A Lattice-Based Consensus Clustering Algorithm

Artem Bocharov, Dmitry Gnatyshak, Dmitry I. Ignatov, Boris G. Mirkin, and Andrey Shestakov

National Research University Higher School of Economics
dignatov@hse.ru
http://www.hse.ru

Abstract. We propose a new FCA-based algorithm for consensus clustering, FCA-Consensus. As the input the algorithm takes T partitions of a certain set of objects obtained by k-means algorithm after its T different executions. The resulting consensus partition is extracted from an antichain of the concept lattice built on a formal context $objects \times classes$, where the classes are the set of all cluster labels from each initial k-means partition. We compare the results of the proposed algorithm in terms of ARI measure with the state-of-the-art algorithms on synthetic datasets. Under certain conditions, the best ARI values are demonstrated by FCA-Consensus.

Keywords: consensus clustering, k-means, Formal Concept Analysis, ensemble clustering, lattice-based clustering

1 Introduction and related work

It seems, consensus clustering approach became popular on the international scene after the paper of A. Strehl and J. Ghosh [1]. Since then consensus clustering is used in such areas as bioinformatics, web-document clustering and categorical data analysis.

As the input the consensus clustering approach usually takes T partitions of a certain set of objects obtained, for example, by k-means algorithm after its T different executions with possibly different k. The resulting consensus partition is build from the matrix $objects \times classes$, where the classes are the set of all cluster labels from each initial k-means partition. Thus, the main goal of consensus clustering is to find (recover) an optimal partition, i.e. to guess the proper number of resulting clusters and put the objects into each block of partition correctly. To evaluate the proposed approach researchers usually hypothesise that if a particular consensus clustering approach is able to guess a proper k and attain high accuracy on labeled datasets, then it can be used in pure unsupervised setting. This task is worth consideration mainly due to two reasons: We do not know a proper k in advance, and k-means is unstable due to randomness of initialisation [2]. However, we can use right guesses of each of the ensemble algorithms to build (recover) a proper partition.

© Marianne Huchard, Sergei O. Kuznetsov (Eds.): CLA 2016, pp. 45–56,
ISBN 978-5-600-01454-1, National Research University Higher School of Economics, 2016.

In [3], consensus clustering algorithms are classified in three main groups: probabilistic approaches [4,5]; direct approaches [1,6,7,8], and pairwise similarity-based approaches [9,10]. In the last category of methods, the (i, j)-th entry a_{ij} of the consensus matrix $A = (a_{ij})$ shows the number of partitions in which objects g_i and g_j belong to the same cluster.

In the previous papers [11,12], a least-squares consensus clustering approach was invoked from the paper [13], to equip it with a more recent clustering procedure for consensus clustering and compare the results on synthetic data of Gaussian clusters with those by the more recent methods. Here, our main goal is to propose a lattice-based consensus clustering algorithm by means of FCA and show its competitive applicability. To the best of our knowledge, a variant of FCA-based consensus approach was firstly proposed to cluster genes into disjoint sets [14]. For those, who are interested theoretical properties of different consensus procedures and its relationship with FCA we could recommend [15].

The paper is organised in five sections. In Section 2, we refresh some definitions from FCA, introduce partitions and their lattice, and prove that any partition lattice can be easily mapped to a concept lattice. In Section 3, we introduce our modification of Close-by-One algorithm for consensus clustering. In Section 4, we provide our experimental results with synthetic data both for individual behaviour of FCA-Consensus and its comparison with the state-of-the-art existing methods. Section 5 concludes the paper and outlines prospective ways of research and developments.

2 Basic definitions

First, we recall several notions related to lattices and partitions.

Definition 1. *A partition of a nonempty set A is a set of its subsets $\sigma = \{B \mid B \subseteq A\}$ such that $\bigcup_{B \in \sigma} B = A$ and $B \cap C = \emptyset$ for all $B, C \in \sigma$. Every element of σ is called* block.

