Wisdom of Crowds Cluster Ensemble

Hosein Alizadeh[1], Muhammad Yousefnezhad[2] and Behrouz Minaei Bidgoli[3]

Abstract: The *Wisdom of Crowds* is a phenomenon described in social science that suggests four criteria applicable to groups of people. It is claimed that, if these criteria are satisfied, then the aggregate decisions made by a group will often be better than those of its individual members. Inspired by this concept, we present a novel feedback framework for the cluster ensemble problem, which we call Wisdom of Crowds Cluster Ensemble (WOCCE). Although many conventional cluster ensemble methods focusing on diversity have recently been proposed, WOCCE analyzes the conditions necessary for a crowd to exhibit this collective wisdom. These include decentralization criteria for generating primary results, independence criteria for the base algorithms, and diversity criteria for the ensemble members. We suggest appropriate procedures for evaluating these measures, and propose a new measure to assess the diversity. We evaluate the performance of WOCCE against some other traditional base algorithms as well as state-of-the-art ensemble methods. The results demonstrate the efficiency of WOCCE's aggregate decision-making compared to other algorithms.

Keywords: *Ensemble Cluster, Wisdom of Crowds, Diversity, Independence.*

1. Introduction

Clustering, one of the main tasks of data mining, is used to group non-labeled data to find meaningful patterns. Generally, different models provide predictions with different accuracy rates. Thus, it would be more efficient to develop a number of models using different data subsets, or utilizing differing conditions within the modeling methodology of choice, to achieve better results. However, selecting the *best* model is not necessarily the ideal choice, because potentially valuable information may be wasted by discarding the results of less-successful models (Perrone and Cooper, 1993; Tumer and Ghosh, 1996; Baker and Ellison, 2008).

[1]Hosein Alizadeh
Department of Computer Eng., Iran University of Science and Technology, Tehran, Iran
e-mail: halizadeh@iust.ac.ir

[2]Muhammad Yousefnezhad
Department of Information Technology, Mazandaran University of Science and Technology, Babol, Iran
e-mail: myousefnezhad@ustmb.ac.ir

[3]Behrouz Minaei Bidgoli
Department of Computer Eng., Iran University of Science and Technology, Tehran, Iran
e-mail: b_minaei@iust.ac.ir

This leads to the concept of *combining*, where the outputs (individual predictions) of several models are pooled to make a better decision (collective prediction) (Tumer and Ghosh, 1996; Baker and Ellison, 2008). Research in the *Clustering Combination* field has shown that these pooled outputs have more strength, novelty, stability, and flexibility than the results provided by single algorithms (Strehl and Ghosh, 2002; Topchy et al., 2003; Fred and Lourenco, 2008; Ayad and Kamel, 2008). In the social science arena, there is a corresponding research field known as the *Wisdom of Crowds*, after the book of the same name (Surowiecki, 2004). Put simply, the Wisdom of Crowds (WOC) is the phenomenon whereby decisions made by aggregating the information of groups usually have better results than those made by any single group member. The book presents numerous case studies and anecdotes to illustrate its argument, and touches on several fields, primarily economics and psychology. Surowiecki justifies his own theory, stating that: "If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, the average of those answers, will cancel individual estimate's errors out. Each person's guess, you might say, has two components: information and error. Subtract the error, and you're left with the information" (Surowiecki, 2004).

In spite of the lack of a well-defined agreement on metrics in cluster ensembles, Surowiecki has suggested a clear structure for building a wise crowd. Supported by many examples from business, economics, societies, and nations, he has argued that a wise crowd must satisfy four conditions, namely diversity, independence, decentralization, and an aggregation mechanism. The goal of this paper is to use the WOC in order to choose a proper subset in a cluster ensemble. Whereas Surowiecki's definition of the WOC is related to social problems, and the decision elements constructed in his definitions are personal opinions, this paper proposes a mapping between cluster ensemble literature and the WOC phenomenon. According to this mapping, a new WOC Cluster Ensemble (WOCCE) framework, which employs the WOC definition of well-organized crowds, is proposed. Experimental results on a number of datasets show that WOCCE efficiently improves the final results compared to similar cluster ensemble methods.

In summary, the main contributions of this paper are:

- A new framework for generating a cluster ensemble from basic (primary) clustering results with *feedback*. WOCCE controls the quality of the ensemble using this feedback.
- A new mapping between the WOC observation (an approach to social problems) and the cluster ensemble problem (one of the main fields in data mining).This allows us to apply the definitions of a wise crowd to the cluster ensemble arena.
- A new heuristic method for measuring independence according to the wise crowd definitions.
- A new diversity metric called A3, which is based on the Alizadeh–Parvin–Moshki–Minaei (APMM) criterion (Alizadeh et al., 2011). A3 measures the diversity of a partition with respect to a reference set (an ensemble).

The rest of this paper is organized as follows. Section 2 reviews some relevant literature. In Section 3, we propose our new framework, and demonstrate the results of a comparison against traditional methods in Section 4. Finally, we present our conclusions in Section 5.

2. Literature review

2.1. Cluster Ensemble

In forecasting, *cluster ensemble* has been demonstrated that better results can be achieved by combining forecasts instead of choosing the best one. This has led to the idea of an ensemble in machine learning, where the component models (also known as members) are redundant in that each provides a solution to the same task, even though this solution may be obtained by different means (Grofman and Owen, 1996; Baker and Ellison, 2008).

Generally, a cluster ensemble has two important steps (Jain et al., 1999; Strehl and Ghosh, 2002):

1- Generating different results from primary clustering methods using different algorithms and changing the number of partitions. This step is called *generating diversity or variety*.

2- Aggregating mechanisms for combining primary results and generating the final ensemble. This step is performed by consensus functions (aggregating algorithms).

