pair counting similarity measures

Methods of Clustering to get multiple outputs:

1) Data Subsets
2) multiple Algorithms
3) Data Feature Subsets

} Solutions via

1) Co-association methods
2) graph partitioning based methods

1) Co-association matrix generated for each data subset
   ⇒ all matrices summed to "consensus" matrix
3) Generates graph representation based on relationship between samples/clusters/partitions ⟹ what does this mean?
   ⇒ searches for consensus solution
   ⇒ evaluated by

can these apply to our clustering outputs? ⟸

1) Pair counting measures:
   - Adjusted Rand Index, Fowlkes-Mallows Index, Jaccard Index
2) Information Theoretic measures
   - mi, nmi, vi, purity
3) set matching metrics
   - Van Dongen criterion, H criterion, $\mathcal{L}$ criterion

* To evaluate if final consensus is better than individuals:
  ⇒ ANMI, PPMI, ARI, (two ARI measures later in paper)

NOTE: useful chart of pair-counting similarity measures w/ def

Pairwise similarity measures are evaluated based on the following properties

1) Symmetry: measures should output same sim. value for 2 cluster outputs no matter the ordering.
2) Possibility of Dist. Computation: can you split into sub-matrices to calc.
3) Detection of un-correlatedness: how well do they do w/ random clustering
4) measure of complementarity: generalization of homogeneity + completeness
5) measure of self similarity: can they compute self similarity

TABLE 3. Comparison of generalized measures based on different properties.

Measures	Range	Base	P1	P2	P3a	P3b	P4a1	P4a2	P4b	P4c	P5a	P5b
Desirable			Y	Y	Y	Y	Min	Min	Base	Min	Max	Max
ARI	[-1, 1]	0	Y	Y	Y	Y	N (0)	N (0)	Y (0)	N	NA	NA
B	[0, 1]	0.5	Y	Y	N	N	N (1)	N (1)	Y (0.5)	N	Y	Y
CZ	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	Y (0.5)	Y	Y	NA
FM	[0, 1]	0.5	Y	Y	N	N	NA	NA	Y (0.5)	Y	Y	NA
G	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
GK	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
GL	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.67)	Y	Y	Y
H	[-1, 1]	0	Y	Y	N	N	Y (-1)	Y (-1)	Y (0)	Y	Y	Y
J	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.33)	Y	Y	NA
K	[0, 1]	0.5	Y	Y	N	N	NA	NA	Y (0.5)	Y	Y	NA
MC	[-1, 1]	0	Y	Y	N	N	NA	NA	Y (0)	Y	Y	NA
P	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	N	NA	NA
PE	[-1, 1]	0	N	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
RAND	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	Y (0.5)	Y	Y	Y
RR	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.25)	Y	Y	N(0)
RT	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.33)	Y	Y	Y
SS1	[0, 1]	0.5	Y	Y	Y	Y	NA	NA	Y (0.5)	Y	NA	NA
SS2	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.2)	Y	Y	NA
SS3	[0, 1]	0.5	Y	Y	N	N	NA	NA	N (0.25)	Y	NA	NA
W1	[0, 1]	0.5	N	Y	N	N	Y (0)	NA	Y (0.5)	Y	Y	NA
W2	[0, 1]	0.5	N	Y	N	N	NA	Y (0)	Y (0.5)	Y	Y	NA
Proportion			$\frac{18}{21}$	$\frac{21}{21}$	$\frac{6}{21}$	$\frac{6}{21}$	$\frac{9}{21}$	$\frac{9}{21}$	$\frac{15}{21}$	$\frac{18}{21}$	$\frac{14}{21}$	$\frac{5}{21}$

TABLE 2. Summary of different properties proposed in this paper.

Property	Brief Descriptions	Desirable Output
P1	Symmetry	$sim(\mathcal{M}^{(P)}, \mathcal{M}^{(Q)}) = sim(\mathcal{M}^{(Q)}, \mathcal{M}^{(P)})$
P2	Possibility of distributed computation	Included
P3	Detection of Uncorrelatedness (DoU)	Included
P3a	DoU between two random partitions	Baseline values
P3b	DoU between two random ensembles	Baseline values
P4	Measure of complementarity	Included
P4a1	$sim(1_{N\times N}, E)$	Minimum values
P4a2	$sim(E, 1_{N\times N})$	Minimum values
P4b	$sim(\mathcal{M}_U, \mathcal{M}_U)$	Baseline values
P4c	$sim(M^{(P)}, E - M^{(P)} + 1)$	Minimum values
P5	Measure of self-similarity	Included
P5a	$sim(1_{N\times N}, E)$	Maximum values
P5b	$sim(E, E)$	Maximum values

TABLE 1. Pair-counting similarity measures investigated in this paper.