Definition 2. *A partition lattice of set A is an ordered set $(Part(A), \vee, \wedge)$ where $Part(A)$ is a set of all possible partitions of A and for all partitions σ and ρ supremum and infimum are defined as follows:*

$$\sigma \vee \rho = \{N_\rho(B) \cup \bigcup_{C \in N_\rho(B)} N_\sigma(C) | B \in \sigma\},$$

$$\sigma \wedge \rho = \{B \cap C \mid \text{ for all } B \in \sigma \text{ and } C \in \rho\}, \text{where}$$

$N_\rho(B) = \{C \mid B \in \sigma, C \in \rho$ and $B \cap C \neq \emptyset\}$ and $N_\sigma(C) = \{B \mid B \in \sigma, C \in \rho$ and $B \cap C \neq \emptyset\}$.

Definition 3. *Let A be a set and let $\rho, \sigma \in Part(A)$. The partition ρ is finer than the partition σ if every block B of σ is a union of blocks of ρ, that is $\rho \leq \sigma$.*

Equivalently one can use traditional connection between supremum, infimum and partial order in the lattice: $\rho \leq \sigma$ iff $\rho \vee \sigma = \sigma$ ($\rho \wedge \sigma = \rho$).

Now, we recall some basic notions of Formal Concept Analysis (FCA) [16]. Let G and M be sets, called the set of objects and attributes, respectively, and let I be a relation $I \subseteq G \times M$: for $g \in G$, $m \in M$, gIm holds iff the object g has the attribute m. The triple $\mathbb{K} = (G, M, I)$ is called a *(formal) context*. If $A \subseteq G$, $B \subseteq M$ are arbitrary subsets, then the *Galois connection* is given by the following *derivation operators*:

$$\begin{aligned} A' &= \{m \in M \mid gIm \text{ for all } g \in A\}, \\ B' &= \{g \in G \mid gIm \text{ for all } m \in B\}. \end{aligned} \tag{1}$$

The pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$ is called a *(formal) concept (of the context K)* with *extent* A and *intent* B (in this case we have also $A'' = A$ and $B'' = B$).

The concepts, ordered by $(A_1, B_1) \geq (A_2, B_2) \iff A_1 \supseteq A_2$ form a complete lattice, called *the concept lattice* $\underline{\mathfrak{B}}(G, M, I)$.

Theorem 1. *(Ganter&Wille [16]) For a given partially ordered set $\mathfrak{P} = (P, \leq)$ the concept lattice of the formal context $\mathbb{K} = (J(P), M(P), \leq)$ is isomorphic to the Dedekind–MacNeille completion of $\mathfrak{P}$, where J(P) and M(P) are set of join-irreducible and meet-irreducible elements of $\mathfrak{P}$.*

Theorem 2. *(this paper) For a given partition lattice $\mathfrak{L} = (Part(A), \vee, \wedge)$ there exist a formal context $\mathbb{K} = (P_2, A_2, I)$, where $P_2 = \{\{a, b\} \mid a, b \in A$ and $a \neq b\}$, $A_2 = \{\sigma \mid \sigma \in Part(A)$ and $|\sigma| = 2\}$ and $\{a, b\}I\sigma$ when a and b belong to the same block of σ. The concept lattice $\underline{\mathfrak{B}}(P_2, A_2, I)$ is isomorphic to the initial lattice $(Part(A), \vee, \wedge)$.*

Proof. According to Theorem 1 the concept lattice of the context $\mathbb{K}_{DM} = (J(\mathfrak{L}), M(\mathfrak{L}), \leq)$ is isomorphic to the Dedekind–McNeille completion of $\mathfrak{L}$. The Dedekind–McNeille completion of a lattice is its isomorphic lattice by the definition (as a minimal completion which forms a lattice). So, we have to show that contexts $\mathbb{K}$ and $\mathbb{K}_{DM}$ (or their concept lattices) are isomorphic.

E.g., from [17] (Lemma 1, Chapter 4, Partition Lattices), we have that the atoms of a partition lattice are those its partitions which have only one block of two elements, the rest are singletons, and its coatoms are partitions into two blocks.

It is evident that all the atoms are meet-irreducible and all the coatoms are join-irreducible and that there are no other irreducible elements of the partition lattice $\mathfrak{L}$.