It is clear that an ensemble with a set of identical models cannot provide any advantages. Thus, the aim is to combine models that predict different outcomes, and there are four parameters that can be changed to achieve this goal. A set of models can be created from two approaches: 1.Choice of data representation, and 2.Choice of clustering algorithms or algorithmic parameters. Strehl and Ghosh (2002) proposed the Mutual Information (MI) metric for measuring the consistency of data partitions; Fred and Jain (2005) proposed Normalized Mutual Information (NMI), which is independent of cluster size. Fern and Lin (2008) developed a method that effectively uses a selection of the basic partitions to participate in the ensemble, and consequently in the final decision. They also employed the Sum NMI and Pairwise NMI as quality and diversity metrics, respectively, between partitions. Alizadeh et al. (2011, 2012) have explored the disadvantages of NMI as a symmetric criterion. They used the APMM and MAX metrics to measure diversity and stability, respectively, and suggested a new method for building a co-association matrix from a subset of base cluster results. This paper uses A3 for diversity measurement which works base on the APMM measure. Additionally, we use the co-association matrix construction scheme of Alizadeh et al. (2011). A3 and the co-association matrix are discussed in detail in Sections 3.1 and 3.4, respectively.

2.2. The Wisdom of Crowds

The Wisdom of Crowds (Surowiecki, 2004) presents numerous case studies, primarily in economics and psychology, to illustrate how the prediction performance of a crowd is better than that of its individual members. The book relates to diverse collections of independent individuals, rather than crowd psychology as traditionally understood. Its central thesis, that a diverse collection of individuals making independent decisions is likely to make certain types of decisions and predictions better than individuals or even experts, draws many parallels with statistical sampling, but there is little overt discussion of statistics in the book. Mackey (Mackey 1841) mentions that not every crowd is wise. These key criteria separate wise crowds from irrational ones (Surowiecki, 2004):

Diversity of opinion: Each person has private information, even if it is only an eccentric interpretation of the known facts.

Independence: People's opinions are not determined by the opinions of those around them.

Decentralization: People are able to specialize and draw on local knowledge.

Aggregation: Some mechanism exists for turning private judgments into a collective decision.

It is important to note that, under some conditions, the cooperation of the crowd will fail because of the consciousness of its members about each other's opinion. This will lead them to conform rather than think differently. Although each member of the crowd may attain greater knowledge and intelligence by this effect, definitely the whole crowd as a whole will become trapped into less unwise (Mackey, 1841; Page, 2007; Hadzikadic and Sun, 2010).

In recent years, the WOC has been used in the field of machine learning. Steyvers et al. (2009) used WOC for recollecting order information, and Miller et al. (2009) proposed an approach to the rank ordering problem. Welinder et al. (2010) used a multidimensional WOC method to estimate the underlying value (e.g., the class) of each image from (noisy) annotations provided by multiple annotators. WOC has also been applied to underwater mine classification with imperfect labels (Williams, 2010) and minimum spanning tree problems (Yi et al., 2010). Finally, Baker and Ellison (2008) proposed a method for using the WOC in ensembles and modules in environmental modeling.

3. The WOCCE approach

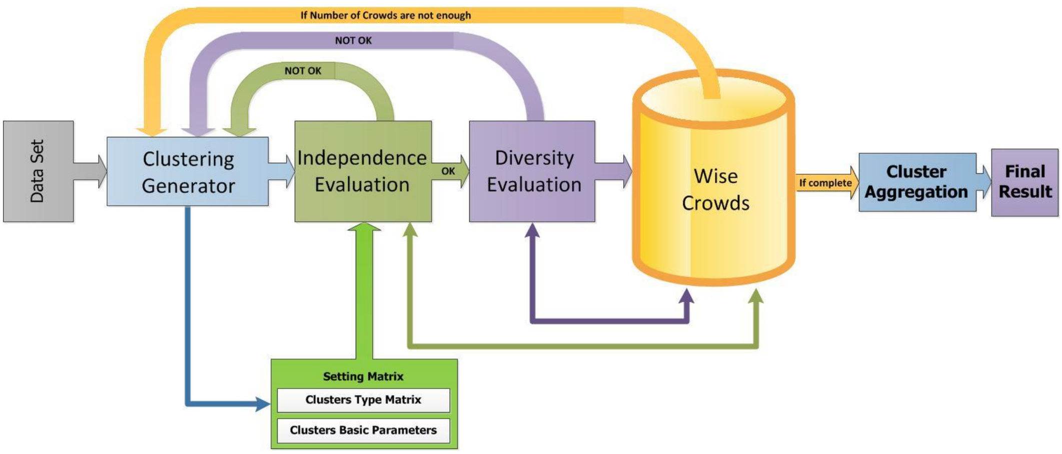

Fig. 1. The WOCCE framework

Surowiecki (2004) has outlined the conditions that are necessary for the crowd to be wise: *diversity, independence, and a particular kind of decentralization*.

To map the WOC to a cluster ensemble, we should restate the wise crowd requirements for the corresponding field. This section discusses these preconditions in detail for the area of clustering. It seems that the best matching between individuals and their opinions in WOC is base clustering algorithms and partitions, respectively, in the context of cluster ensembles. The goal of WOCCE, as illustrated in Figure (1), is to construct a wise crowd in the primary partition via a recursive procedure.

3.1. Diversity of Opinions

To define the diversity of opinion in cluster ensembles, which utilize base partitions, we use the term *diversity of base partitions*. According to this assumption, and Surowiecki's definition of diversity of opinion, it should be rephrased as:

If the result of a base clustering algorithm has less similarity value than a defined threshold in comparison with other partitions existing in the ensemble, it is eligible to be added to the ensemble.

Similarity and repetition of specific answers can be controlled by tracing errors. This paper proposes a new method, The A3, based on AAPMM in order to evaluate the diversity of each primary cluster algorithm. To calculate the similarity of cluster C with respect to a set of partitions in the reference set contacting M partitions, This paper uses AAPMM which is shown in equation (1) (Alizadeh et al., 2011):

$$AAPMM(C) = \frac{1}{M}\sum_{j=1}^{M} APMM\left(C, P_j^{b*}\right), \quad (1)$$

where P_j^{b*} is the corresponding derivation from the *j-th* partition in the reference set. *APMM*(*C*,*P*) is the similarity between cluster C and a specific partition (Alizadeh et al., 2011).

