problems [18]. For selecting contrast patterns using our proposal we used the following  $k$  values: 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, and 80%. By using these values, we can investigate if our proposal is able to attain statistically similar AUC results at using fewer patterns than the number of patterns used in [18].

We also used the Shaffer and Finner post-hoc procedures, and the Friedman test to compare all the classification results, as suggested in [5, 6]. Post-hoc results will be shown by using CD (*critical distance*) diagrams [5]. In a CD diagram, the rightmost classifier is the best classifier. The position of the classifier within the segment represents its rank value, and if two or more classifiers share a thick line it means that they have statistically similar behavior.

## 5 Experimental Results

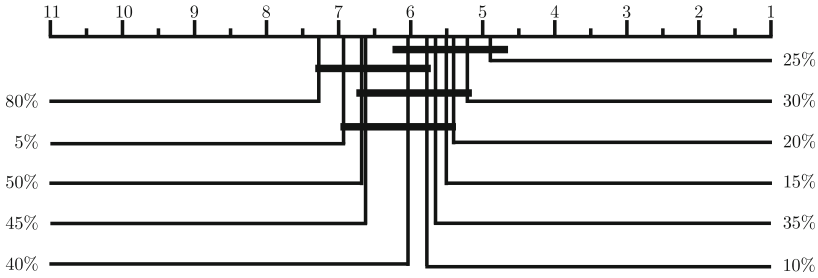
This section is devoted to analyzing and discussing about the classification results achieved by the contrast pattern selection methods described in Sect. 2, using all the imbalanced databases shown in Table 1.

---

<sup>1</sup> <http://www.keel.es/datasets.php>.

**Table 1.** Summary of the imbalanced databases used in our study. Containing the name in the KEEL dataset repository (name), the number of objects (#Objects) and features (#Feat.), and the IR [21].

Name	#Objects	#Feat.	IR	Name	#Objects	#Feat.	IR
glass1	214	9	1.82	ecoli0146vs5	280	6	13.00
ecoli0vs1	220	7	1.86	shuttlec0vsc4	1829	9	13.87
wisconsin	683	9	1.86	yeast1vs7	459	7	14.30
pima	768	8	1.87	glass4	214	9	15.46
iris0	150	4	2.00	ecoli4	336	7	15.80
glass0	214	9	2.06	pageblocks13vs4	472	10	15.86
yeast1	1484	8	2.46	abalone9vs18	731	8	16.40
haberman	306	3	2.78	dermatology6	358	34	16.90
vehicle2	846	18	2.88	zoo3	101	16	19.20
vehicle1	846	18	2.90	glass016vs5	184	9	19.44
vehicle3	846	18	2.99	shuttlec2vsc4	129	9	20.50
glass0123vs456	214	9	3.20	shuttle6vs23	230	9	22.00
vehicle0	846	18	3.25	yeast1458vs7	693	8	22.10
ecoli1	336	7	3.36	glass5	214	9	22.78
newthyroid1	215	5	5.14	yeast2vs8	482	8	23.10
newthyroid2	215	5	5.14	lymphography	148	18	23.67
ecoli2	336	7	5.46	flareF	1066	11	23.79
segment0	2308	19	6.02	cargood	1728	6	24.04
glass6	214	9	6.38	carvgood	1728	6	25.58
yeast3	1484	8	8.10	krvskzeroonevsdraw	2901	6	26.63
ecoli3	336	7	8.60	krvskonevsfifteen	2244	6	27.77
pageblocks0	5472	10	8.79	yeast4	1484	8	28.10
ecoli034vs5	200	7	9.00	winequalityred4	1599	11	29.17
yeast2vs4	514	8	9.08	poker9vs7	244	10	29.50
ecoli067vs35	222	7	9.09	yeast1289vs7	947	8	30.57
ecoli0234vs5	202	7	9.10	abalone3vs11	502	8	32.47
glass015vs2	172	9	9.12	winequalitywhite9vs4	168	11	32.60
yeast0359vs78	506	8	9.12	yeast5	1484	8	32.73
yeast0256vs3789	1004	8	9.14	krvskthreevseleven	2935	6	35.23
yeast02579vs368	1004	8	9.14	winequalityred8vs6	656	11	35.44
ecoli046vs5	203	6	9.15	ecoli0137vs26	281	7	39.14
ecoli01vs235	244	7	9.17	abalone17vs78910	2338	8	39.31
ecoli0267vs35	224	7	9.18	abalone21vs8	581	8	40.50
glass04vs5	92	9	9.22	yeast6	1484	8	41.40
ecoli0346vs5	205	7	9.25	winequalitywhite3vs7	900	11	44.00
ecoli0347vs56	257	7	9.28	winequalityred8vs67	855	11	46.50
yeast05679vs4	528	8	9.35	abalone19vs10111213	1622	8	49.69
vowel0	988	13	9.98	krvskzerovseight	1460	6	53.07
ecoli067vs5	220	6	10.00	winequalitywhite39vs5	1482	11	58.28
glass016vs2	192	9	10.29	poker89vs6	1485	10	58.40
ecoli0147vs2356	336	7	10.59	shuttle2vs5	3316	9	66.67
led7digit02456789vs1	443	7	10.97	winequalityred3vs5	691	11	68.10
ecoli01vs5	240	6	11.00	abalone20vs8910	1916	8	72.69
glass06vs5	108	9	11.00	krvskzerovsfifteen	2193	6	80.22
glass0146vs2	205	9	11.06	poker89vs5	2075	10	82.00
glass2	214	9	11.59	poker8vs6	1477	10	85.88
ecoli0147vs56	332	6	12.28	abalone19	4174	8	129.44
cleveland0vs4	177	13	12.62				



