Multilayer clustering: A discovery experiment on country level trading data

Dragan Gamberger Matej Mihelčić Nada Lavrač

Friday 1st August, 2014

Abstract

The topic of this work is the presentation of a novel clustering methodology based on instance similarity in two or more attribute layers. The work is motivated by multi-view clustering and redescription mining algorithms. In our approach we do not construct descriptions of subsets of instances and we do not use conditional independence assumption of different views. We do bottom up merging of clusters only if it enables reduction of an example variability score for *all* layers. The score is defined as a two component sum of squared deviates of example similarity values. For a given set of instances, the similarity values are computed by execution of an artificially constructed supervised classification problem. As a final result we identify a small but coherent clusters. The methodology is illustrated on a real life discovery task aimed at identification of relevant subgroups of countries with similar trading characteristics in respect of the type of commodities they export.

1 Introduction

Clustering is an optimisation task which tries to construct subpopulations of instances so that distances between instances within each subpopulation are small while distances between instances in different subpopulations are as large as possible [1]. The main problem of clustering algorithms is to define an appropriate measure of distance between instances. It is well-known that different measures may result in identification of different clusters [2, 3].

The most common measure is Euclidean distance that is well defined for numerical attributes [1]. Nominal attributes can be handled only after some transformations. When dealing with numerical attributes it is necessary to normalize the data in the preprocessing step in order to ensure equal relevancy of all attributes regardless of their absolute values [1]. Results obtained by clustering are unreliable in the sense of the number of constructed clusters and in the sense of instances included in clusters.

A multi-view learning uses more than one set of attributes in order to improve quality of both supervised and unsupervised techniques [4, 5]. Redescription mining can be interpreted as a clustering approach in which the quality of the results is ensured by the condition that resulting clusters must have meaningful interpretations in independent attribute layers [6, 7]. In this work we present an approach to reliable clustering that reuses the basic ideas of multi-view clustering and redescription mining in a novel setting. We call it multilayer clustering because it has been originally developed for analysis of network data available in more than one layer [8]. In contrast to redescription mining, we do not construct descriptions of subsets of instances and in contrast to multi-view clustering we do not assume conditional independence of layers.

The first step is to determine the similarity of instances by executing a supervised machine learning task on an artificial problem in which the target set of instances are positive examples and negative examples are obtained by random shuffling of positive examples. We compute similarity tables for each attribute layer independently and then search for clusters that satisfy similarity conditions in *all* available layers. The main characteristic of the approach is that the resulting clusters are small but very coherent. Additionally, the methodology can be directly implemented on original attribute values without any transformation and normalization. When compared to redescription mining, results are less sensitive in respect of noise. The novel methodology is presented in Section 2. Its application is illustrated on a real world problem of recognizing groups of countries with similar economical profile based on export data for 106 different commodity types. It is a good example of clustering in a domain with a lot of noisy and imprecise data. Besides export data, we have a separate set of 105 attributes describing socio-economic characteristics of the countries. In this way we have a typical multi-view setting with two independent attribute layers for a fixed set of examples consisting of 155 countries. The obtained results are presented in Section 3.

2 Clustering related variability reduction algorithm

In machine learning we have a set of examples E that are described by a set of attributes A. Specifically in redescription mining, it is assumed that the set of attributes may be partitioned in at least two disjoint parts (layers). The partitioning is not random but a consequence of the meaning of the attributes or the way the data have been collected. For example, in a medical domain the first layer may contain anamnestic data (medical history of patients) while the second layer may contain laboratory measurements. In some other domain, different layers may contain the same attributes but collected in various time periods. The goal is to construct coherent clusters, that are as large as possible, in the complete attribute space.

2.1 Single layer clustering

Let us assume a basic clustering task in which we have only one layer of attributes. The approach consists of two steps. In the first step we compute the so called example similarity table. It is an N times N symmetric matrix, where N is the number of examples. All its values are in the range 0.0 - 1.0. A large value at a position i, j ($i \neq j$) denotes large similarity between examples i and j. In the second step we use the table in order to construct clusters.

