The Similarity between Dissimilarities

David M. J. Tax[1], Veronika Cheplygina[2,1], Robert P. W. Duin[1],
Jan van de Poll[3]

[1] Pattern Recognition Laboratory, Delft University of Technology, The Netherlands
[2] Biomedical Imaging Group Rotterdam, Erasmus Medical Center, The Netherlands
[3] Transparency Lab, Amsterdam, The Netherlands (transparencylab.com)

Abstract. When characterizing teams of people, molecules, or general graphs, it is difficult to encode all information using a single feature vector only. For these objects dissimilarity matrices that do capture the interaction or similarity between the sub-elements (people, atoms, nodes), can be used. This paper compares several representations of dissimilarity matrices, that encode the cluster characteristics, latent dimensionality, or outliers of these matrices. It appears that both the simple eigenvalue spectrum, or histogram of distances are already quite effective, and are able to reach high classification performances in multiple instance learning (MIL) problems. Finally, an analysis on teams of people is given, illustrating the potential use of dissimilarity matrix characterization for business consultancy.

1 Introduction

Consider the problem of evaluating and improving performances of teams in organizations based on the employee responses to questionnaires. The teams differ in size, and the roles of employees may be different for every organization. A key question for an organizations top management is how to support the autonomy of these teams while still keeping an eye on the overall process and the coherency of the teams performance. Assuming a span of control of 10-15 direct reports for an average manager, a middlesize organization may easily comprise of hundreds of teams. So, pattern recognition in organizational development may supply fundamentally important information of how similar - or dissimilar - teams are [1, 20, 15]. A possible solution is to focus at the diversity within a team – is there a large group of people who are all doing a similar job, or are there some isolated groups of people who are doing very different from the rest? Identifying such groups – clusters of employees – would help to compare the organizational structures on a higher level.

More formally, in this paper we focus on comparing sets (teams) of different samples (employees), residing in different feature spaces (evaluation questions). Comparing the team structures would be equivalent to comparing similarity matrices, with each similarity matrix originating from a single team. Comparing similarities alleviates the problem of different feature spaces, yet is still not trivial because the sets can be of different sizes, and there are no natural correspondences between the samples.

Comparing distance matrices has links with comparing graph structures: a distance matrix between N samples can be seen as a fully connected graph with N nodes, where the nodes are unlabeled and the edges are associated with weights. In graph-based pattern recognition, approaches such as graph edit distance [3, 21] or graph kernels [12, 10] have been used to define distance or similarity measures between graphs. Graph matching approaches search for a best correspondence between the nodes and define the graph distance as a measure of discrepancy between the matched nodes and edges. Graph kernels define similarity by considering all possible correspondences. However, the search space for correspondences becomes very large if the nodes are unlabeled, and the graph is fully connected. In [16] we used a threshold on the distances to reduce the number of edges. However, this threshold had a large influence on the results, suggesting that the larger distances can, too, be informative.

To avoid removing informative edges and to present a computationally efficient solution, in this paper we focus on finding feature representations to represent distance matrices. By representing each distance matrix in the same feature space, they can be compared with each other, for example, using the Euclidean distance. We investigate several representations in this paper, based on spectra [6], histograms of all distances [18], histograms of nearest neighbor distances, and hubness properties [23]. A detailed description of the representations is given in Section 2.

In Section 3 we investigate how well these features representations can encode the class information for some artificial examples. In Section 4 we investigate how good these representations are for multiple instance learning (MIL) datasets, where the goal is to classify sets of feature vectors. In Section 4.2 we apply the representation on real-world organisational data, and discuss some of the insights that arise from comparing teams of people.

2 Dissimilarity matrix representation

We assume we have a collection of N square dissimilarity matrices $\{D_n \in \mathbb{R}^{m_n \times m_n}; n = 1...N\}$ of size $m_n \times m_n$. One element of matrix D_n is indicated by $D_n(i, j)$. We assume that the matrices have the following characteristics:

- The dissimilarities of objects to themselves is zero (i.e. the D_n have zeros on the diagonal), and the dissimilarity is symmetric ($D_n(i, j) = D_n(j, i)$). In situations the matrices are not symmetric, they are made symmetric by averaging D_n and its transpose: $\tilde{D}_n = (D_n + D_n^\top)/2$.
- The size $m_n \times m_n$ of the matrices can be different for each D_n. It is assumed that the matrices have a minimum size of 3×3.
- The order of the rows and columns is arbitrary and may be permuted without altering the information that is stored in the dissimilarity matrix.