**Fig. 1.** CD diagram with a statistical comparison (using  $\alpha = 0.05$ ) of the AUC results of our proposal for selecting contrast patterns, using different  $k$  values over all the tested databases.

In order to simplify the presentation, a supplementary material website<sup>2</sup> has been created for this paper, which contains several tables from experimental results as well as detailed tables from the statistical test results.

Figure 1 shows a CD diagram with a statistical comparison of the AUC results obtained by our proposal using different  $k$  values and considering all the imbalanced databases shown in Table 1. From this figure, we can conclude that our proposal using  $k = 25\%$  obtained the best position into the Friedman's ranking. However, the difference of the AUC results of our proposal using  $k = 25\%$  against using  $k$  as: 10%, 15%, 20%, 30%, 35%, and 40% is not statistically significant. Therefore, we selected  $k = 10\%$  since it allows selecting the fewest number of patterns.

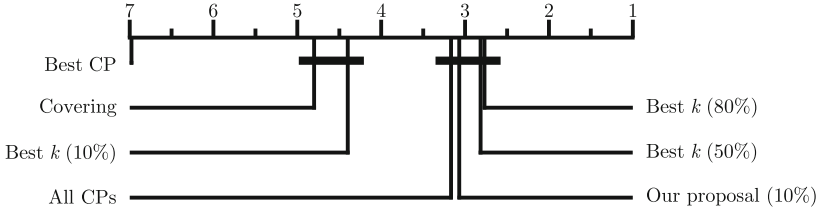
Figure 2 shows a CD diagram with a statistical comparison of the AUC results obtained by our proposal, using  $k = 10\%$ , against the other contrast pattern selection methods reviewed in the Sect. 2, as well as by using all the contrast patterns (All CPs). From this figure, we can see that the AUC results of our proposal against those AUC results archived by Best  $k = 50\%$ , Best  $k = 80\%$ , and All CPs are not statistically significant. However, our proposal obtains a better position into the Friedman's ranking than using all the contrast patterns (All CPs). Also, our proposal uses fewer contrast patterns for classification than the other approaches having statistically similar behavior. On the other hand, notice that the contrast pattern selection method Best CP statistically obtained the worst results. This is because, in class imbalance problems, commonly the best contrast pattern according to the ranking comes from the majority class and consequently the accuracy for the minority class is greatly affected.

## 5.1 Regarding Different Class Imbalance Levels

For studying the effect of the class imbalance level on the contrast pattern selection methods previously analyzed, we divided the databases into equal-frequency groups depending on the IR of each one. For doing this, we used the *Discretize*<sup>3</sup>

<sup>2</sup> <https://sites.google.com/site/octavioloyola/papers/PSM4MajClass>.