Name	Notation	Reference	Definition	Range
Adjusted Rand Index	ARI	[15]	$\frac{a-\frac{(a+b)(a+c)}{\lambda}}{0.5(a+b+a+c)-\frac{(a+b)(a+c)}{\lambda}}$	(-1, 1]
Baulieu	B	[41]	$\frac{\lambda^2-\lambda(b+c)+(b-c)^2}{\lambda^2}$	[0, 1]
Czekanowski	CZ	[42]	$\frac{2a}{2a+b+c}$	[0, 1]
Fowlkes-Mallows	FM	[16]	$\frac{a}{\sqrt{(a+b)(a+c)}}$	[0, 1]
Gamma (Γ)	G	[15]	$\frac{\lambda a-(a+b)(a+c)}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$	[-1, 1]
Goodman and Kruskal	GK	[43]	$\frac{ad-bc}{ad+bc}$	[-1, 1]
Gower and Legendre	GL	[44]	$\frac{a+d}{a+0.5(b+c)+d}$	[0, 1]
Hamann	H	[45]	$\frac{(a+d)-(b+c)}{\lambda}$	[-1, 1]
Jaccard	J	[17]	$\frac{a}{a+b+c}$	[0, 1]
Kulczynski	K	[46]	$\frac{1}{2}(\frac{a}{a+b}+\frac{a}{a+c})$	[0, 1]
McConnaughey	MC	[47]	$\frac{a^2-bc}{(a+b)(a+c)}$	[-1, 1]
Pearson	P	[18]	$\frac{ad-bc}{(a+b)(a+c)(c+d)(b+d)}$	[-1, 1]
Peirce	PE	[48]	$\frac{ad-bc}{(a+c)(b+d)}$	[-1, 1]
RAND	R	[14]	$\frac{a+d}{\lambda}$	(0, 1]
Russel and Rao	RR	[49]	$\frac{a}{\lambda}$	[0, 1]
Rogers and Tanimoto	RT	[50]	$\frac{a+d}{a+2(b+c)+d}$	[0, 1]
Sokal and Sneath 1	SS1	[51]	$\frac{1}{4}(\frac{a}{a+b}+\frac{a}{a+c}+\frac{d}{d+b}+\frac{d}{d+c})$	[0, 1]
Sokal and Sneath 2	SS2	[51]	$\frac{a}{a+2(b+c)}$	[0, 1]
Sokal and Sneath 3	SS3	[51]	$\frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	[0, 1]
Wallance 1	W1	[16]	$\frac{a}{a+b}$	[0, 1]
Wallance 2	W2	[16]	$\frac{a}{a+c}$	[0, 1]

TABLE 3. Comparison of generalized measures based on different properties.

Measures	Range	Base	P1	P2	P3a	P3b	P4a1	P4a2	P4b	P4c	P5a	P5b
Desirable			Y	Y	Y	Y	Min	Min	Base	Min	Max	Max
ARI	[-1, 1]	0	Y	Y	Y	Y	N (0)	N (0)	Y (0)	N	NA	NA
B	[0, 1]	0.5	Y	Y	N	N	N (1)	N (1)	Y (0.5)	N	Y	Y
CZ	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	Y (0.5)	Y	Y	NA
FM	[0, 1]	0.5	Y	Y	N	N	NA	NA	Y (0.5)	Y	Y	NA
G	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
GK	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
GL	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.67)	Y	Y	Y
H	[-1, 1]	0	Y	Y	N	N	Y (-1)	Y (-1)	Y (0)	Y	Y	Y
J	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.33)	Y	Y	NA
K	[0, 1]	0.5	Y	Y	N	N	NA	NA	Y (0.5)	Y	Y	NA
MC	[-1, 1]	0	Y	Y	N	N	NA	NA	Y (0)	Y	Y	NA
P	[-1, 1]	0	Y	Y	Y	Y	NA	NA	Y (0)	N	NA	NA
PE	[-1, 1]	0	N	Y	Y	Y	NA	NA	Y (0)	Y	NA	NA
RAND	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	Y (0.5)	Y	Y	Y
RR	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.25)	Y	Y	N(0)
RT	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.33)	Y	Y	Y
SS1	[0, 1]	0.5	Y	Y	Y	Y	NA	NA	Y (0.5)	Y	NA	NA
SS2	[0, 1]	0.5	Y	Y	N	N	Y (0)	Y (0)	N (0.2)	Y	Y	NA
SS3	[0, 1]	0.5	Y	Y	N	N	NA	NA	N (0.25)	Y	NA	NA
W1	[0, 1]	0.5	N	Y	N	N	Y (0)	NA	Y (0.5)	Y	Y	NA
W2	[0, 1]	0.5	N	Y	N	N	NA	Y (0)	Y (0.5)	Y	Y	NA
Proportion			$\frac{18}{21}$	$\frac{21}{21}$	$\frac{6}{21}$	$\frac{6}{21}$	$\frac{9}{21}$	$\frac{9}{21}$	$\frac{15}{21}$	$\frac{18}{21}$	$\frac{14}{21}$	$\frac{5}{21}$

- Property P2 (Possibility of distributed computation) will facilitate the distributed computation of the measure values;
- Property P3 (Detection of Uncorrelatedness (DoU)) is well recognized in previous works in distinguishing meaningful partitions/ensembles from random partitions, while in this paper we extend this notion to the scenario of ensembles;
- Property P4 (Measure of complementarity) is important in identifying the most discriminative/uncertain complementary pairs;
- Property P5 (Measure of self-similarity) is important in measuring the degree of self-similarity of partitions/ensembles with different levels of uncertainty.

These observations might also provide an explanation for a number of problems discussed in previous works but not yet solved, which include: (1) Why Fowlkes-Mallows (FM) tends to vary within $[0.6, 1]$ and Rand (R) tends to vary within $[0.5, 0.95]$ for partitions with unbalanced data point distributions [24]; (2) Under what scenarios will negative values be observed for these generalized measures [8], [24]; (3) How these measures perform when applied under different conditions, e.g., between two random partitions (or two random ensembles)(Property 4), or between pairs of different consensus matrices (Property 5)?

V. EXPERIMENTS

We have conducted a number of experiments to investigate the properties of the generalized measures, as well as some of their applications. More specifically, we will mainly investigate the property, Detection of Uncorrelatedness (DoU), for different generalized measures. The dependence of DoU on different variants is consequently presented. We then conduct further experiments to compare these generalized measures based on a number of public data sets. Application of generalized measures to characterize the diversity of cluster ensembles is also discussed.