Let σ and ρ be two partitions from $\mathfrak{L}$, $\sigma \in J(\mathfrak{L})$ and $\rho \in M(\mathfrak{L})$, and $\sigma \leq \rho$. It means that all blocks of σ are subsets of blocks of ρ and the non-trivial block $\{i, j\} \in \sigma$ is a subset of one of the blocks of ρ. Note that A_2 coincides with the coatom set. It directly implies that $\{i, j\}I\rho$ iff an atom σ with block $\{i, j\}$ is finer than a coatom ρ.□

In addition we can show the correspondence between elements of $\mathfrak{L} = (Part(A), \vee, \wedge)$ and formal concepts of $\underline{\mathfrak{B}}(P_2, A_2, I)$. Every $(A, B) \in \underline{\mathfrak{B}}(P_2, A_2, I)$ corresponds to $\sigma = \bigwedge B$ and every pair $\{i, j\}$ from A is in one of σ blocks, where $\sigma \in Part(A)$. Every $(A, B) \in \underline{\mathfrak{B}}_{DM}(J(\mathfrak{L}), M(\mathfrak{L}), \leq)$ corresponds to $\sigma = \bigwedge B = \bigvee A$.

Example 1. In Fig. 1, one can see the diagram of a concept lattice isomorphic to partition lattice of 4-element set.

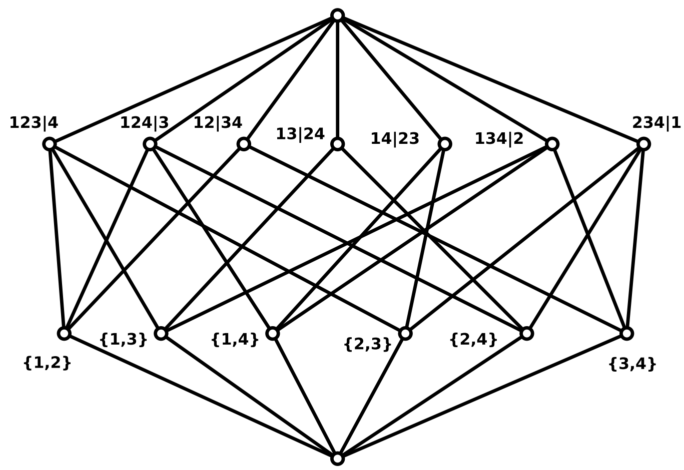

Fig. 1. The line diagram of a concept lattice isomorphic to the partition lattice of 4-element set (reduced labeling).

3 FCA-Consensus: close by object

To work in FCA terms we need to introduce a (formal) *partition context* that corresponds to the matrix X from the previous subsection. Let us consider such a context $\mathbb{K}_{\mathcal{R}} = (G, \sqcup M_t, I \subseteq G \times \sqcup M_t)$, where G is a set of objects, $t = 1, \ldots, T$, and each M_t consists of labels of all clusters in the t-th k-means partition from the ensemble. For example, gIm_{t1} means that object g has been clustered to the first cluster by t-th clustering algorithm in the ensemble.

Our FCA-Consensus algorithm looks for $\mathfrak{S}$, an antichain of concepts of $\mathbb{K}_{\mathcal{R}}$, such that for every (A, B) and (C, D) the condition $A \cap C = \emptyset$ is fulfilled. Here, the concept extent A corresponds to one of the resulting clusters, and its intent contains all labels of the ensemble members that voted for the objects from A being in one cluster. The input cluster sizes may vary, but it is a reasonable consensus hypothesis that at least $\lceil T/2 \rceil$ should vote for a set of objects to be in cluster.

One can prove a theorem below, where by *true partition* we mean the original partition into clusters to be recovered.

Theorem 3. *In the concept lattice of a partition context* $\mathbb{K}_{\mathcal{R}} = (G, \sqcup M_t, I \subseteq G \times \sqcup M_t)$*, there is the antichain of concepts* $\mathfrak{S}$ *such that all extents of its concepts* A_i *coincide with* S_i *from* σ*, the true partition, if and only if* $S_i'' = S_i$ *where* $i = 1, \ldots, |\sigma|$.

Proof. The proof is trivial by noting the fact that blocks of partitions are non-intersected and each block should be closed to form a concept extent.□

In fact, it happens if all ensemble algorithms has voted for all objects from S_i being in one concept (cluster). However, this is rather strong requirement and we should experimentally study good candidates for such an antichain.

The algorithm below works as Close by One (CbO) [18] adding objects one by one and checking a new canonicity conditions. Here it is modified in the following way: we need to stop adding objects to a particular concept in our candidate antichain $\mathfrak{S}$ until $|Y| \geq \lceil T/2 \rceil$, where Y is the intent of this current concept. Moreover, the covered objects at a particular step should not be added with any concept to the antichain $\mathfrak{S}$ further.