Example similarity table (EST) computation We start from the original set of N examples represented by nominal and numerical attributes that may contain unknown values. The next step is to define an artificial classification problem so that the examples from the original set make positive examples while we artificially construct negative examples by shuffling values of the positive examples. Shuffling is done at the level of attributes so that we randomly mix values among examples. The values remain within the same attribute as in the original example. As a result, we have the same values in positive and negative examples but in negative examples we have randomized connections between attributes. Typically we construct 4 times more negative examples than positive examples.

Next, we use a supervised machine learning to build a predictive model for the discrimination between positive cases (original examples) and negative cases (examples with shuffled attribute values). The goal of learning is not the predictive model itself but information on similarity of examples. Machine learning approaches in which we can determine if some examples are classified in the same way are appropriate for this task. For example, in decision tree learning it means that examples end in the same leaf node while in covering rule set induction it means that examples are covered by the same rule. In order to estimate similarity between examples it is necessary to do a statistics over a potentially large set of classifiers. Additionally, a necessary condition for a good result is that classifiers are as diverse as possible and that each of them is better than random. All these conditions are satisfied by Random Forest [9] and Random Rules algorithms [10]. We use the latter approach in which we typically construct about 1500 rules for each EST computation.

Similarity of examples is determined so that for each pair of examples we count how many rules are true for both examples. The example similarity table presents the statistics for positive examples (original set of examples). A pair of similar examples will be covered by many rules while no rules or a very small number of rules will cover pairs that are very different in respect of their attribute values. Final EST values are obtained by the normalization of the determined counts by the largest detected value.

Table 1 presents an example of the similarity table for a set of 6 examples extracted from a real case with 155 examples. On the left side is the table with number of rules covering pairs of examples. Diagonal elements represent total number of rules covering each example. By the normalization of this table we obtain EST that is presented on the right side. It can be noticed that we have two very similar examples (examples 2 and 5), three similar (examples 1,3, and 4), and one very different example 6. The maximal value in the table on the left side is 97 and EST values (the table on the right side) are obtained by normalization with this value.

Clustering related variability (CRV) score

The second step in the process of clustering starts from the EST. The goal is to identify subsets of examples that can reduce variability of values in the EST. For this purpose we define a so called Clustering Related Variability (CRV) score. It is the basic measure which guides the search for iterative bottom up clustering. CRV score is not the other name for some type of example similarity measure. It is defined for a single example but so that the value depends on the examples it is clustered with. A cluster may consist of a single example.

Table 1: Example of an EST.

	ex1	ex2	ex3	ex4	ex5	ex6
ex1	38	0	27	28	0	7
ex2	0	97	3	1	97	3
ex3	27	3	47	16	3	1
ex4	28	1	16	45	1	4
ex5	0	97	3	1	97	3
ex6	7	3	1	4	3	39

	ex1	ex2	ex3	ex4	ex5	ex6
ex1	0.39	0.0	0.28	0.29	0.0	0.07
ex2	0.0	1.0	0.03	0.01	1.0	0.03
ex3	0.28	0.03	0.48	0.16	0.03	0.01
ex4	0.29	0.01	0.16	0.46	0.01	0.04
ex5	0.0	1.0	0.03	0.01	1.0	0.03
ex6	0.07	0.03	0.01	0.04	0.03	0.40

Clustering related variability for an element i contained in a cluster C is denoted by CRV_i. It is the sum of squared deviates of EST values in row i ($X_i = \{x_{i,j},\ j \in \{1, \ldots, N\}\}$) computed separately for examples that are within and outside cluster C. $CRV_i = CRV_{i,wc} + CRV_{i,oc}$.

Within cluster value $CRV_{i,wc} = \sum_{j \in C}(x_{i,j} - x_{mean,wc})^2$ is computed as a summation over columns j of row i corresponding to examples included in the same cluster with example i. In this expression $x_{mean,wc}$ is the mean value of all $x_{i,j}$ in the cluster. When example i is the only example in cluster C then $CRV_{i,wc} = 0$ because we compute the sum only for value $x_{i,i}$ and $x_{mean,wc} = x_{i,i}$.

Outside cluster value $CRV_{i,oc}$ is defined in the same way as $CRV_{i,wc}$ but for $x_{i,j}$ values of row i not included in cluster C. The used $x_{mean,oc}$ is the mean value of the EST element values not included in the cluster and it is different from the $x_{mean,wc}$ used to compute $CRV_{i,wc}$. When example i is the only example in a cluster then $CRV_{i,oc}$ is the sum of squared deviates for all values in row i except $x_{i,i}$.