A. EXPERIMENTS: DETECTION OF UNCORRELATEDNESS (DoU)

The experiments in this subsection are conducted for the following purposes: (i) to verify the detection of uncorrelatedness property for different generalized measures; (ii) to investigate the effect of different factors, such as the number of clustering solutions L, the number of points N and the number of clusters K. In previous sections, we use vague descriptions such as "if N is sufficiently large". These experiments can shed some light on issues such as "What value of N is large enough?".

We first compare two groups of measures on the detection of uncorrelatedness property. Specifically, Group 1 contains six measures, which include 'ARI', 'G', 'GK', 'P', 'PE' and 'SS1'. Group 2 contains the other measures. We generate two random ensembles according to the following specification: the number of data points $N = 100$, the number of partitions for each ensemble $L = 10$, the maximum possible number of clusters $K = 5$ and the number of repeated trials $T = 20$. Figure 1 shows the absolute residual error values of different similarity values after their baseline values are subtracted. From this figure, we can observe that the mean values of all six measures in Group 1 are close to zero, while those in Group 2 are quite different from zero. The experiment results agree well with our analysis in the previous sections.

We further investigate which parameters affect the results corresponding to this property of DoU: the number of data points N, the number of partitions in ensembles L, or the maximum number of clusters K. We only study the performance of the measures in Group 1 for different parameter

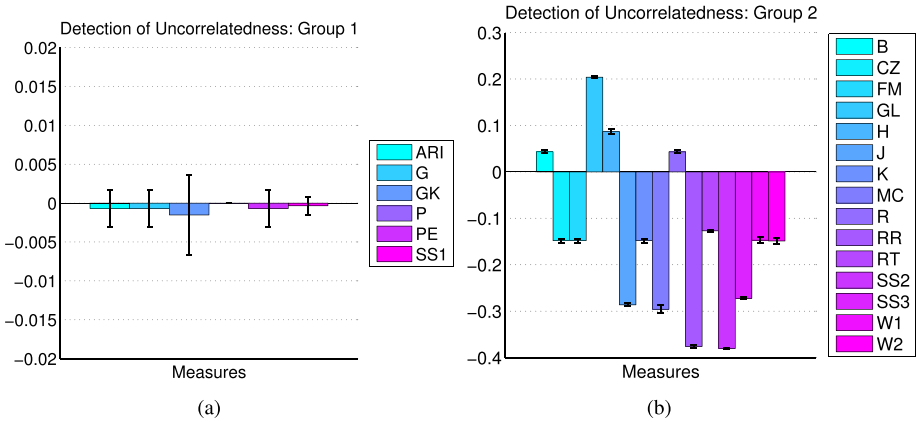


FIGURE 1. Detection of uncorrelatedness for two random ensembles. Group 1 includes six measures which have the desirable property(close to zero), while measures in Group 2 do not have the property (different from zero). (a) Group 1. (b) Group 2.

settings, ranging from $N_{min} = 100, L_{min} = 10, K_{min} = 5$ to $N_{max} = 1000, L_{max} = 100, K_{max} = 20$. To study the effect of a single parameter, we fix the other two parameters. The number of repeated trials is set to $T = 20$. Results of six different scenarios are shown in Figure 2. From this figure, we can observe that (i) the values of all six measures decrease when N increases (Figure 2(a) and Figure 2(b)) and when L increases (Figure 2(e) and Figure 2(f)), but are insensitive to K (Figure 2(c) and Figure 2(d)); (ii) the GK curves are higher than others; (iii) the ARI, G, and PE curves are very close to each other, and (iv) the P curve almost overlaps with the horizontal axis. The first observation indicates that the detection of uncorrelatedness property for well-behaved measures become more obvious with a larger number of data points and/or with a larger number of partitions. The last three observations are in good agreement with the definitions of the measures. These differences come from the fact that their numerators converge to zero while they have different denominators. Specifically, the denominators of ARI, G, GK, and PE are second-order factors of a, b, c, d, while that of P is fourth-order. Moreover, GK has the smallest denominator.

B. FURTHER EXPERIMENTS

We conduct further experiments using nine well-known public data sets from the UCI machine learning repository,[1] including UCI-Breast-Cancer-Wisconsin, UCI-BCW, UCI-Chart, UCI-Glass, UCI-Iris, UCI-Image-Segmentation, UCI-Pima, UCI-Vehicle, and UCI-Wine. These have been used to evaluate the performance of different previous clustering and cluster ensemble techniques.