For this data type we investigate a few simple vector representations. Additional similarities between the dissimilarity matrices are also possible (such as embedding each matrix into a low-dimensional space and using the earth-movers distance, or matching the rows and columns of the dissimilarity matrices and computing the Frobenius norm [14]), but these tend to be computationally expensive. Here we focus on vector representations of the dissimilarity matrices.

We consider the following representations:

1. spectrum features spect: use the k largest eigenvalues σ of the centered matrix D_n:
$$\mathbf{x}_n = \sigma_{1:k}(C^\top D_n C), \quad \text{where } C = \mathbb{I}_{m_n} - \frac{1}{m_n}\mathbf{1}\mathbf{1}^\top \tag{1}$$
2. histogram of distances hist: collect all dissimilarities from all matrices, split the range of dissimilarity values from 0 to the maximum into k equally-sized bins, and count for each D_n the number of occurrences into each bin. Optionally, the count can be converted into a frequency by dividing by $m_n(m_n + 1)/2$.
3. equalized histogram of distances histeq: split the range of dissimilarity values into k bins with an equal number of counts (instead of using equally wide bins). The bins become wider when dissimilarity values do not occur often, and they become smaller for frequently appearing values.
4. histogram of the k-nearest neighbor distances distnn: instead of collecting all dissimilarities, only the dissimilarities up to the k-nearest neighbors are used. Per row of D_n, only $k < m_n$ dissimilarities are used; the total number of dissimilarities is therefore reduced from $m_n(m_n + 1)/2$ to $m_n k$. By this variations in local densities are captured better.
5. histogram of how often samples are the k-th nearest neigbor of other samples disthub: a measure used in hub analysis [24]. First the dataset is represented by a k-occurence histogram which stores how often each sample is the k-th nearest neighbor of others. To make this representation comparable across datasets of different sizes, it is summarized by q quantiles of the histogram. For the final representation, we concatenate the quantile-histograms for different values of $k \in \{1, 3, \ldots, |K|\}$, resulting in a $|K| \times q$ dim. feature vector.

In some situations we might want to be invariant to (non-linear) scaling of the dissimilarity values. For example, the expert may only have provided a relative ranking, but not an exact dissimilarity between two elements of a set. In this case, the extracted features should be invariant to the scaling of the dissimilarities. In the above representations, only disthub is invariant.

3 Illustrative examples

To show the characteristics of the different representations, we construct some multi-class artificial datasets. Depending on the experiment we perform, the number of dissimilarity matrices, and the sizes of the matrices are varied.

- The `cluster` dataset is constructed to investigate how well the clustering structure can be characterized. In `cluster` the dissimilarity matrices are computed from 2-dimensional datasets, containing samples belonging to a varying number of clusters (up to four clusters). The class label of a dissimilarity matrix is equal to the number of clusters, and therefore this defines a 4-class classification problem.
- The `subspace` dataset is constructed to investigate how well the subspace structure can be characterized. In `subspace` the dissimilarity matrices are derived from p-dimensional Gaussian distributions, where the dimensionality is one ($p = 1$) for class 1, $p = 2$ for class 2, up to class 4.
- The `outlier` dataset is used to investigate the sensitivity to outliers. In `outlier` the matrices are derived from 2-dimensional Gaussian distributions (zero mean, identity covariance matrix). Class 1 does not contain outliers. Class 2 contains an outlier from a Gaussian distribution with a 10 times larger covariance matrix, and for class 3 contains two such outliers.

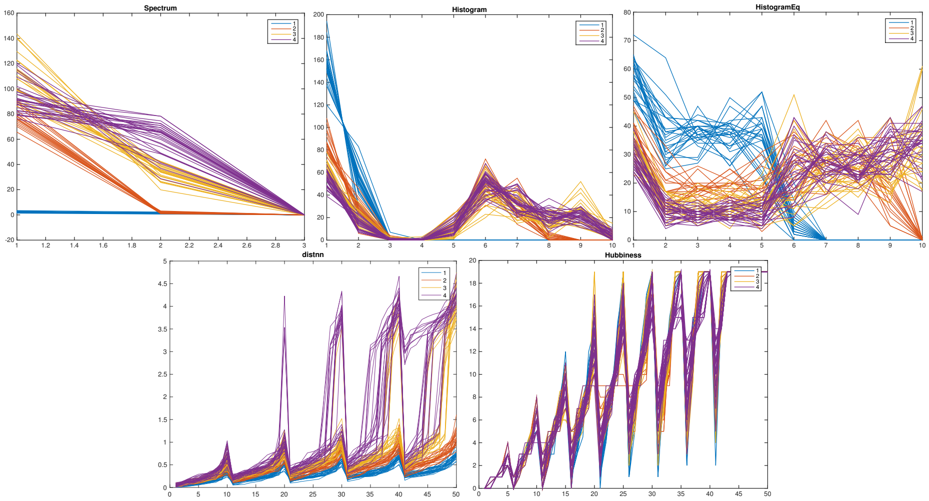

Fig. 1. The different feature representations for dissimilarity matrices derived from the `cluster` dataset with one, two, three or four clusters.