The final CRV value of a cluster C is the average sum of all the CRV values for the elements contained in the cluster. That is, $CRV_C = \frac{\sum_{i \in C} CRV_i}{|C|}$

Example of CRV computation

We will use the data from the EST, presented in Table 1, to compute the CRV value for the example (ex1) contained in the cluster C. In this demonstration we will concentrate on three main cases: when a cluster contains only example ex1, when ex1 is clustered with ex3, and finally when it is clustered both with ex3 and ex4. By visual inspection of EST we can immediately notice some similarity among examples $\{ex1, ex3, ex4\}$. The goal is to demonstrate the CRV value computation and to show how its value decreases when clusters contain similar examples.

If example $ex1$ is the only example in a cluster: $C = \{ex1\}$ then:

$CRV_{ex1,wc} = (0.39 - 0.39)^2 = 0$

$CRV_{ex1,oc} = (0.0 - 0.13)^2 + (0.28 - 0.13)^2 + (0.29 - 0.13)^2 + (0.0 - 0.13)^2 + (0.07 - 0.13)^2 = 0.08$

$CRV_{ex1} = 0.08$

When we add a new element (ex3) to this cluster: $C = \{ex1, ex3\}$

$CRV_{ex1,wc} = (0.39 - 0.34)^2 + (0.28 - 0.34)^2 = 0.01$

$CRV_{ex1,oc} = (0.0-0.09)^2+(0.29-0.09)^2+(0.0-0.09)^2+(0.07-0.09)^2 = 0.06$

$CRV_{ex1} = 0.07$

Finally, when we have: $C = \{ex1, ex3, ex4\}$

$CRV_{ex1,wc} = (0.39 - 0.32)^2 + (0.28 - 0.32)^2 + (0.29 - 0.32)^2 = 0.01$

$CRV_{ex1,oc} = (0.0 - 0.02)^2 + (0.0 - 0.02)^2 + (0.07 - 0.02)^2 = 0.00$

$CRV_{ex1} = 0.01$

Single layer algorithm

It is possible to define the following bottom up clustering algorithm that is based on the CRV score.

CRV score based single layer clustering

1) Each example is in its own cluster
2) Iteratively repeat steps 3-6
3) For each pair of clusters x,y compute
   - $CRVx$ (mean CRV_i for examples in cluster x)
   - $CRVy$ (mean CRV_i for examples in cluster y)
   - $CRVxy$ (mean CRV_i score in union of clusters x and y)
   - $DIFF = mean(CRVx, CRVy) - CRVxy$
4) Select pair of clusters x,y with maximal $DIFF$ value
5) If maximal DIFF is positive then merge clusters x and y

6) Else stop.

The algorithm has a property that at first most similar examples will be merged together. In this way it produces a hierarchy of clusters. It may be noticed that in contrast to most other clustering algorithms, it has a very well defined stopping criteria. The process stops when further merging cannot result in reduction of the example variability measured by the CRV score. It means that the algorithm automatically determines the optimal number of clusters and that some examples may stay unclustered (more precisely, they remain as clusters consisting of only one example).

2.2 Multilayer algorithm

The basic lesson learnt from redescription mining and multi-view clustering is that the reliability of clustering can be significantly improved by a requirement that the result should be confirmed in two or more attribute layers. The approach for clustering based on example similarity has been presented in the previous section for a single layer case. It can be easily extended to clustering in multilayer domains.

If we have more than one attribute layer then for each of them we compute the example similarity table independently. For each layer we have to construct its own artificial classification problem and execute the supervised learning process in order to determine similarity between examples. Regardless of the number and type of attributes in different layers, the tables will be always matrices of dimension N times N. The reason is that by definition we have the same set of N examples in all layers.

After the computation of similarity tables, we execute the second step of the clustering process. Conceptually it is identical to a single layer approach. The main difference is that merging of two clusters is possible only if there is variability reduction in all layers. For each possible pair of clusters we have to compute potential variability reduction for all attribute layers and to select the smallest value for this pair. If this minimal value is positive it means that merging of the clusters enables variability reduction in all layers. When there are more pairs with positive minimal value, we chose the pair with the largest minimal value and then we merge these clusters in the current iteration.