1) COMPARISON BETWEEN A PARTITION AND A SIMILARITY MATRIX

An intuitive application of the generalized measures is to evaluate similarity between a partition P and an ensemble $\mathcal{Q}$ (with consensus matrix $\mathcal{M}^{(\mathcal{Q})}$). Application of two generalized Adjusted Rand indices (ARImp and ARImm) under this scenario were explored in our recent works [20], [21]. Note that for other measures but not including the pair-counting similarity measures, researchers tend to use the mean value of these measures between the partition P and each partition $Q^{(l)}$, i.e., $sim(P, \mathcal{Q}) = \frac{1}{L}\sum sim(P, Q^{(l)})$ for the same purpose [5], [10]. Thus, a study of the relationship between the results based on $sim(P, \mathcal{M}^{(\mathcal{Q})}))$ and the traditional method (i.e., $sim(P, \mathcal{M}^{(\mathcal{Q})})) = \frac{1}{L}\sum sim(P, Q^{(l)}))$ is of great interest. For each UCI data set, we run Kmeans with the true number of clusters to obtain an initial clustering solution in each trial. Next, we generate a random value $L \in [25, 250]$, and run Kmeans L times to obtain a corresponding number of partitions. We apply each measure to these partitions using our method and the traditional one respectively, and the mean Pearson correlation coefficient between the two different sets of results averaged across 20 trials are reported in Figure 3. Despite the different characteristics of the data sets, experimental results of most of the measures are highly correlated to those of the traditional method, except for B, GK, W1 and W2. These results suggest that most of our generalized measures can achieve results similar to those of the traditional method, while our approach only requires access to the consensus matrix of the ensemble, without the need to observe each individual partition. This advantage becomes more important when the similarity matrix is not constructed from a partition but directly specified. It is also interesting to note that the values for W1 and W2 are uniformly low across all the datasets. This observation might be due to how these two measures are formulated as follows: (i) their definitions do not include the factor d, which is usually larger than the other three factors; and (ii) their numerators include either b or c only ($a + b$ for W1 and $a + c$ for W2), while most of the other measures have both the b and c factors present in an interchangeable way.

[1] http://archive.ics.uci.edu/ml/datasets.html

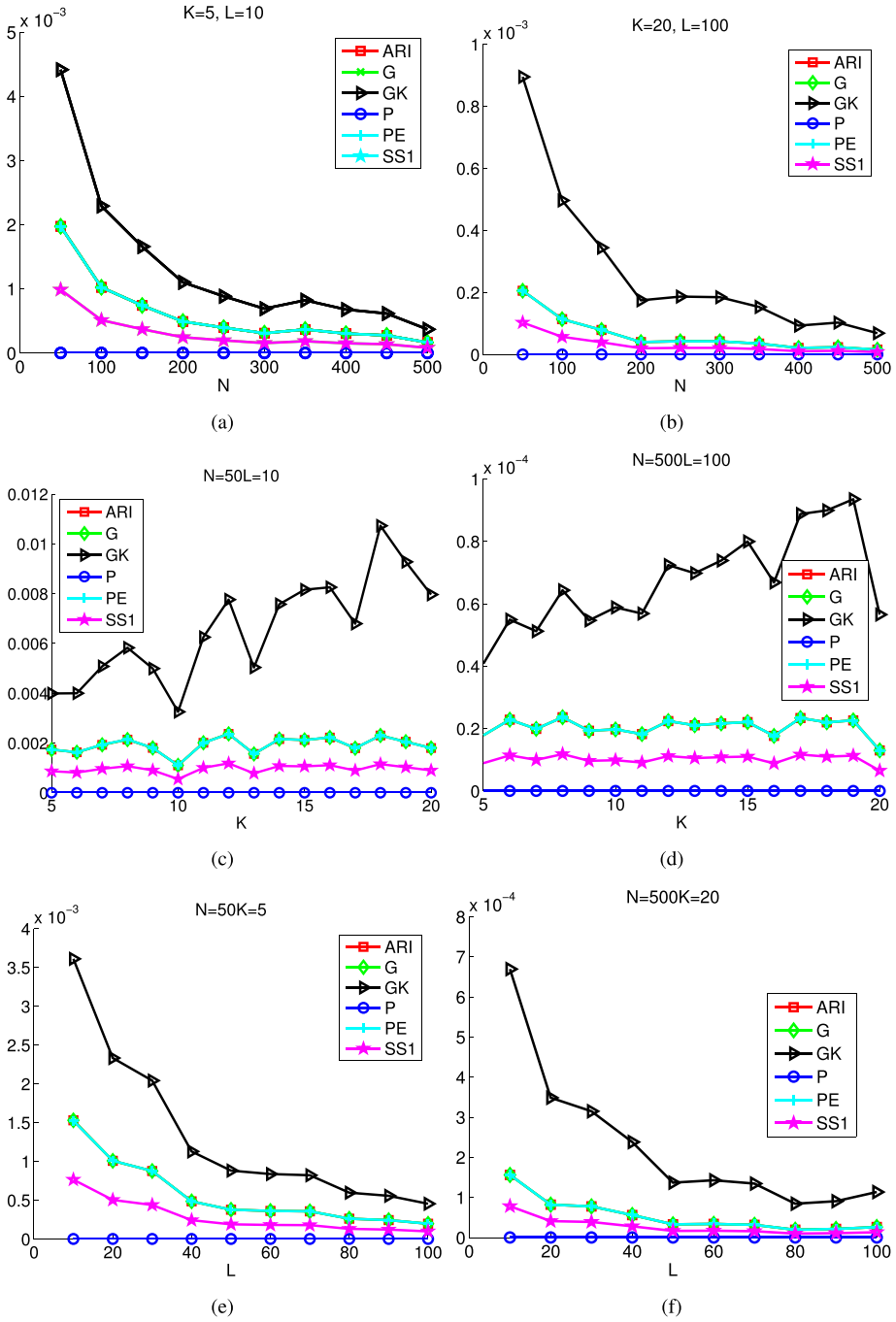


FIGURE 2. Investigation of the effect of different parameters on the detection of uncorrelatedness property: the number of data points N, the number of partitions in ensembles L, and the maximum number of clusters K.