Algorithm 1: Main($(G, M, I), T$)

Input: a partition context (G, M, I) and the number of ensemble clusterers T
Output: $\mathfrak{S}$

$C = \emptyset$
for all $g \in G$ **do**
  if $g \notin C$ **then**
    $gpp = g''$
    $gp = g'$
    $\mathfrak{S}.enqueue(gpp, gp)$
    $C = C \cup gpp$
  end if
end for
return Process($(G, M, I), k, \mathfrak{S}$)

Thus, the resulting antichain $\mathfrak{S}$ may not cover all objects but we can add each non-covered object g to a concept $(A, B) \in \mathfrak{S}$ with maximal size of the intersection, $|B \cap g'|$. Traditionally, the algorithm consists of two parts, a wrapper procedure, Main, and a recursive procedure, Process.

4 Experimental results

All evaluations are done on synthetic datasets that have been generated using Matlab. Each of the datasets consists of 300 five-dimensional objects comprising three randomly generated spherical Gaussian clusters. The variance of each

Algorithm 2: Process$((G, M, I), T, \mathfrak{S})$

```
𝔗 = 𝔖
Cover = ∅ While 𝔗 ≠ ∅
𝔗.dequeue(A, B)
if A ∩ Cover = ∅ then
  Cover = Cover ∪ A
  𝔓.enqueue(A, B)
  for all g ∈ min(G \ Cover) do
    X = A ∪ {g}
    Y = X′
    if |Y| ≥ ⌈T/2⌉ then
      Z = Y′
      if {h|h ∈ Z \ X,  h < g} = ∅ then
        𝔓.dequeue(A, B)
        𝔓.enqueue(Z, Y)
        Cover = Cover ∪ Z
      end if
    end if
  end for
end if
if 𝔖 = 𝔓 then
  return 𝔓
end if
𝔖 = 𝔓
return Process((G, M, I), T, 𝔓)
```

cluster lies in $0.1 - 0.3$ and its center components are independently generated from the Gaussian distribution $\mathcal{N}(0, 0.7)$.
Let us denote thus generated partition as λ with k_λ clusters. The *profile* of partitions $\mathcal{R} = \{\rho^1, \rho^2, \ldots, \rho^T\}$ for consensus algorithms is constructed as a result of T runs of k-means clustering algorithm starting from random k centers.

We carry out the experiments in four settings:

1. Investigation of influence of the number of clusters $k_\lambda \in \{2, 3, 5, 9\}$ under various numbers of minimal votes (Fig. 2),
   a) two clusters case $k_\lambda = 2, k \in \{2, 3, 4, 5\}$,
   b) three clusters case $k_\lambda = 3, k \in \{2, 3\}$,
   c) five clusters case $k_\lambda = 5, k \in \{2, 5\}$,
   d) nine clusters case $k_\lambda = 9, k \in \{2, 3, 4, 5, 6, 7, 8, 9\}$;
2. Investigation of the numbers of clusters of ensemble clusterers with fixed number of true clusters $k_\lambda = 5$ (Fig. 3),
   a) $k = 2$,
   b) $k \in \{2, 3, 4, 5\}$,
   c) $k \in \{5\}$,
   d) $k \in \{5, 6, 7, 8, 9\}$
   e) $k = 9$;
3. Investigation of the number of objects $N \in \{100, 300, 500, 1000\}$ (Fig. 4);
4. Comparison with other state-of-the-art algorithms (Fig. 5–8),
   a) two clusters case $k_\lambda = 2, k \in \{2, 3, 4, 5\}$,
   b) three clusters case $k_\lambda = 3, k \in \{2, 3\}$,
   c) five clusters case $k_\lambda = 5, k \in \{2, 5\}$,
   d) nine clusters case $k_\lambda = 9, k \in \{2, 3, 4, 5, 6, 7, 8, 9\}$.

Each experiment encompasses 10 runs for every of 10 generated datasets. Such meta-parameters as the dimension number $p = 3$, the number of partitions (clusterers) in the ensemble $T = 100$, and the parameters of Gaussian distribution have been fixed for each experiment. After applying consensus algorithms, Adjusted Rand Index (ARI) [3] for the obtained consensus partition σ and the generated partition λ is computed as $ARI(\sigma, \lambda)$.