In order to measure the similarity of a whole partition, this paper propose saver aging AAPMM over all of the clusters that exist in a specific partition. We call this average measure *A3*. In other words, A3 is a weighted average of the AAPMMs of one partition's clusters:

$$A3(p) = \frac{1}{n}\sum_{i=1}^{k} n_i \times AAPMM(C_i)\ . \qquad (2)$$

In equation (3), C_i is the *i-th* cluster in partition *p*, and C_i has size n_i. *n* is the number of members in partition *p* and *k* is the number of clusters in the partition. A3 measures information between a partition and those partitions in a reference set. In fact, it counts the repetition of clusters in the corresponding set. Therefore, A3 measures the similarity of a partition with respect to a set. As it is normalized between zero and one, we use *1 – A3* to represent the diversity:

$$Diversity(p) = 1 - A3(p)\ . \qquad (3)$$

According to the above definitions, one of the conditions for appending a partition to the crowd (known as the diversity condition) is:

$$Diversity(p) \geq dT\ . \qquad (4)$$

This means that if the diversity of a generated partition with respect to a crowd satisfies the minimum threshold of *dT* (diversity threshold), it will be added to the crowd.

3.2. Independence of Opinions

According to Surowiecki's definition, independence means that an opinion must not be influenced by an individual or certain group. By mapping this to cluster ensembles, we have the following definition:

The decision making mechanism of each base clustering algorithm must be different. In the case of using similar algorithms, the basic parameters that determine their final decisions must be sufficiently different.

In other words, a new partition generated by a primary clustering algorithm is independent if and only if it satisfies the following conditions:

1) Every two partitions that are generated by different methods are independent because their algorithm's mechanisms are different.
2) Every two partitions that are generated by the same method with different basic parameters are independent.

This suggests that the independence of the results generated by a single algorithm should be investigated by checking the basic parameters. As most of the base algorithms are quite sensitive to their initial conditions, we propose a system of initialization checking to ensure that independent results are generated by each algorithm. The procedure Basic-Partition-Independence (BPI) illustrated in Figure (2) has been developed to calculate the independence of two partitions.

```
Function BPI (P1, P2) Return Result
        If (Algorithm-Type (P1) == Algorithm-Type (P2) then
                Result = 1 - Likeness (Basic-Parameter (P1), Basic-Parameter (P2))
        Else
                Result = 1
        End if
End Function
```

Fig. 2. Measuring the degree of independence between two clusters

In Figure (2), *P1* and *P2* are base partitions, the *Algorithm-Type* function returns the type of base algorithm that created those partitions, and the *Basic-Parameter* function returns the basic parameters of the algorithm that generated the partition (for example, the seed points of Kmeans). These values can be defined according to two factors: the nature of the problem and the type of base algorithms. This paper proposes a heuristic function (*Likeness*) for measuring a cluster's independence.

In order to calculate the *Likeness*, we assume that *MaxDis* is the maximum value in the distance matrix (we use a Euclidean metric to calculate distance). The matrices $\mathbf{MAT}_A$ and $\mathbf{MAT}_B$ contain the basic parameters of partitions P_A and P_B, respectively. $\mathbf{LMAT}_t$ is the similarity matrix of $\mathbf{MAT}_A$ and $\mathbf{MAT}_B$, and Sim_t is its minimum value. By removing the row and column that contain Sim_t, we generate $\mathbf{LMAT}_{t+1}$. This procedure of removing a row and column should be repeated until **LMAT** has the size 1×1. Equation (6) explains the *Likeness* function:

$$Likeness = 1 - \left(\frac{1}{MaxDis}\sum_{t=0}^{n} Sim_t\right) \quad (5)$$

$\mathbf{LMAT}_0$ is an $n \times n$ matrix. In other words, *n* is the number of basic parameters in the algorithm, e.g., the number of clusters in Kmeans (because the basic parameters of Kmeans give a matrix of *k* seed points). The independence of each partition is calculated as follows:

$$Independence(P) = \frac{1}{M}\sum_{i=1}^{M} \mathrm{BPI}(P, P_i) \qquad (6)$$

where M refers to the number of members in the crowd. Thus, according to the above definitions, one of the conditions for entering the result of a clustering into the crowd is given by equation (8), which we call the independence condition:

$$Independence(C) \geq iT \qquad (7)$$

where $0 \leq iT \leq 1$ is the threshold value for independence.

3.3. Decentralization of Opinions

Surowiecki explains the necessary conditions for generating a wise crowd as follows (Surowiecki, 2004): “If you set a crowd of self-interested, independent people to <u>work in a decentralized way on the same problem, instead of trying to direct their efforts from the top down</u>, their collective solution is likely to be better than any other solution you could come up with.”

According to Surowiecki’s explanation of decentralization and his examples on the CIA and Linux, it can be inferred that decentralization is a quality metric. The WOC cluster should be implemented such that decentralization is established across all of its parts. According to the above, we define decentralization in a cluster ensemble as follows:

The primary algorithm must not be influenced by any mechanism or predefined parameters; in this way, each base algorithm has a chance to reveal a ‘very good result’ with its own customization and local knowledge.

In the above definition, a *very goodresult* is one that has good performance, as well as enough diversity and independence to be added to the crowd. Regarding this new definition, we should consider the following conditions when designing a cluster ensemble mechanism, in order to retain decentralization:

1- The number of primary algorithms participating in the crowd should be greater than one.

2- The threshold parameter cT, which we call the coefficient of decentralization, is a coefficient which is multiplied in the number of clusters. Every base algorithm clusters the dataset into at most $cT \times k$ clusters. i.e. it clusters the dataset into a number of clusters between cT to $cT \times k$.