2) OVERALL CORRELATION OF DIFFERENT MEASURE PAIRS IN THE COMPARISON BETWEEN A PARTITION AND A SIMILARITY MATRIX

We have also obtained the pairwise correlation of the 21 generalized measures based on the comparison between a partition and a similarity matrix in the last subsection, which we visualize in Figure 4. Although we only compare these measures for the nine UCI data sets, we can already observe their diverse behavior. It is interesting to note that some measure pairs have quite large correlation, e.g., ARI and G, CZ and J, CZ and FM, SS1 and SS3. On the other hand, H, W1, and W2 appear to be the most different measures compared to the others. We hope that this comparison might provide some help when practitioners choose the desirable subset of measures for their own task. These observations might also be useful when we need to select multiple

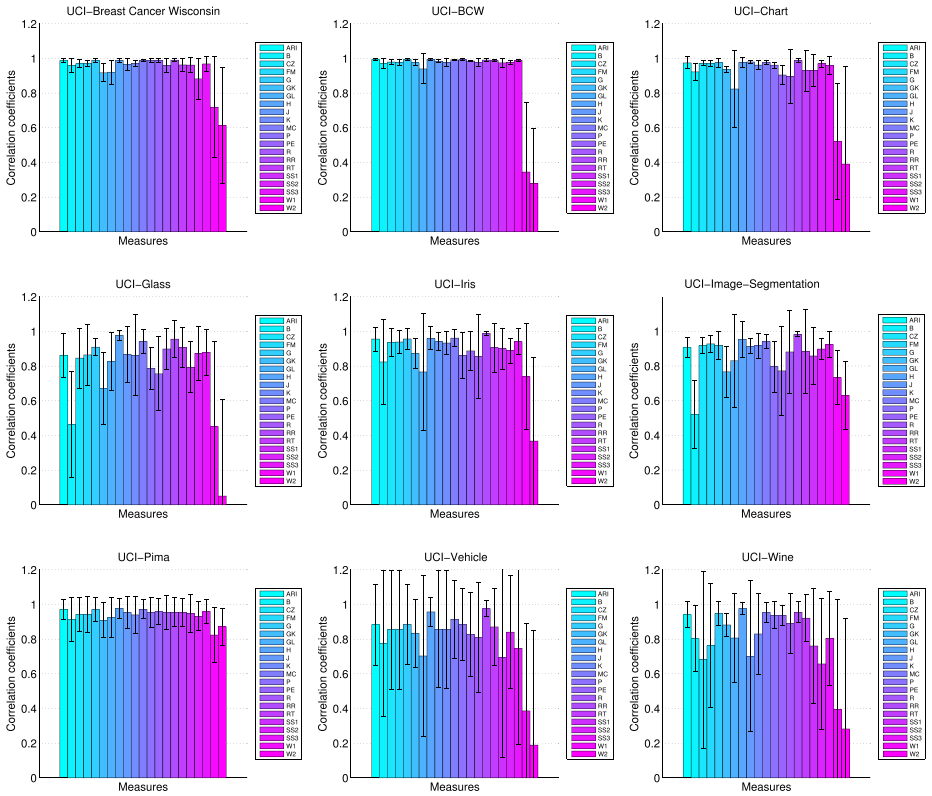


FIGURE 3. **Application of generalized measures to characterize the similarity between a partition and an ensemble. Results show correlation coefficient values between our methods and traditional methods.**

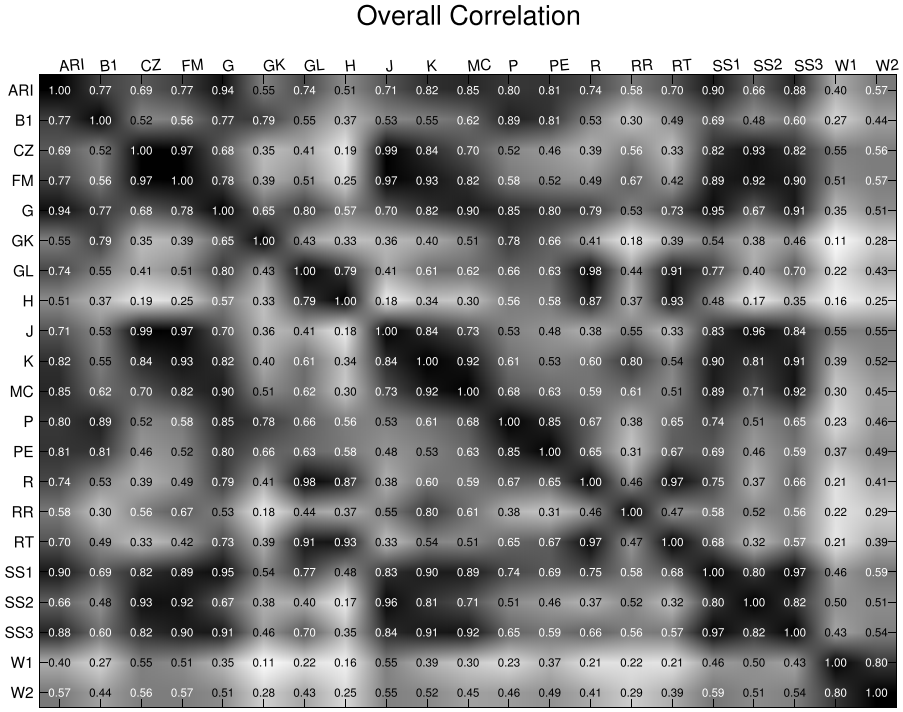


FIGURE 4. **Overall pairwise correlation of the 21 generalized measures on nine datasets.**

generalized measures to perform a more objective comparison between different clustering (or cluster ensemble) algorithms.