Given two partitions $\rho^a = \{R_1^a, \ldots, R_{k_a}^a\}$ and $\rho^b = \{R_1^b, \ldots, R_{k_b}^b\}$, where $N_h^a = |R_h^a|$ is the cardinality of R_h^a, $N_{hm} = |R_h^a \bigcap R_m^b|$, N is the number of objects, $C_a = \sum\limits_h \binom{N_h^a}{2} = \sum\limits_h \frac{N_h^a(N_h^a - 1)}{2}$.

$$ARI(\rho^a, \rho^b) = \frac{\sum\limits_{hm} \binom{N_{hm}}{2} - C_a C_b \Big/ \binom{N}{2}}{\frac{1}{2}(C_a + C_b) - C_a C_b \Big/ \binom{N}{2}} \tag{2}$$

This criterion expresses similarity of two partitions; its values vary from 0 to 1, where 1 means identical partitions, and 0 means totally different ones.

4.1 Comparing consensus algorithms

The lattice-based consensus results have been compared with the results of the following algorithms (Fig. 5–8):

– AddRemAdd ([19,11])
– Voting Scheme (Dimitriadou, Weingessel and Hornik, 2002) [6]
– cVote (Ayad, 2010) [7]
– Condorcet and Borda Consensus (Dominguez, Carrie and Pujol, 2008) [8]
– Meta-CLustering Algorithm (Strehl and Ghosh, 2002) [1]
– Hyper Graph Partitioning Algorithm [1]
– Cluster-based Similarity Partitioning Algorithm [1]

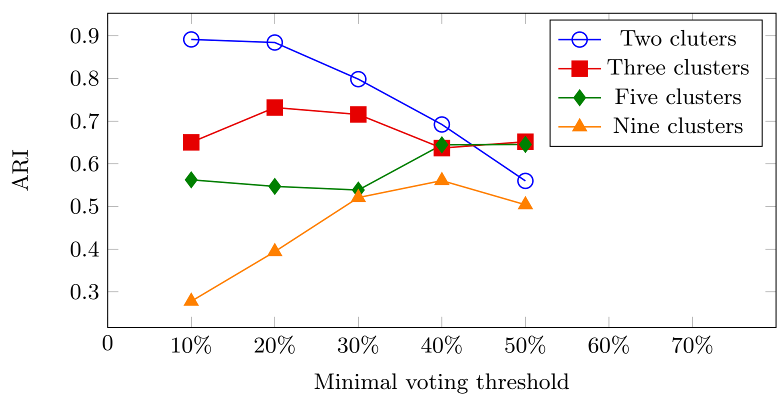

Fig. 2. Influence of minimal voting threshold to ARI for different number of true clusters

To provide the reader with more details we show the values of ARI graphically for each dataset out of ten used. The summarised conclusions are given in the next section.

5 Conclusion

Through experimentation we have draw the following conclusions:

– Optimal voting threshold in terms of minimal intent size for the resulting antichain of concepts is not constant; moreover, it is not usually a majority of votes of ensemble members (see Fig. 2).
– A rather expected conclusion: FCA-based consensus clustering method works better if set the number of blocks for the ensemble clusterers to be equal to the size of the original (true) partition (see Fig. 3).

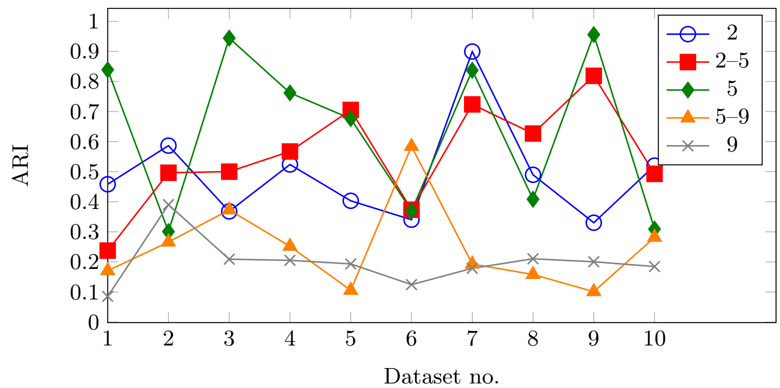

Fig. 3. Influence of minimal voting threshold to ARI for different numbers of clusters of the ensemble clusterers (each point is averaged over 10 datasets)