CRV score based multilayer clustering

1) Each example is in its own cluster
2) Iteratively repeat steps 3-8
3) For each pair of clusters x,y do
4) For each attribute layer do
$CRVx$ (mean CRV_i for examples in cluster x)
$CRVy$ (mean CRV_i for examples in cluster y)
$CRVxy$ (mean CRV_i score in union of clusters x and y)
$DIFF = mean(CRVx, CRVy) - CRVxy$
5) For the given pair x,y select minimal $DIFF$ for all layers
6) Select pair of clusters x,y with maximal $DIFF$ value
7) If maximal DIFF is positive then merge clusters x and y
8) Else stop.

When we do clustering in two or more layers we have a conjunction of necessary conditions for merging two clusters. A typical consequence is that resulting clusters are smaller than in the case of a single layer clustering. This is illustrated by the experiment presented in the next section.

3 Experimental data and results

Our experimental work was conducted on the trading data that are publicly available from UNCTAD [11]. This database contains information for each pair of countries about the value of trade for 106 different commodity types. We have selected 155 countries from the database with relatively small number of unknown values for the year 2012. For them we have computed the total export value for the 106 different commodities. Finally, for each country we normalized indicator values by the value of country's total export in the year 2012. The result is a table with 155 rows and 106 columns. All known values are in the range 0-100 representing the percentage of export that a country has in the respective commodity type. Primary commodities, food and live animals, meat and meat preparations, machinery and transport equipment are some examples of aggregated commodity types. Some of the commodity types overlap. The prepared data table is publicly available from http://lis.irb.hr/DS2014data/ accompanied with the complete list of countries and the list of commodities.

The discovery task is to identify relevant subgroups of countries with a similar export patterns. The results are potentially relevant for understanding global trends, for example, by comparing the current subgroups with those obtained from data in year 2000. Our work has been motivated by the necessity to analyse and predict partial interests of EU countries in respect of a potential free trade agreement with China.

Table 2: Three largest clusters from export data.

Cluster with 27 countries: Exporters of primary commodities
Gambia, Seychelles, Zambia, Burkina Faso, Guyana, Ethiopia, Mali, Paraguay, Malawi, Chile, DR Congo, Tajikistan, Afghanistan, Benin, Peru, Belize, Cote d'Ivoire, Mozambique, Guinea, Papua New Guinea, Ghana, Australia, Bolivia, Oman, Russian Federation, Kazakhstan, Romania
Cluster with 24 countries: Exporters of manufactured goods
Germany, Japan, Czech Republic, Slovakia, Italy, Slovenia, Austria, China-Taiwan, R. Korea, Hungary, Poland, Portugal, Turkey, Finland, Sweden, Bangladesh, Cambodia, Luxembourg, France, China, Thailand, USA, Mexico, United Kingdom
Cluster with 17 countries: Fuel exporters
Algeria, Libya, Nigeria, Iraq, Angola, Congo, Brunei, Azerbaijan, Aruba, Gabon, Venezuela, Yemen, Iran, Saudi Arabia, Kuwait, Qatar, Mongolia

Table 2 presents the three largest clusters constructed from the export data layer by the single layer methodology described in Section 2. We have given a name to each cluster based on the common properties of included countries that have been identified by a simple statistical analyses. The largest cluster includes 27 countries that are mainly primary commodity exporters. The other two clusters contain exporters of manufactured goods and fuel exporters. It can be recognized from the lists of countries included in these clusters that the algorithm has been successful in identification of similarities between countries. However, clusters also include some unexpected results such as: Australia, Russia, and Romania being in the cluster of primary commodity exporters together with Guyana and Ethiopia, Bangladesh and Cambodia being in the cluster with Germany and Japan, while Mongolia participates in the cluster of fuel exporters.

Table 3: Three largest clusters from socio-economic data.