3) MEASURING THE DIVERSITY OF CLUSTER ENSEMBLES

Previous works suggest that the quality of a cluster ensemble is related to its diversity [5], [37], [53]. In general,

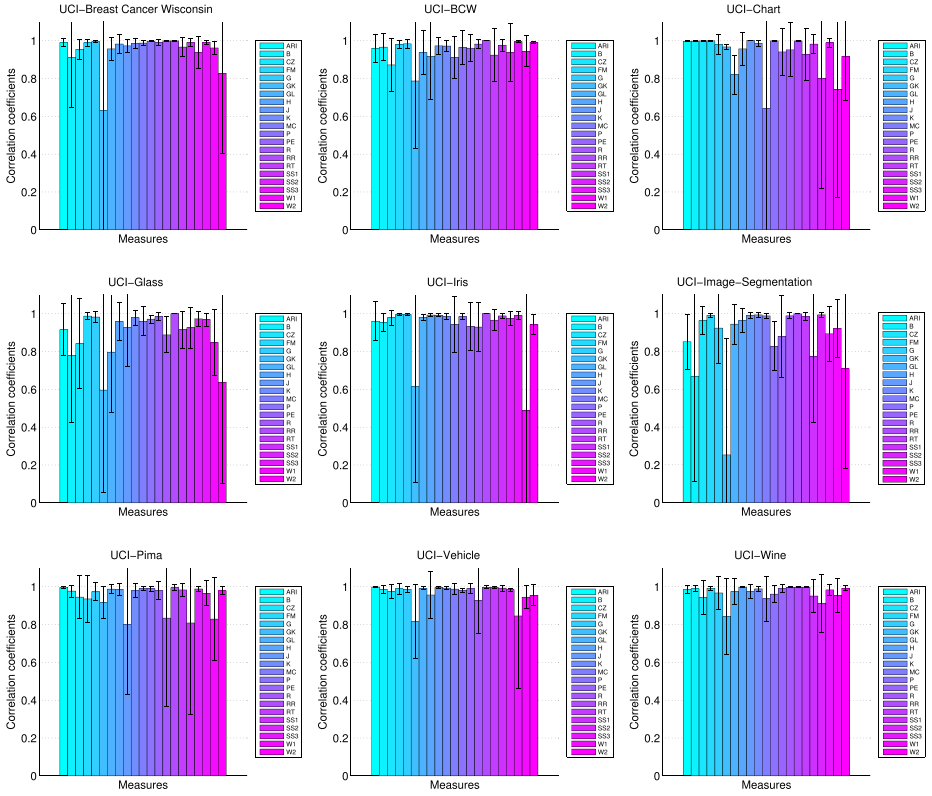


FIGURE 5. Application of generalized measures to characterize the diversity of cluster ensembles. Results show correlation coefficient values between our methods and traditional methods.

traditional methods use pairwise similarity between clustering solutions to measure the diversity of a cluster ensemble (i.e., $\frac{1}{\lambda}\sum_{i=1}^{N}\sum_{j=i+1}^{N} sim(P^{(i)}, P^{(j)})$) such as Pairwise Normalized Mutual Information (PNMI) [5], [37]. In this application, motivated by the traditional method, we use $sim(\mathcal{M}^{(P)}, \mathcal{M}^{(P)})$ between ensembles, rather than partition pairs, to approximate the traditional method. The relation between these two methods is then discussed.

Specifically, we generate 600 clustering solutions for different UCI data sets with the Kmeans algorithm. We use different cluster numbers sampled at random, and group these solutions into three classes using the spectral clustering algorithm as performed in [5]. We refer to these three classes as the small cluster class, the medium cluster class, and the large cluster class according to their sizes. Initially, we add the small cluster class into a base group, and compute the similarity between the clustering solutions in the base group using the different generalized measures. Then we divide the medium cluster class into four different groups at random, and add these groups to the base group one by one in ascending order of their sizes. The corresponding similarity at this stage is also computed using the different generalized measures. Finally, the large cluster class is also divided and added to the base group, followed by the computation of their corresponding similarity values. In this way, the diversity of the 600 clustering solutions can be investigated under nine different conditions. For each condition, the similarity values are computed with our method and the traditional method, respectively, and the Pearson correlation coefficient between the two different computation results are reported in Figure 5.

Interestingly, except for some measures such as B, GK, W1 and W2 which behave in an unstable way, most of the measure values such as ARI (the 1st column), FM (the 4th column) and G (the 5th column) are highly correlated to the results of the traditional methods. Note that our approach only needs to access the consensus matrix of the ensemble, while the traditional computational method requires access to each individual partition. Thus our approach is more general, and its unique advantage is especially useful when access to the individual clustering solutions is difficult.

Could be usefull to use this method instead of more traditional pairwise methods -> requires access to less data

C. FURTHER DISCUSSIONS

We have analyzed our proposed properties for generalized pair-counting similarity measures in the scenarios of partitions and of cluster ensembles from both the perspectives of theoretical analysis and experimental study. As can be seen, each proposed property has its merit, and is possessed by a number of popular measures. Notably, we do not aim at discovering a complete set of properties to evaluate these measures. Instead, we propose a number of important properties to investigate these measures, especially in the scenario of cluster ensembles. These properties can thus serve as important criteria for the design and selection of evaluation

We should evaluate our clusters with a few, figure out what we are trying to measure most

measures for clustering solutions. In general, there is no single measure which possesses all the desirable properties. Thus, we do not incline to recommend any particular measure for evaluation. Instead, using multiple measures for clustering evaluation is more reasonable and less biased. Therefore, it is important to select a number of diverse measures. In addition, our analysis and experimental results discover measures that have a high correlation with each other. For example, we can find measure pairs which are highly correlated from Figure 4, e.g., ARI vs. G (0.94), CZ vs. FM (0.97), CZ vs. J (0.99), CZ vs. SS2 (0.93), FM vs. J (0.97), FM vs. K (0.93), FM vs.SS2 (0.92), FM vs. SS3 (0.90), G vs. MC (0.90), GL vs. R (0.98), GL vs. RT (0.91), H vs. RT (0.93), J vs. SS2 (0.96), K vs. MC (0.92), K vs. SS1 (0.90), K vs. SS3 (0.91), MC vs. SS3 (0.92), R vs. RT (0.97), and SS1 vs. SS3 (0.97).