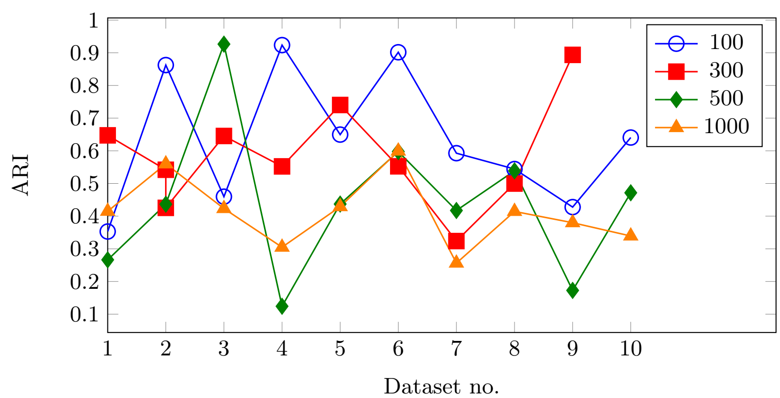

Fig. 4. Influence of different numbers of objects to ARI

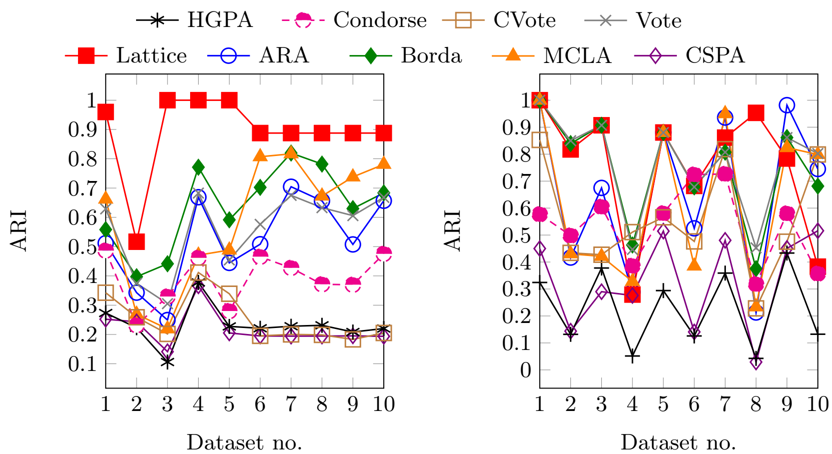

Fig. 5. Two clusters

Fig. 6. Three clusters

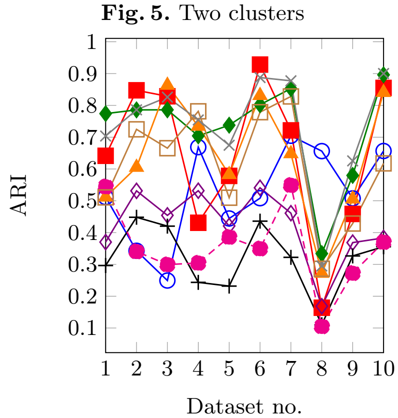

Fig. 7. Five clusters

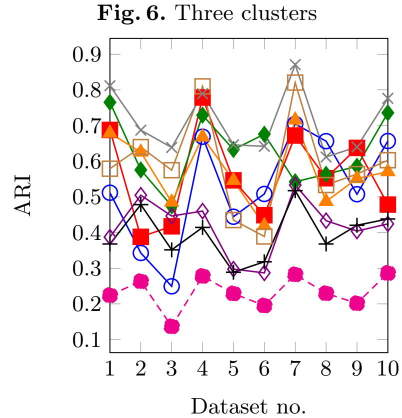

Fig. 8. Nine clusters

- ARI depends on the number of objects: The higher the number, the lower ARI (see Fig. 4).
- For two (and almost for all three) true clusters our method beats the other compared algorithms and in some cases consensus clustering task is solved with 100% accuracy (see Fig. 5–6).
- For larger number of clusters, our method is positioned as the median among the compared methods (see Fig. 7–8).