Cluster with 22 countries: Rural and young population
Ethiopia, Malawi, Uganda, Rwanda, Papua New Guinea, Niger, Burkina Faso, Tanzania, Afghanistan, Kenya, Tajikistan, Mozambique, Yemen, Togo, Zambia, Zimbabwe, DR Congo, Guinea, Madagascar, Mali, Benin, Senegal
Cluster with 18 countries: Modest level of rural population
Estonia, Hungary, Ukraine, Latvia, Austria, Italy, Lithuania, Czech Republic, Germany, Bulgaria, Belarus, Cuba, Spain, Greece, Poland, Croatia, Portugal, Switzerland
Cluster with 16 countries: Urban population
Denmark, France, Sweden, Finland, Netherlands, New Zealand, Iceland, Uruguay, Japan, Belgium, Malta, Australia, Canada, Norway, UK, USA

One possible interpretation is that only export data is insufficient information for effective and very consistent clustering of countries. In order to increase the quality of the results we have prepared the second layer of attributes. It consists of 105 World Bank indicators [12] that describe socio-economic characteristics of countries in the year 2012. We have selected indicators from economic policy, health, agriculture, and gender sets of public World Bank data. Our goal has been to select the most representative indicators from each field. The additional criterion was to use only relative indicators, that do not need normalization, in order to be comparable between countries of different size. We present a small sample of selected indicators for better insight: "Life expectancy at birth", "Percentage of population ages 15-64", "Public health expenditure as percentage of gross domestic product", and "Central government debt as percentage of gross domestic product" etc. The constructed attributes are all numeric and there is a noticeable amount of missing values. The data set is prepared for the same set of 155 countries as in the UNCTAD dataset and it is publicly available from our web site.

Before using both layers, we will present the clustering result obtained with socio-economic data in Table 3. It is interesting because it demonstrates that a dominant socio-economic characteristic of a country is the ratio of rural and urban population. The result is not coherent and it happens that the cluster with moderate number of rural population includes countries like

Germany and Switzerland but also Cuba and Belarus. In the same way, the cluster of countries with high percentage of urban population includes USA and Norway together with Uruguay and Malta. From the methodological point of view this is not a bad result but constructed clusters are not very useful for our discovery task because they tend to group economically very different countries.

Next, we have merged the export and socio-economic data into a single layer consisting of 211 attributes. The result obtained by the single layer methodology on this data has been very similar to the result obtained only on export data. Again, the three largest clusters represent primary commodity exporters, manufactured goods exporters, and fuel exporters. The results are now more consistent, Mongolia is discarded from the fuel exporters cluster and Romania and Russia are not in the cluster of primary commodity exporters. However, Australia and Iceland have been included in this cluster!

Finally, we present the result obtained by the multilayer approach in Table 4. In this approach, export and socio-economic data have been treated as separate layers. At first glance it can be noticed that the constructed clusters are significantly smaller but more coherent. The largest cluster has 8 countries that can be described as a group of countries with rural population that export primary commodities. Their basic common characteristic is that more than 87% of their exports are primary commodities. For Mozambique it is aluminium, beryllium, and tantalum, Ghana exports gold and diamonds, Zambia copper, Mali exports gold and kaolin while Cote d'Ivoire is one important exporter of cocoa. Some other common characteristics of these countries are that they export a low amount of manufactured goods (less than 11%) and a low amount of other food staff excluding tea, coffee, cocoa and spices (less than 24.5%).

For our discovery task, much more relevant result is the identification of a group of four EU countries: Czech Republic, Germany, Austria and Italy. At first, it may be a bit surprising that these countries have been identified as a most coherent group of EU countries. Recognition of their common characteristics is not a simple task because it is a small cluster and each of these four countries share a lot of common characteristics with other developed economies, especially those in EU. A potential solution is a simple statistical comparison of properties with most similar examples *not included* in the cluster. In multilayer methodology most similar examples may be identified as those included in larger clusters constructed for single layers that contain examples from multilayer clusters. Figure 1 illustrates the relations for our

Table 4: Clusters detected by the multilayer approach in which export data and socio-economic data are in different layers.