An interesting extension of generalized similarity measures to be explored is the case of data labels (or similarity matrices) with missing values. This kind of problems can be dealt with using two different methods: 1) prediction-based methods: we can predict the missing values based on other related entries in the matrices. 2) dropout-based methods: we can simply remove the missing entries for both matrices, and revise the normalized factors in related equations, e.g., (4).

VI. CONCLUSION

In this paper, we have compared 21 pair-counting similarity measures in a generalized setting based on both single clustering solutions and cluster ensemble results, and analyzed their desirable properties from both the perspectives of theoretical analysis and experimental study. We identify their different behaviors and their correlations in different scenarios. It is interesting to observe that each property is possessed by at least a few generalized measures. Notably, some measures have similar performance with regard to these properties in spite of their different formulations. These properties can also serve as important criteria for the design and selection of evaluation measures. We have also performed a number of experiments to verify the comparison criteria, and to demonstrate the performance of the different generalized measures in practical applications.

REFERENCES

[1] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proc. 2nd Eur. Conf. Comput. Learn. Theory*, 1995, pp. 23–37.

[2] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[3] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.

[4] X. Z. Fern and C. E. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004. p. 36.

[5] X. Z. Fern and W. Lin, "Cluster ensemble selection," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, 2008, pp. 787–797.

[6] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," *Mach. Learn.*, vol. 52, nos. 1–2, pp. 91–118, Jul. 2003.

[7] A. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, Jun. 2005.

[8] L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of K-means cluster ensembles with respect to random initialization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1798–1808, Nov. 2006.

[9] H. G. Ayad and M. S. Kamel, "Cumulative voting consensus method for partitions with variable number of clusters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 160–173, Jan. 2008.

[10] J. Azimi and X. Fern, "Adaptive cluster ensemble selection," in *Proc. 21st Int. Joint Conf. Artif. Intell.*, 2009, pp. 992–997.

[11] N. Iam-On, T. Boongoen, and S. Garrett, "LCE: A link-based cluster ensemble method for improved gene expression data analysis," *Bioinformatics*, vol. 26, no. 12, pp. 1513–1519, 2010.

[12] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2396–2409, Dec. 2011.

[13] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Trans. Knowl. Discovery From Data*, vol. 1, no. 1, p. 4, 2007.

[14] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[15] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[16] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *J. Amer. Stat. Assoc.*, vol. 78, no. 383, pp. 553–569, 1983.

[17] P. Jaccard, "Etude de la distribution florale dans une portion des Alpes et du Jura," *Bull. Soc. Vaudoise Sci. Naturelles*, vol. 37, no. 142, pp. 547–579, 1901.

[18] A. N. Albatineh, M. Niewiadomska-Bugaj, and D. Mihalko, "On similarity indices and correction for chance agreement," *J. Classification*, vol. 23, no. 2, pp. 301–313, 2006.

[19] R. J. G. B. Campello, "A fuzzy extension of the rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, no. 7, pp. 833–841, May 2007.

[20] S. Zhang and H.-S. Wong, "ARImp: A generalized adjusted rand index for cluster ensembles," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Istanbul, Turkey, Aug. 2010, pp. 778–781.

[21] S. Zhang, H.-S. Wong, and Y. Shen, "Generalized adjusted rand indices for cluster ensembles," *Pattern Recognit.*, vol. 45, no. 6, pp. 2214–2226, Jun. 2012.

[22] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on Web-page clustering," in *Proc. Workshop Artif. Intell. Web Search (AAAI)*, 2000, pp. 58–64.

[23] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," in *Proc. KDD Workshop Text Mining*, vol. 400. Boston, MA, USA, 2000, pp. 525–526.

[24] M. Meilă, "Comparing clusterings—An information based distance," *J. Multivariate Anal.*, vol. 98, no. 5, pp. 873–895, May 2007.

[25] N. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Is a correction for chance necessary?" in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 1073–1080.

[26] P. Luo, H. Xiong, G. Zhan, J. Wu, and Z. Shi, "Information-theoretic distance measures for clustering validation: Generalization and normalization," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1249–1262, Sep. 2009.

[27] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 1, pp. 2837–2854, Oct. 2010.

[28] S. Van Dongen, "Performance criteria for graph clustering and markov cluster experiments," Centrum Wiskunde & Inform., Amsterdam, The Netherlands, Tech. Rep. INSR0012, 2000.

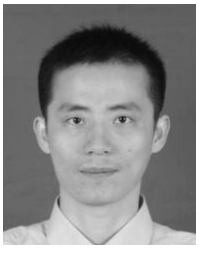
[29] M. Meilă and D. Heckerman, "An experimental comparison of model-based clustering methods," *Mach. Learn.*, vol. 42, nos. 1–2, pp. 9–29, Jan. 2001.

[30] B. Larsen and C. Aone, "Fast and effective text mining using linear-time document clustering," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 1999, pp. 16–22.

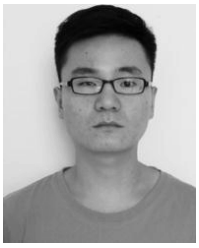
[31] D. T. Anderson, J. C. Bezdek, M. Popescu, and J. M. Keller, "Comparing fuzzy, probabilistic, and possibilistic partitions," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 906–918, Oct. 2010.