Figure 1 shows the five different representations for a sample of 100 dissimilarities drawn from the `cluster` dataset. For the spectrum representation three features are computed, for the (equalized) histograms $k = 10$ and for the `disthub` representation in total 75 features are computed. This `cluster` dataset has a very clear structure, and all the representations are able to distinguish well between the four different classes. In particular, the distinction between 1 cluster and more-than-1 cluster datasets are easy to make. For the `disthub` representation the distinction between the classes is less visible in the figure, due to the large difference in scales between the different features.

For each dataset, we compute the dissimilarity matrices using the Euclidean distances between the samples. We then compute the different representations, and use a linear classifier (LDA) to distinguish the different classes per dataset.

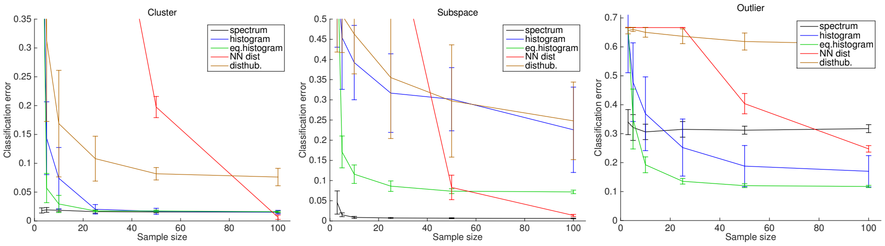

Fig. 2. Learning curves showing the error as a function of the number of training dissimilarity matrices. From left to right: results on `cluster`, `subspace` and `outlier` data.

Figure 2 shows the classification performance on the three artificial sets as function of the number of training matrices. The size of the individual dissimilarity matrices is fixed to $m_n = 30$. For many situations the (equalized) histogram is able to capture the information needed for good generalization. The histogram estimates start to suffer from noise for very small datasets and for situations where there is no clustering structure, and only the subspace dimensionality is informative. In these situations a spectrum representation is to be preferred. When very large training sizes are available, it is advantageous to use the nearest-neighbor distance histograms. Because `distnn` combines the histograms of the first-, second-, and all higher-nearest neighbors, this representation becomes very high-dimensional, but also very rich.

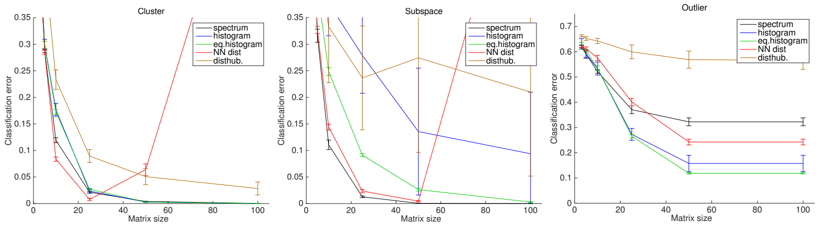

Fig. 3. The classification performance as function of the size m_n of the dissimilarity matrices, for the cluster data (left), the subspace data (middle) and the outlier data (right).

In Figure 3 a similar curve to Figure 2 is shown, only here the sizes of the individual dissimilarity matrices are varied while the number of training matrices is fixed to $N = 100$ per class. Here as well the (equalized) histograms perform well when the dissimilarity matrices are large. Then there is a sufficient number of values available to estimate histograms well. For very small matrices, and characterizing subspace structure or outliers, the spectrum performs well. Somewhat surprising, to characterize the clustering structure with small dissimilarity sizes, the nearest neighbor distances are most effective, although this tends to overfit with larger matrices.

4 Experiments

We distinguish between two sets of experiments, a supervised and an unsupervised set. For the supervised set, we have a collection of labeled dissimilarity matrices. Here we use the bags from MIL data, where the distances between the instances in one bag give one dissimilarity matrix, and each matrix is labeled according to the original bag label (positive or negative). For the unsupervised set we only have a collection of dissimilarities between teams of people, for which we want to investigate how much variability is present in the teams, and what constitutes this variability.