Cluster 1 with 8 countries
Coted'Ivoire, Ghana, Guinea, Mozambique, Papua New Guinea, Mali, DR Congo, Zambia
Cluster 2 with 4 countries
Czech Republic, Germany, Austria, Italy
Cluster 3 with 3 countries
Congo, Iraq, Angola
Cluster 4 with 3 countries
Poland, Portugal, Hungary
Cluster 5 with 3 countries
Finland, Sweden, Japan
Cluster 6 with 2 countries
Kuwait, Qatar
Cluster 7 with 2 countries
Ethiopia, Malawi
Cluster 8 with 2 countries
Latvia, Lithuania

domain in which, for example, the cluster consisting of Czech Republic, Germany, Austria and Italy is a subset of the clusters of manufactured goods exporters (layer 1) and the cluster of countries with modest rural population (level 2). In this figure arrows denote superset/subset relation and numbers denote sizes of clusters. Clusters at the basic layers are identified by the given names representing dominant characteristic of included countries while the clusters obtained by the combination of layers are represented by lists of included countries.

By using this approach we have identified the following decisive characteristics for the cluster consisting of Czech Republic, Germany, Austria, and Italy: a) high export of medium-skill and technology-intensive manufactures, b) low export of primary commodities, precious stones and non-monetary gold, c) low but always present export of beverages and tobacco, d) very low percentage of young population, d) low market capitalization of companies relative to gross domestic product. Figure 2 presents distributions of these

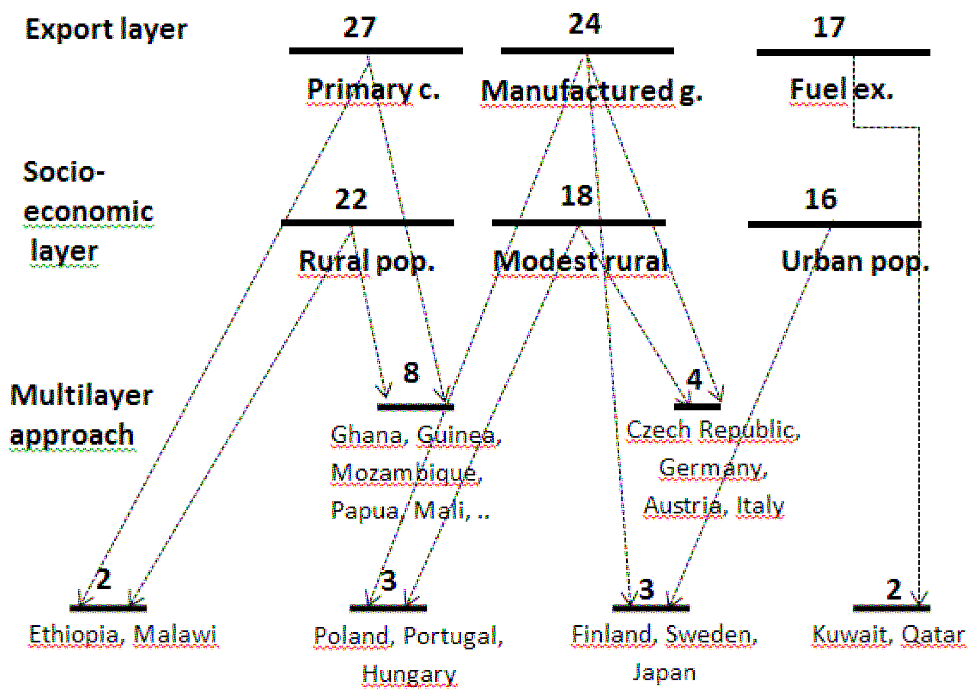

Figure 1: Approximative superset/subset relations among constructed clusters.

five characteristics in three different clusters: in the cluster of 24 countries representing manufactured goods exporters, cluster of 18 countries that have modest level of rural population, and finally for the target cluster consisting of four countries. This figure demonstrates that the resulting multilayer cluster has very narrow range of values for some relevant attributes. Furthermore, this fact is also true for some properties which do not occur in the supersets. In this way we identified that low percentage of young population and low market capitalization of companies as percentage of GDP are additional properties of this cluster of countries. Identification of these properties may present a potentially relevant discovery result.