[32] X. He, C. H. Q. Ding, H. Zha, and H. D. Simon, "Automatic topic identification using webpage clustering," in *Proc. IEEE Int. Conf. Data Mining*, Nov./Dec. 2001, pp. 195–202.

[33] J. Neville, M. Adler, and D. Jensen, "Clustering relational data using attribute and link information," in *Proc. IJCAI Text Mining Link Anal. Workshop*, 2003, pp. 9–15.
[34] P. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[35] P. W. Lord, R. D. Stevens, A. Brass, and C. A. Goble, "Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation," *Bioinformatics*, vol. 19, no. 10, pp. 1275–1283, Jul. 2003.
[36] A. Schlicker, F. Domingues, J. Rahnenführer, and T. Lengauer, "A new measure for functional similarity of gene products based on gene ontology," *BMC Bioinformatics*, vol. 7, no. 1, p. 302, 2006.
[37] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 186–193.
[38] J. Wu, S. Zhu, H. Xiong, J. Chen, and J. Zhu, "Adapting the right measures for pattern discovery: A unified view," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1203–1214, Aug. 2012.
[39] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.
[40] J. Wu, H. Xiong, and J. Chen, "Adapting the right measures for K-means clustering," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 877–886.
[41] F. B. Baulieu, "A classification of presence/absence based dissimilarity coefficients," *J. Classification*, vol. 6, no. 1, pp. 233–246, Dec. 1989.
[42] J. Czekanowski, "Coefficient of Racial Likeness und Durchschnittliche Differenz," *Anothropologidcher*, vol. 14, pp. 227–249, Jan. 1932.
[43] L. Goodman and W. H. Kruskal, "Measures of association for cross classifications," *J. Amer. Stat. Assoc.*, vol. 49, no. 268, pp. 732–764, 1954.
[44] J. C. Gower and P. Legendre, "Metric and Euclidean properties of dissimilarity coefficients," *J. Classification*, vol. 3, no. 1, pp. 5–48, 1986.
[45] U. Hamann, "Weiteres über merkmalsbestand und verwandtschaftsbeziehungen der 'Farinosae,'" *Willdenowia*, vol. 3, no. 1, pp. 169–207, 1962.
[46] S. Kulczynski, "Die pflanzenassociationen der pienenen," *Bull. Int. Acad. Polonaise Sci. Lett., Classe Sci. Math. Naturelles B*, vol. 2, pp. 57–203, 1927.
[47] B. H. McConnaughey and L. P. Laut, *The Determination and Analysis of Plankton Communities*. Indonesia: Lembaga Penelitian Laut, 1964.
[48] C. S. Peirce, "The numerical measure of the success of predictions," *Science*, vol. 4, no. 93, pp. 453–454, Nov. 1884.
[49] P. Russel and T. Rao, "On habitat and association of species of anopheline larvae in south-eastern madras," *J. Malaria Inst. India*, vol. 3, no. 1, pp. 153–178, 1940.
[50] D. J. Rogers and T. Tanimoto, "A computer program for classifying plants," *Science*, vol. 132, no. 3434, pp. 1115–1118, 1960.
[51] R. Sokal and P. Sneath, *Principles of Numerical Taxonomy*. San Francisco, CA, USA: Freeman, 1963.
[52] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn. (EMNLP-CoNLL)*, 2007, pp. 410–420.
[53] L. I. Kuncheva, S. T. Hadjitodorov, and L. P. Todorova, "Experimental comparison of cluster ensemble methods," in *Proc. 9th Int. Conf. Inf. Fusion*, Jul. 2006, pp. 1–7.



SHAOHONG ZHANG received the Ph.D. degree from the Department of Computer Science, City University of Hong Kong. He was a Post-Doctoral Fellow with the Department of Computer Science, City University of Hong Kong. He is currently an Associate Professor with the Department of Computer Science, Guangzhou University. His research interests include pattern recognition, data mining, and bioinformatics.

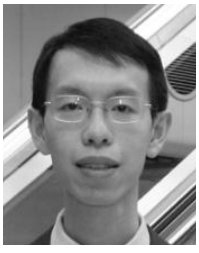


ZONGBAO YANG received the B.S. degree from Guangdong Medical University, Dongguan, China, in 2016. He is currently pursuing the M.S. degree with Guangzhou University, Guangzhou, China. His research interests include data mining and citation network.

XIAOFEI XING (M'17) received the Ph.D. degree in computer science from Central South University, China, in 2012. He has been a Visiting Research Fellow with the University of Tsukuba, Japan. He is an Assistant Professor with the Department of Computer Science, Guangzhou University. His research interests include modeling and performance evaluation in wireless sensor networks and mobile computing. He is a member of the China Computer Federation.

YING GAO received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2002. He is currently a Professor with the School of Computer Science and Educational Software, Guangzhou University, Guangzhou. His main research interests include intelligent optimization algorithms, pattern recognition, and signal processing.

DONGQING XIE was born in Hunan, China, in 1965. He received the M.S. degree from Xidian University in 1988 and Ph.D. degree from Hunan University in 1999. He has been a Professor with Hunan University since 2001. He is currently a Professor with the Department of Computer Science, Guangzhou University. His research interest includes pattern recognition, information security, algorithm analysis and design.



HAU-SAN WONG was a Research Associate with the School of Electrical and Information Engineering, The University of Sydney, and a Post-Doctoral Teaching Fellow with the Department of Computer Science, Hong Kong Baptist University. He is currently an Associate Professor with the Department of Computer Science, City University of Hong Kong. His research interests include bioinformatics and machine learning.