<sup>3</sup> Path in WEKA: `weka.filters.unsupervised.attribute.Discretize`.



**Fig. 2.** CD diagram with a statistical comparison (using  $\alpha = 0.05$ ) of the AUC results of our proposal and the other contrast pattern selection methods reported in the literature.

method, taken from the WEKA Data Mining Tool [12], to create six equal-frequency groups depending on the IR of the databases. These groups are shown in Table 1 using horizontal thin lines.

**Table 2.** Results of the best contrast pattern selection for each bin

Name	Bin interval	#Databases	Best selection method
Bin1	(1.820, 5.300]	16	Best $k = 80\%$
Bin2	(5.300, 9.175]	16	Best $k = 50\%$
Bin3	(9.175, 12.810]	16	Our proposal ( $k = 10\%$ )
Bin4	(12.810, 23.730]	16	Best $k = 80\%$
Bin5	(23.730, 39.905]	16	Best $k = 50\%$
Bin6	(39.905, 129.440]	15	Best $k = 80\%$

Table 2 shows the best contrast pattern selection method for each bin. From this table, we can conclude that for the less imbalanced databases (Bin1), Bin4 and for the Bin6, the best contrast pattern selection method is Best  $k = 80\%$ . For Bin3, the best selection method is our proposed method using  $k = 10\%$ . Finally, for Bin2 and Bin5 the best contrast pattern selection is Best  $k = 50\%$ . These results help us to select the best contrast pattern selection method depending on the class imbalance level of the database.

## 6 Conclusions and Future Work

Selecting a collection of high-quality patterns is an important task for pattern-based classification. The main aim is to achieve good classification results using as few patterns as possible in order to obtain a model easier to understand by experts in the application domain. Following this idea, the main contribution of this paper is a new contrast pattern selection method for contrast pattern-based classification in class imbalance problems. Our proposal selects all the patterns



of the minority class and based on a ranking computed through the *Jaccard* measure, it selects a percent of the best patterns of the majority class.

From our experiments using several imbalanced databases, we can conclude that our proposal performs significantly better, when it uses the 25% of the contrast patterns of the majority class, regarding other tested percents. Also, our proposal using  $k = 10\%$  outperforms significantly other contrast pattern selection methods reported in the state-of-the-art, like Best CP, Covering, and Best  $k = 10\%$ . Moreover, our proposal using  $k = 10\%$  have not statistical differences with other contrast pattern selection methods, like Best  $k = 50\%$ , Best  $k = 80\%$ , and All CPs, but these methods need more patterns.

On the other hand, based on our experiments regarding the class imbalance ratio of the databases, we suggest that: if the database has an IR smaller than or equal to 5.3, or its IR ranges in (12.810, 23.730], or its IR ranges in (39.905, 129.440], then Best  $k = 80\%$  is the best contrast pattern selection method. If the database has an IR in (9.175, 12.810] then our proposal using  $k = 10\%$  is recommended. And finally, if the database has an IR in (5.300, 9.175] or its IR ranges in (23.730, 39.905] then we suggest using the contrast pattern selection method Best  $k = 50\%$ .

Finally, as future work, we will explore the use of maximal or closed contrast patterns as an alternative for selecting a reduced subset of contrast patterns for classification in class imbalance problems.

**Acknowledgment.** This work was partly supported by National Council of Science and Technology of Mexico under the scholarship grant 370272.

## References

1. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S.: KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *J. Multiple-Valued Logic Soft Comput.* **17**(2–3), 255–287 (2011)
2. Alhammady, H.: A novel approach for mining emerging patterns in rare-class datasets. In: Sobh, T. (ed.) *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering*, pp. 207–211. Springer, Dordrecht (2007)
3. Cieslak, D., Hoens, T., Chawla, N., Kegelmeyer, W.: Hellinger distance decision trees are robust and skew-insensitive. *Data Min. Knowl. Disc.* **24**(1), 136–158 (2012)
4. Coenen, F., Leng, P.: An evaluation of approaches to classification rule selection. In: *Fourth IEEE International Conference on Data Mining*, pp. 359–362 (2004)
5. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006)
6. Derrac, J., García, S., Molina, D., Herrera, F.: A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol. Comp.* **1**(1), 3–18 (2011)
7. Dong, G., Bailey, J.: *Contrast Data Mining: Concepts, Algorithms, and Applications*. Chapman and Hall/CRC, 1st edn. (2012)