3- The method of entering a primary algorithm into the crowd should ensure that the final results will not be affected by its errors. In other words, the decision making of the final ensemble should not be centralized.

From the above discussion, it is clear that decentralization checking should be performed during the generation of the base results. In other words, we try to satisfy the decentralization conditions in the first phase, while producing the base partitions. Therefore, unlike diversity and independence, there is no evaluation of decentralization during the assessment of the initial partitions.

3.4. Aggregation Mechanism

In this step, the opinions in the wise crowd are combined to reach a final consensus partition. In some of clustering method, the consensus partition uses a $n \times n$ co-association matrix that counts the number of groupings in the same cluster for all data points. In these methods, the primary clustering results are first used to construct the co-association matrix. The most prominent of these methods is EAC[4] (Fred and Jain, 2005). Each entry in the co-association matrix is computed as:

$$C(i,j) = \frac{n_{i,j}}{m_{i,j}} \quad (8)$$

where $m_{i,j}$ is the number of partitions in which this pair of objects are simultaneously present and $n_{i,j}$ counts the number of clusters shared by objects with indices *i* and *j*. WOCCE uses the co-association matrix to aggregate the results, then employs the Average-Linkage algorithm to derive the final partition from the matrix.

3.5. Summing up

In WOCCE, the process starts with an evaluation of the diversity and independence of the partitions which it is shown in Figure (1). As stated earlier, the necessary decentralization conditions are satisfied in the cluster generation phase by constructing the base partitions. Therefore, there is no component for assessing the decentralization of the generated partitions. In the WOCCE

[4]Evidence Accumulation Clustering

framework, only the decentralized partitions that pass both the independence and diversity filters are eligible to join the wise crowd.

In summary, the differences between these two approaches are:

1- The method of evaluating the clustering algorithm. In the WOCCE approach, the diversity and independence of each primary algorithm is compared with other algorithms in the crowd after execution. If they have the necessary conditions, they are added to the crowd.

2- Most importantly, in the WOCCE approach, each primary clustering can be inserted into the crowd without affecting other algorithms' results. This approach can detect errors and identify information in the results (by checking the diversity and independence values), and then compensate for these errors with true information from all the results in the crowd, guaranteeing that the errors will not be spread to other members (by changing the total diversity and independence values in each step).

3- In the WOCCE approach, the selection and measurement of independence and diversity are performed in one step. This causes the independence and diversity values to be retained in the final ensemble.

Figure (3) shows the pseudo code for the WOCCE procedure:

```
Function WOCCE (Dataset, Kb, iT, dT, cT) Return [Result, nCrowd]
nCrowd = 0;
While we have base cluster
   [idx, Basic-Parameter] = Generate-Basic-Algorithm (Dataset, Kb*cT)
   If (Independent (Basic-Parameter) > iT)
      If (Diversity (idx) >dT)
         Insert idx to Crowd-Partitions
         Crowd = Crowd + 1
      End if
   End if
End while
Co-Acc = Make-Correlation-Matrix (Crowd-Partition)
Z = Average-Linkage (Co-Acc)
Result = Cluster (Z, Kb)
```

Fig. 3. Pseudo code for the WOCCE framework

In Figure (3), *Kb* is the number of clusters given by the base algorithm. The *Generate-Basic-Algorithm* function builds the partitions of base clusters (primary results), *Make-Correlation-Matrix* builds the co-association matrix according to equation (8), and the *Linkage* and *Cluster* functions build the final ensemble in accordance with the Average Linkage method. The parameter *Result* is our final ensemble, and *nCrowd* is the number of members in the crowd.

Cluster ensemble selection for selecting optimizes partitions in the final result does the auditing of initial result of partitions usually by choosing one or many consensus metrics. In this method there might be two major problems:

Firstly, although the final result is always in accordance with the selected metrics, providing the optimized result, there could be other metrics by which the best final result can be generated. Secondly, since in this auditing method only the first results are analyzed (including the correct data as well as errors), it is possible that all aspects and specifications of data are not considered or since for precise auditing. Thus, it is necessary for more attention to be spent on other contractive entities in any cluster ensemble algorithm.

Unlike other available methods wise clustering is a structural perspective for generating the best result based on all aspects and specifications of data which operates in relation to "The Wisdom of Crowds" theory. In this method all needed information from clustering problems is gathered by controlling all entities within cluster ensemble as the result errors in each entity is optimized by information from other entities which consequently reduces the possibility of occurrence of errors in complex data dramatically.

In WOCCE, for the first time, the issue of independency algorithms is incorporated. This technique redundant repeated results in a particular algorithm and ensure that the similar results created by algorithms with acceptable degree of independency. For this reason in WOCCE the number of selected initial results is much less than that in other methods. Unlike the cluster ensemble, the framework of WOCCE always includes the following four main conditions: Independency of algorithms, diversity of initial (basic) results, decentralization of framework's structure for attending the quality of final results and method of feedback combination for safeguarding the auditing results of partitions in the wise crowd (initial results for combination). This structure makes WOCCE a flexible technique and capable of being programed, so that by altering the value of thresholds, It can be adjusted for any data which will be discus in section 5.3.Table (1) presents a brief mapping between terminologies in WOC and the corresponding cluster ensemble area.

Table1. Mapping between WOC and cluster ensemble terminologies

WOC Terminology	Cluster Ensemble Terminology
Primary opinion	Primary partition
People	Base algorithm
Wise crowd	Primary clustering results
Diversity of Opinion	Diversity of primary clustering results
Opinion independence	Independence of clustering algorithms that generate primary partitions
Decentralization	Decentralization in cluster generation

4. Evaluation

4.1. Datasets

This section describes a series of empirical studies and reports their results. The proposed method is examined over 11 different standard datasets. We have chosen datasets that are as diverse as possible in their numbers of true classes, features, and samples, as this variety better validates the results obtained. Brief information about these datasets is listed in Table (2). More information is available in Newman et al. (1998) and Alizadeh et al. (2012).