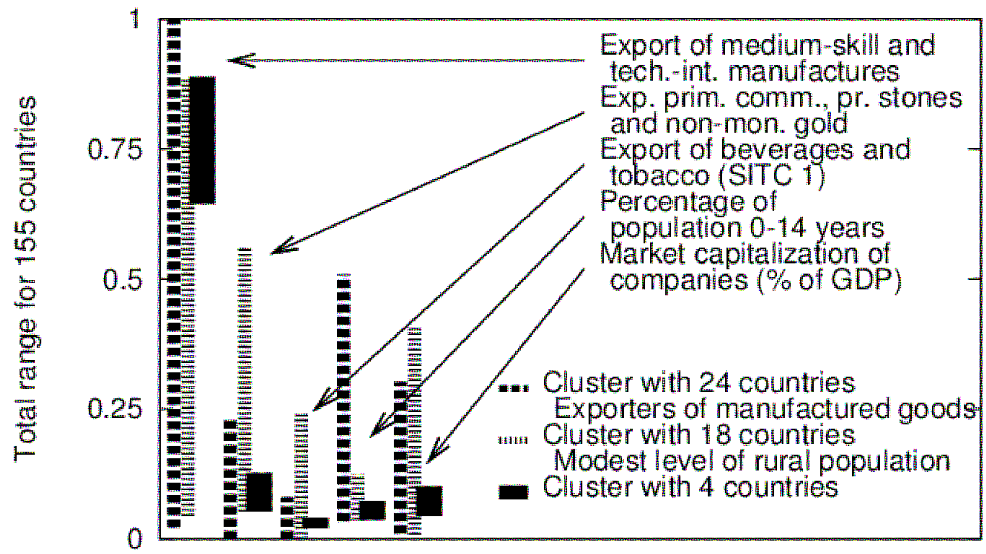

Figure 2: Distribution of values in three different clusters for five attributes: three from the export layer and two from the socio-economic layer.

4 Conclusions

In this work we have presented a novel clustering methodology that may be useful in different discovery tasks. The most decisive advantages are that it may be successfully used on instances described by both numeric and nominal attributes and that it has a well defined stopping criteria. Experimental evaluation of this methodology and its comparison with other known approaches will be a topic of our future work. In this paper we used the country level trading data for the illustration of the results one can expect from this novel methodology. The results are encouraging because we succeeded to get coherent clusters with examples that have narrow ranges of attribute values in some relevant attributes. In the interpretation of the common properties of the included examples, countries in our case, we have used the property that clusters constructed by the multilayer approach are typically subsets of clusters obtained on single layers. This approach enables us to undertake a statistical comparison with most similar examples that are *not* included in the resulting clusters. The most relevant problem of the methodology is that constructed clusters are small and that they will tend to be even smaller if additional data layers are included.

Acknowledgements

This work was partially supported by EU projects MULTIPLEX 317532 "Foundational Research on Multilevel Complex Networks and Systems" and and MAESTRA 612944 "Learning from Massive, Incompletely annotated, and Structured Data". It has been supported in part by the Croatian Science Foundation under the project number I-1701-2014.

References

[1] Gan G. and Ma C. and Wu J. (2007) Data Clustering: Theory, Algorithms, and Applications. *Society for Industrial and Applied Mathematics.*

[2] Cha S. H. (2007) Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1:300-307.

[3] Kumar V. and Chhabra J.K. and Kumar D. (2014) Impact of Distance Measures on the Performance of Clustering Algorithms. *Intelligent Computing, Networking, and Informatics*, 243:183-190.

[4] Sun S. (2013) A survey of multi-view machine learning. *Neural Computing and Applications*, 23:2031-2038.

[5] Bickel, S. and Scheffer T. (2004) Multiview clustering. *Proc. of the Fourth IEEE Int. Conf. on Data Mining*, 19-26.

[6] Parida L. and Ramakrishnan N. (2005) Redescription mining: Structure theory and algorithms. *In Proc. of Association for the Advancement of Artificial Intelligence (AAAI '05)*, pp. 837-844.

[7] Galbrun E. and Miettinen P. (2012) From black and white to full color: extending redescription mining outside the boolean world. *Statistical Analysis and Data Mining*, pp. 284-303.

[8] Caldarelli G. (2007) Scale-Free Networks: Complex Webs in Nature and Technology. Oxford University Press.

[9] Breiman, L. (2001) Random forests. *Machine Learning* 45(1):5-32.

[10] Pfahringer, B., Holmes, G., Wang, C. (2004) Millions of random rules. In *Proc. of the Workshop on Advances in Inductive Rule Learning, 15th European Conference on Machine Learning (ECML).*

[11] *UNCTAD database*, http://unctadstat.unctad.org/.

[12] *World Bank*, http://data.worldbank.org/indicator.