8. Dong, G., Zhang, X., Wong, L., Li, J.: CAEP: classification by aggregating emerging patterns. In: Arikawa, S., Furukawa, K. (eds.) DS 1999. LNCS, vol. 1721, pp. 30–42. Springer, Heidelberg (1999). doi:[10.1007/3-540-46846-3\\_4](https://doi.org/10.1007/3-540-46846-3_4)
9. Fürnkranz, J., Flach, P.: An analysis of stopping and filtering criteria for rule learning. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS, vol. 3201, pp. 123–133. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-30115-8\\_14](https://doi.org/10.1007/978-3-540-30115-8_14)
10. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: Finding the best diversity generation procedures for mining contrast patterns. *Expert Syst. Appl.* **42**(11), 4859–4866 (2015)
11. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Medina-Pérez, M.A., Ruiz-Shulcloper, J.: LCMine: an efficient algorithm for mining discriminative regularities and its application in supervised classification. *Pattern Recogn.* **43**(9), 3025–3034 (2010)
12. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *SIGKDD Expl.* **11**(1), 10–18 (2009)
13. Huang, J., Ling, C.X.: Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **17**(3), 299–310 (2005)
14. Kundu, G., Islam, M., Munir, S., Bari, M.: ACN: an associative classifier with negative rules. In: Proceedings of the 11th IEEE International Conference on Computational Science and Engineering, pp. 369–375. IEEE Xplore Press (2008)
15. Li, W., Han, J., Pei, J.: CMAR: accurate and efficient classification based on multiple class-association rules. In: Proceedings of the International Conference on Data Mining, ICDM 2001, pp. 369–376. IEEE (2001)
16. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, KDD 1998, pp. 80–86. AAAI (1998)
17. Loyola-González, O., García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: An empirical comparison among quality measures for pattern based classifiers. *Intell. Data Anal.* **18**, S5–S17 (2014)
18. Loyola-González, O., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., García-Borroto, M.: Effect of class imbalance on quality measures for contrast patterns: an experimental study. *Inform. Sci.* **374**, 179–192 (2016)
19. Loyola-González, O., Medina-Pérez, M.A., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Monroy, R., García-Borroto, M.: PBC4cip: a new contrast pattern-based classifier for class imbalance problems. *Knowl.-Based Syst.* **115**, 100–109 (2016)
20. Moreno-Torres, J.G., Saez, J.A., Herrera, F.: Study on the impact of partition-induced dataset shift on k-fold cross-validation. *IEEE Trans. Neural Networks Learn. Syst.* **23**(8), 1304–1312 (2012)
21. Orriols-Puig, A., Bernadó-Mansilla, E.: Evolutionary rule-based systems for imbalanced data sets. *Soft Comput.* **13**(3), 213–225 (2009)
22. Refai, M.H., Yusof, Y.: Partial rule match for filtering rules in associative classification. *J. Comput. Sci.* **10**(4), 570 (2014)
23. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. *Inf. Syst.* **29**(4), 293–313 (2004)
24. Wang, Y.J., Xin, Q., Coenen, F.: A novel rule weighting approach in classification association rule mining. In: Seventh IEEE International Conference on Data Mining Workshops, pp. 271–276 (2007)

25. Ye, Y., Li, T., Jiang, Q., Wang, Y.: CIMDS: adapting postprocessing techniques of associative classification for malware detection. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **40**(3), 298–307 (2010)
26. Yin, X., Han, J.: CPAR: classification based on predictive association rules. In: *Proceedings of the Third SIAM International Conference on Data Mining, SDM 2003*, pp. 331–335. SIAM (2003)
27. Zhang, X., Dong, G.: Overview and Analysis of Contrast Pattern Based Classification. In: Dong, G., Bailey, J. (eds.) *Contrast Data Mining: Concepts, Algorithms, and Applications*, Chap. 11. *Data Mining and Knowledge Discovery Series*, pp. 151–170. Chapman and Hall/CRC (2012)