Table2. Information about the datasets used in our simulations

Name	Feature	Class	Sample
Half Ring	2	2	400
Iris	4	3	150
Balance Scale*	4	3	625
Breast Cancer*	9	2	683
Bupa*	6	2	345
Galaxy*	4	7	323
Glass*	9	6	214
Ionosphere*	34	2	351
SA Heart*	9	2	462
Wine*	13	2	178
Yeast*	8	10	1484
Pendigits[5]	16	10	10992
Statlog	36	7	6435
Optdigits[6]	62	10	5620

[5] Pen-Based Recognition of Handwritten Digits Data Set
[6] Optical Recognition of Handwritten Digits Data Set

The features of the datasets marked with an asterisk are normalized to a mean of 0 and variance of 1, i.e., N(0, 1).

4.2. Experimental Method for Calculating Thresholds

This paper proposes an experimental method for determining the threshold values *iT*, *dT*, and *cT*. First, we check the relationships between the thresholds and WOCCE factors:

- *iT* has a relation with the number of base clustering algorithms, the variety of base clustering algorithm types, and the runtime of WOCCE.
- *dT* has a relation with the variety of base clustering algorithm types, the decentralization threshold (*cT*), and the number of partitions in the crowd.
- *cT* has a relation with the number of data in the dataset, the number of features in the dataset, and the number of partitions in the clustering.

In this paper, *cT*is chosen based on the number of data,and *iT* and *dT*are chosen such that each WOCCE algorithm's runtime is approximately 30 min on a PC with certain specifications[7].

4.3. Results

We used MATLAB R2012a (7.14.0.739) in order to generateour experimental results. The algorithms in Table (3)were used to generate the wise crowd.

Table3. List of base algorithms used in WOCCE

No.	Algorithm Name
1	K-Means
2	Fuzzy C-Means
3	Median K-Flats
4	Gaussian Mixture
5	Subtract Clustering
6	Single-Linkage Euclidean
7	Single-Linkage Hamming
8	Single-Linkage Cosine
9	Average-Linkage Euclidean
10	Average-Linkage Hamming

[7] CPU = Intel X9775 (4*3.2 GHz), RAM = 16 GB, OS = Windows Server 2012 RTM x64

11	Average-Linkage Cosine
12	Complete-Linkage Euclidean
13	Complete-Linkage Hamming
14	Complete-Linkage Cosine
15	Ward-Linkage Euclidean
16	Ward-Linkage Hamming
17	Ward-Linkage Cosine
18	Spectral clustering using a sparse similarity matrix
19	Spectral clustering using Nystrom method with orthogonalization
20	Spectral clustering using Nystrom method without orthogonalization

The Average-Linkage method was used to generate the final dendrogram. The distances were measured by a Euclidean metric. The value of *cT* was determined according to the dataset, number of samples, and runtime. All results are reported as the average of 10 independent runs of the algorithm. The final clustering performance was evaluated by re-labeling between obtained clusters and the ground truth labels and then counting the percentage of correctly classified samples. The WOCCE results are compared with well-known base algorithms including K-means, Fuzzy C-means, Subtract Clustering, and Single-Linkage, as well as two state-of-the-art cluster ensemble methods (EAC and MAX). Table (4) shows the results.

Table4. Experimental results

	Primary methods					*Cluster Ensemble methods*							
	Kmeans	**FCM**	**Subtract Clustering**	**Single Linkage**	**EAC**	**MAX**	**CSPA**	**HGPA**	**MCLA**	**WOCCE**			
										iT	*dT*	*cT*	Result
Half Ring	75.75	78	86	75.75	77.17	78.48	74.5	50	74.5	0.2	0.06	3	**87.2**
Iris	65.3	82.66	55.3	68	**96**	72.89	85.34	48.66	89.34	0.2	0.06	1	**96**
Balance Scale	40.32	44	45.32	46.4	52	52.1	51.84	41.28	51.36	0.23	0.063	3	**54.88**
Breast Cancer	93.7	94.43	65	65.15	95.02	75.72	80.97	50.37	96.05	0.18	0.02	1	**96.92**
Bupa	54.49	50.1	**57.97**	57.68	55.18	56.17	56.23	50.72	55.36	0.21	0.04	3	57.42
Galaxy	30.03	34.98	29.72	25.07	31.95	32.78	29.41	31.27	28.48	0.2	0.05	2	**35.88**
Glass	42.05	47.19	36.44	36.44	45.93	44.17	38.78	41.12	51.4	0.19	0.06	3	**51.82**
Ionosphere	69.51	67.8	**77**	64.38	70.48	64.48	67.8	58.4	71.22	0.3	0.1	3	70.52
SA Heart	64.51	63.41	67.26	65.15	65.19	63.96	58.42	50.93	62.54	0.65	0.8	1	**68.7**
Yeast	31.19	29.98	31.2	31.73	31.74	32.4	14	15.23	17.56	0.5	0.5	1	**34.76**
Wine	65.73	**71.34**	67.23	37.64	70.56	69.17	67.41	62.36	70.22	0.2	0.05	3	**71.34**
Pendigits	46.97	36.77	10.4	10.46	10.47		58.32	11.14	58.62	0.02	0.12	1	**58.68**
Optdigit	52.52	38.33	47.72	10.28	20		75.21	64.77	77.15	0.01	0.1	1	**77.16**
Statlog	50.93	49.91	23.8	23.8	23.9		54.23	52.94	55.71	0.01	0.1	1	**55.77**

The best result obtained for each dataset is highlighted in bold. Even though WOCCE was outperformed in two datasets (Bupa and Ionosphere), the majority of the results demonstrate that the proposed method showed a superior performance. To compare WOCCE with its powerful ensemble rivals more accurately, Figure (4) and Figure (5) illustrates the number of misclassification errors made by ensemble methods and average of accuracy for each technique respectively.