Thus, the first step on synthetic datasets has been done and we need to test the approach on real datasets. The used version of CbO can be modified for usage on the space of all partition labels for the cases when we have more objects than those labels. The algorithm complexity and time-efficiency should carefully studied and compared with those of the existing algorithms. An interesting venue is to use partition lattices as a search space to find an optimal partition. For example, one can build a pattern structure [20] over partitions similar to one in [21] and analyse the correlation of stability indicies [22] of the partitions as pattern concepts with ARI measure. By so doing it is possible to understand what are the good regions in the lattice for searching an optimal partition that can be built from existing ones via partition union and intersection operations.

Acknowledgments. We would like to thank Jaume Baixeries, Amedeo Napoli, Alexei Buzmakov, Mehdi Kaytoue, and Oleg Anshakov for their comments, remarks and explicit and implicit help during paper preparations. The paper was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of a subsidy by the Russian Academic Excellence Project "5-100".

References

1. Strehl, A., Ghosh, J.: Cluster ensembles — A knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research **3** (2002) 583–617
2. Kuncheva, L.I., Vetrov, D.P.: Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(11) (2006) 1798–1808
3. Ghosh, J., Acharya, A.: Cluster ensembles. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery **1**(4) (2011) 305–315
4. Topchy, A.P., Jain, A.K., Punch, W.F.: A mixture model for clustering ensembles. In: Proceedings of the Fourth SIAM International Conference on Data Mining, Lake Buena Vista, Florida, USA, April 22-24, 2004. (2004) 379–390
5. Wang, H., Shan, H., Banerjee, A.: Bayesian cluster ensembles. Statistical Analysis and Data Mining **4**(1) (2011) 54–70
6. Dimitriadou, E., Weingessel, A., Hornik, K.: A combination scheme for fuzzy clustering. IJPRAI **16**(7) (2002) 901–912
7. Ayad, H., Kamel, M.S.: On voting-based consensus of cluster ensembles. Pattern Recognition **43**(5) (2010) 1943–1953

8. Dominguez, X.S., Carrie, J.S., Pujol, F.A.: Fuzzy clusterers combination by positional voting for robust document clustering. Procesamiento del lenguaje natural **43** (2009) 245–253
9. Mirkin, B.: Clustering: A Data Recovery Approach. CRC Press (2012)
10. Guénoche, A.: Consensus of partitions : a constructive approach. Advances in Data Analysis and Classification **5**(3) (2011) 215–229
11. Mirkin, B.G., Shestakov, A.: Least square consensus clustering: Criteria, methods, experiments. In: Advances in Information Retrieval - 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, March 24-27, 2013. Proceedings. (2013) 764–767
12. Mirkin, B., Shestakov, A. In: A Note on the Effectiveness of the Least Squares Consensus Clustering. Springer New York, New York, NY (2014) 181–185
13. Mirkin, B., Muchnik, I.: Geometrical interpretation of clustering scoring functions. In: Methods for the Analysis of Multivariate Data in Economics, Novosibirsk, Nauka Publisher (1981) 3 – 11 (In Russian).
14. Hristoskova, A., Boeva, V., Tsiporkova, E.: A Formal Concept Analysis Approach to Consensus Clustering of Multi-Experiment Expression Data. BMC Bioinformatics **15** (2014) 151
15. Domenach, F., Tayari, A.: Implications of Axiomatic Consensus Properties. In: Algorithms from and for Nature and Life: Classification and Data Analysis. Springer International Publishing, Cham (2013) 59–67
16. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. 1st edn. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1999)
17. Graetzer, G.: General Lattice Theory. Akademie-Verlag (1978)
18. Kuznetsov, S.O.: A fast algorithm for computing all intersections of objects from an arbitrary semilattice. Nauchno-Tekhnicheskaya Informatsiya Seriya 2 - Informatsionnye protsessy i sistemy (1) (1993) 17–20
19. Mirkin, B.: Core concepts in Data Analysis: summarization, correlation, visualization. Springer (2011)
20. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings. (2001) 129–142
21. Codocedo, V., Napoli, A.: Lattice-based biclustering using partition pattern structures. In: ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014). (2014) 213–218
22. Kuznetsov, S.O.: Stability as an estimate of degree of substantiation of hypotheses derived on the basis of operational similarity. Nauchno-Tekhnicheskaya Informatsiya Seriya 2 - Informatsionnye protsessy i sistemy (12) (1990) 21–29