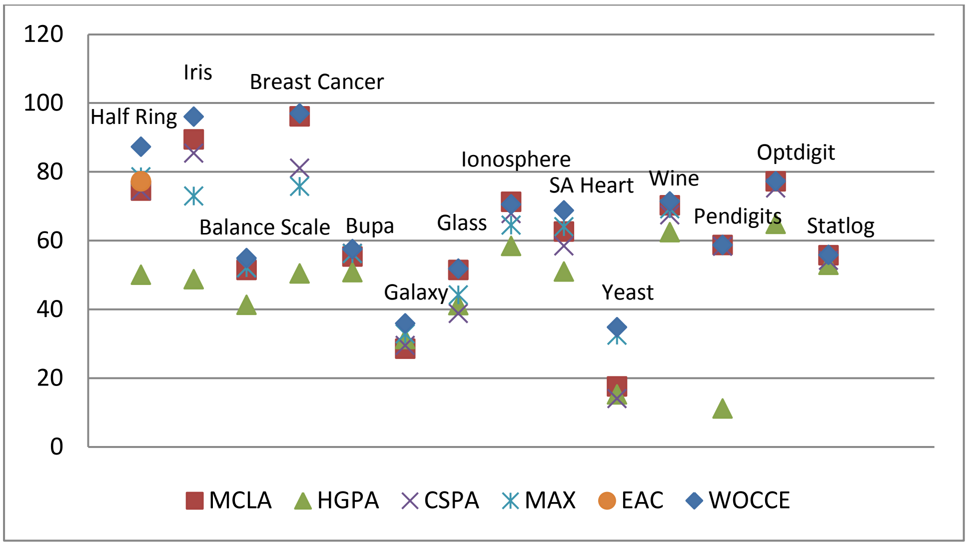

Fig. 4. Misclassification errors made by ensemble methods for each dataset

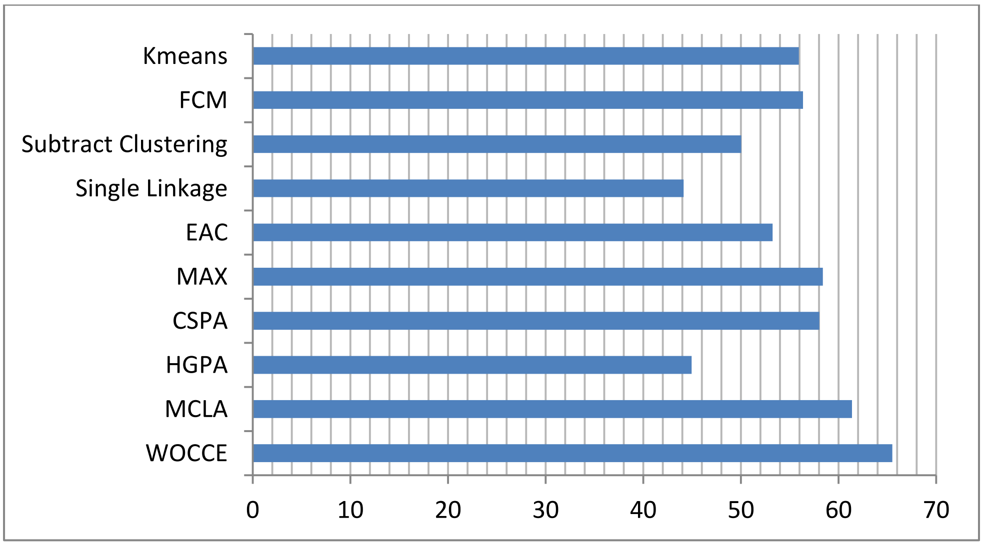

Fig. 5. Average of accuracy for each technique

From Figure (4), it is difficult to separate the WOCCE and EAC methods. However, the average performance over all eleven datasets reveals that WOCCE outperformed EAC by over 2%. Figure (6) illustrates the relationship between iT and the runtime of WOCCE. This experiment was performed with $dT = 0$ in order to remove the effect of diversity on the final results. The vertical axis refers to time and the horizontal axis refers to the independence threshold.

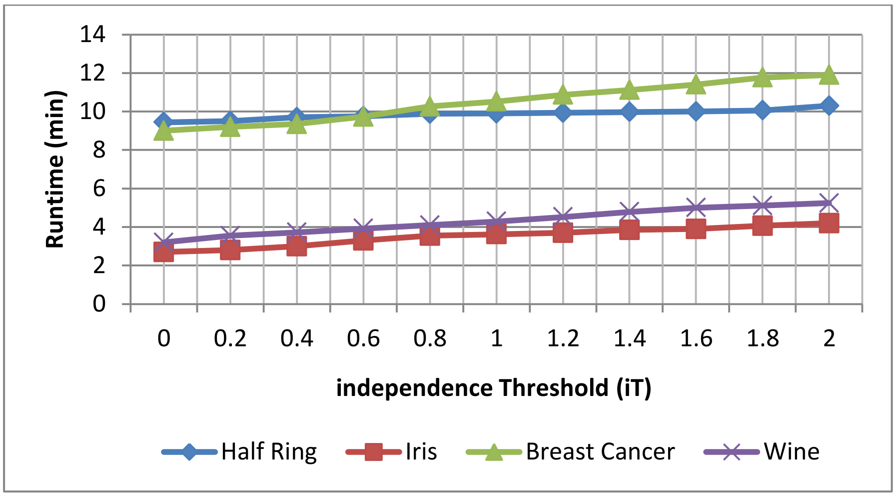

Fig. 6. Relation between Independence and Runtime

Figure (7) illustrates the relationship between dT and the runtime of WOCCE. This experiment was performed with $iT = 0$ in order to remove the effect of independence on the final results. The vertical axis refers to time and the horizontal axis refers to the diversity threshold.

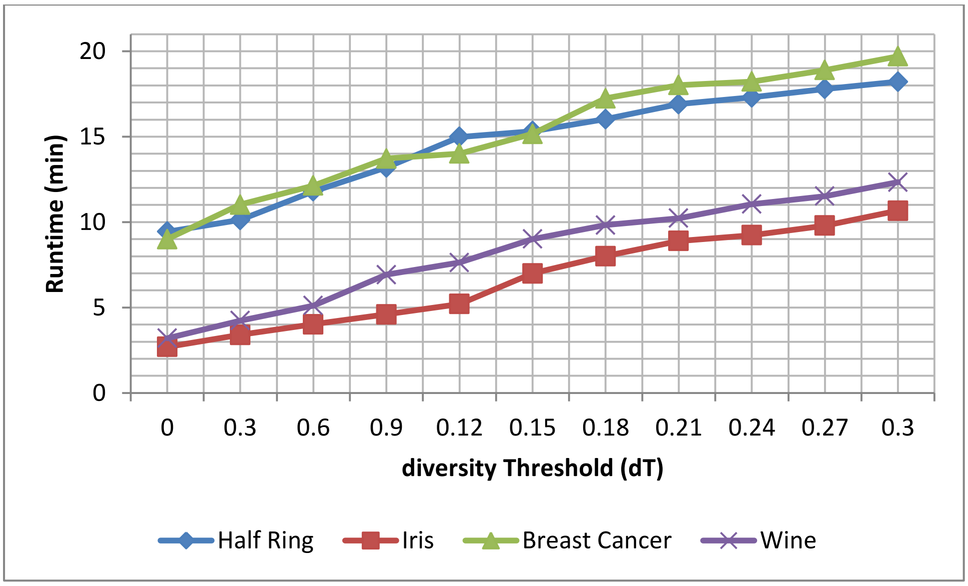

Fig. 7. Relation between Diversity and Runtime

Figures (8) and (9) illustrate the relationship between the performance of WOCCE, based on the number of correctly classified samples, and the independence and diversity thresholds, respectively. To plot Figure (8), a fixed value was assigned to *dT* in order to measure the effect of independence on the final results. The vertical axis refers to the performance and the horizontal axis refers to independence. Similarly for Figure (9), *iT* was fixed in order to plot performance with respect to diversity.

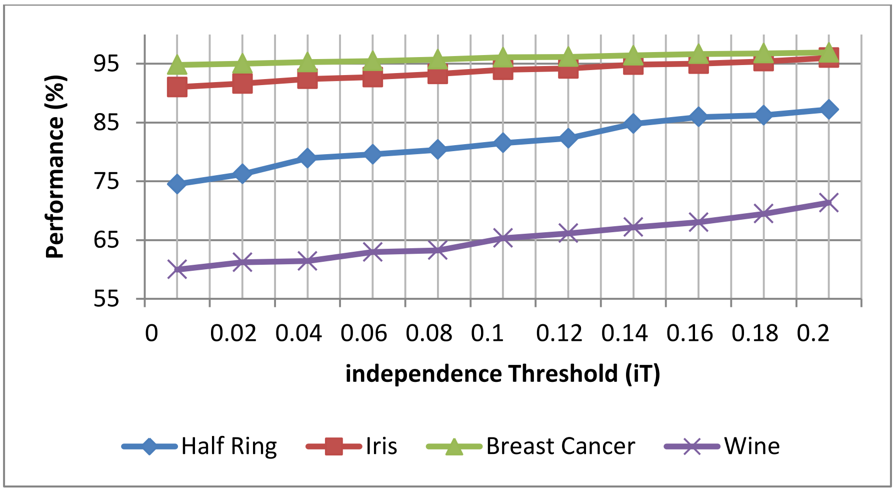

Fig. 8. Relation between Independence and Performance

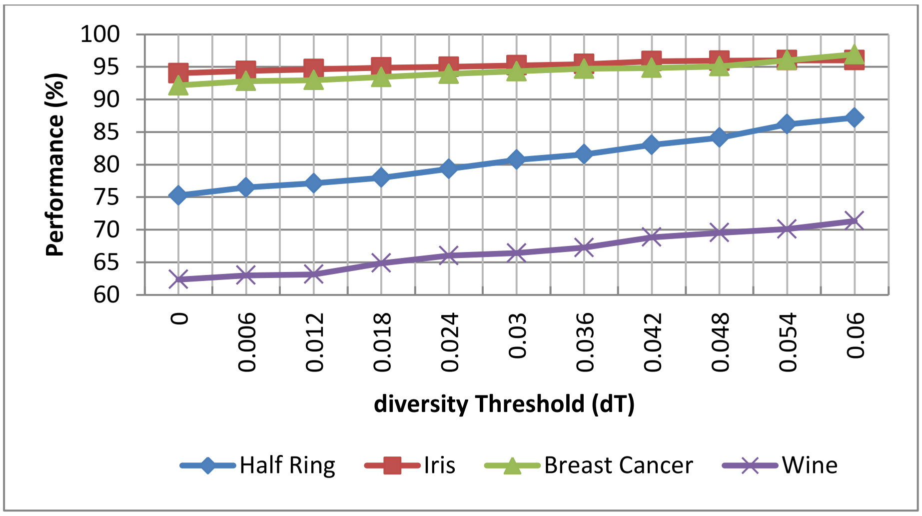

Fig. 9. Relation between Diversity and Performance

These figures show that the performance increased with the respective threshold value. They illustrate the effect of independence and diversity in the performance of our cluster ensemble, and confirm that, along with diversity, the independence is an important factor that should be considered when creating the ensemble.

Table (5) compares the attributes of WOCCE with some contemporary algorithms.

Table5. Comparison of attributes

Name	Selection Procedure	Diversity Check	Independence Check	Decentralization Check	Proof of the Optimized Solution	Complexity
(Fred and Jain, 2005)	NOT Supported	NOT Supported	NOT Supported	NOT Supported	NOT Supported	**Low**
(Fern and Lin, 2008)	**Support**	**NMI Criterion**	NOT Supported	NOT Supported	NOT Supported	Medium
(Singh et al., 2010)	NOT Supported	NOT Supported	NOT Supported	NOT Supported	Supported	High
(Alizadeh et al., 2011)	**Support**	**MAX criterion**	NOT Supported	NOT Supported	NOT Supported	**Low**
(Alizadeh et al., 2012)	**Support**	**APMM criterion**	NOT Supported	NOT Supported	NOT Supported	**Low**
WOCCE	**Support**	**A3**	**Supported**	**Supported**	NOT Supported	Medium

As stated in Table (5), the other combinational methods do not investigate the independence and decentralization when building the ensemble. Instead, most of them focus on the diversity of the primary partitions. WOCCE is the first system to date that adds these two conditions to the cluster ensemble field. Although we have not presented a mathematical proof to support our method, the experimental results confirm its superior performance with respect to other cluster ensemble methods for most of the benchmark datasets.

5. Conclusion

In this paper, the WOC phenomenon was mapped to the cluster ensemble problem. The primary advantage of this mapping was the addition of two new aspects, the independence and decentralization, as well as a new framework to investigate them. Until now, common cluster ensemble methods have concentrated on the diversity of the primary partitions. Inspired by the WOC research in the social sciences, this paper introduced the conditions of independence and decentralization to the field of cluster ensemble research. The proposed WOCCE framework uses a feedback procedure to investigate all three conditions, yielding a wise crowd incorporating decentralization, independence, and diversity.

Similar to other pioneering ideas, the WOCCE framework can be improved later.

This paper suggested employing as different as base algorithm to satisfy the decentralization condition. We also proposed a procedure to assess the independence of the base algorithms, and introduced the A3 criterion to measure the diversity of the partitions. Our suggestions for satisfying the corresponding conditions will be investigated further in future work. The main drawback of the WOCCE algorithm is that it has three threshold parameters that must be set to appropriate values. This parameterization can be considered as another area for future work.

References

Alizadeh H., Parvin H. and Parvin S., (2012)."A Framework for Cluster Ensemble Based on a Max Metric as Cluster Evaluator".International Journal of Computer Science (IAENG), pp.1-39.

Alizadeh, H., Minae, i.M. and Parvin, H., (2011)."A New Asymmetric Criterion for Cluster Validation.Springer-Verlag Berlin Heidelberg", pp.320-30.

Ayad H.G. and Kamel M.S., (2008)."Cumulative Voting Consensus Method for Partitions with a Variable Number of Clusters". IEEE Trans. On Pattern Analysis and Machine Intelligence, 30(1), pp.160-73.

Baker, L. and Ellison D., (2008)."The wisdom of crowds — ensembles and modules in environmental modeling".Geoderma, 147, pp.1-7.

Fern X. and Lin W. (2008), "Cluster Ensemble Selection".SIAM International Conference on Data Mining (SDM08).

Fred A. and Lourenco A., (2008). Cluster Ensemble Methods: from Single Clusterings to Combined Solutions. Studies in Computational Intelligence (SCI), 126, pp.3-30.

Fred, A. and Jain, A.K., (2005). "Combining Multiple Clusterings Using Evidence Accumulation". IEEE Trans. on Pattern Analysis and Machine Intelligence, 27(6):835–850.

Grofman, B. and Owen, G., (1996). "Information Pooling and Group Decision Making". Proceedings of the Second University of California, Irvine Conference on Political Economy. JAI Press, Inc., Greenwich, Connecticut.

Hadzikadic M. and Sun M., (2010)."Wisdom of Crowds in the Prisoner's Dilemma Context". Presented at Advances in Machine Learning II, pp.110-18.

Jain A., Murty M. N. and Flynn P., (1999). "Data clustering: A review". ACM Computing Surveys.31(3), pp.264-323.

Mackey C., (1841). "Extraordinary Popular Delusions and the Madness of Crowds". ISBN: 978-1897597323.

Miller, B., Hemmer, P., Steyvers, M. and Lee, M. D., (2009)."The Wisdom of Crowds in Rank Ordering Problems".International Conference on Computing and Mission.

Newman, C.B.D.J., Hettich, S. and Merz, C., (1998).UCI repository of machine learning databases. [Online] Available at: http://www.ics.uci.edu/~mlearn/MLSummary.html.

Page, S.E., (2007). "The difference: How the power of diversity creates better groups, firms, schools, and societies". Princeton University Press.

Perrone, M.P. and Cooper, L.N., (1993). "When networks disagree: Ensemble method for neural".

Steyvers, M., Lee, M., Miller, B. and Hemmer, P., (2009)."The Wisdom of Crowds in the Recollection of Order Information".Advances in Neural Information Processing Systems, 22, pp.1785-93.

Strehl A. and Ghosh J., (2002). "Cluster ensembles - a knowledge reuse framework for combining multiple partitions". Journal of Machine Learning Research, 3, pp.583-617.

Surowiecki, J., (2004). "The Wisdom of Crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations". Little, Brown. ISBN: 0-316-86173-1.

Topchy, A., Jain, A.K. & Punch, W.F., (2003)."Combining Multiple Weak Clusterings".Proc. 3d IEEE Intl. Conf. on Data Mining, pp.331-38.

Tumer, K. &Ghosh, J., (1996)."Error correlation and error reduction in ensemble classifiers Connection Science". Special issue on combining artificial neural networks: ensemble approaches, 8(3&4), pp.385-402.

Welinder, P., Branson, S., Belongie, S. &Perona, P., (2010)."The Multidimensional Wisdom of Crowds".Conference on Neural Information Processing Systems (NIPS).

Williams, D. P., (2010)."Underwater Mine Classification with Imperfect Labels".International Conference on Pattern Recognition.

Yi, S. K. M., Steyvers, M. & Lee M. D., (2010). "Wisdom of the Crowds in Minimum Spanning Tree Problems". In Proceedings of the 32nd Annual Conference of the Cognitive Science Society