4.1 Supervised experiments: multiple instance learning

We look at a wide variety of multiple instance learning (MIL) problems. In MIL, the i-th sample is a bag $B_n = \{\mathbf{x}_{n1}, \mathbf{x}_{n2}, \ldots, \mathbf{x}_{nm_n}\}$ of m_n instances. The goal is to classify bags, based on the presence of *concept* feature vectors, or based on the overall distribution of the bag's instances. Consider image classification, where a bag is an image, and an instance is an image patch. When classifying images of tigers, a patch containing a tiger is an example of a concept instance. When classifying images of scenes, it might be more reasonable to examine several patches before deciding what type of environment the image is depicting.

Characteristics of the datasets are listed in Table 1. From our previous experiences with these datasets [5, 25], we expect these datasets to contain a mix of concept-like and distribution-like problems. Note that in our previous work [5, 25] we represented each bag by its dissimilarities relative to a set of prototype bags, whereas here we use an absolute representation where each bag is represented by dissimilarities between its own instances.

Dataset	#bags	#instances	Dataset	#bags	#instances
	neg/pos	min-mean-max		neg/pos	min-mean-max
Musk 1	23/37	3-7-40	alt.atheism	50/50	22-54-76
Musk 2	53/37	4-73-1044	comp.graphics	51/49	12-31-58
Corel African	1410/93	3-5-13	Harddrive	178/190	3-186-299
AjaxOrange	1440/60	31-32-32	Brown Creeper	350/197	3-19-43
Web recomm. 1	55/20	9-46-229	Biocreative comp.	2591/396	3-12-53

Table 1. Characteristics of MIL Datasets. Most of the datasets are available for download from `http://www.miproblems.org`

We removed bags that contained only 1 or 2 instances. We then represented each bag by a $m_n \times m_n$ dissimilarity matrix between its instances, where the dissimilarity is simply the Euclidean distance between the feature vectors. We represented each dissimilarity matrix with the representations described in Section 2. We used two classifiers: a linear discriminant classifier and a 1-nearest neighbor classifier. The experiments were performed using 10-fold cross-validation, where the best hyper parameter for each representation type (the optimal value for k), was determined on the training set using a second internal 10-fold cross-validation. We choose $k \in \{5, 10, 25, 50, 100\}$.

We report the AUC performances of both classifiers, using the best parameters for each representation type. For reference, we also list the best performance of traditional MIL classifiers[1]. The classifiers that often perform well are MILES [4], MI-SVM [2], EM-DD [27], a logistic classifier trained on a bag summary representation (based on the mean instance, or the min/max values per feature) [11], and p-posterior classifier [26].

The results are similar to those on artificial data: when the dissimilarity matrices are small, a spectrum representation is preferred. When larger training sets are available, it is often good to choose for an equalized histogram. These histograms tend to become relatively high dimensional, and the classifier can therefore not be too complex, so a linear classifier is a good choice.

What is also surprising is that, although these representations remove the absolute locations of the instances in the feature space, it is still possible to achieve very reasonable classification performance. For some datasets classification performances exceed the best performances achieved up to now (comp.graphics, Biocreative) or are comparable (Corel African, alt.atheism, Harddrive). For datasets that contain a specific concept (Musk1, Musk2, AjaxOrange, Web recomm. 1, Brown Creeper), the classifier that has access to individual feature vectors is better off.

4.2 Unsupervised experiments: analysis of teams of people

Given the required speed in a strategic decision making process, we used an online survey for the unsupervised gathering of a strategic status update from 20,191 employees in 1,378 teams in 277 different client projects on, for example, Human Resource Management, Information Technology and Marketing and Sales. We did not use a Likert scale given the subsequent need for statistical corrections for the structure of the survey [9], for various response styles [7], for a variety of sampling errors [19] and for a wide variety of biases [22]. Instead, we opted to use a Guttman scale with objective verifiable answers [8, 13]. The assessment questions were different for different teams. Four different types of assessment can be distinguished: (1) human resource (HR): focusing on team effectiveness, competency assessments, cultural aspects, (2) strategy: how strategy is finally incorporated, innovation assessment, (3) marketing and sales: analysis

[1] Available from `http://homepage.tudelft.nl/n9d04/milweb/`

Musk 1
MI-SVM 92.9 (1.3)

repr	LDA	1-NN
spect 25D	**74.2 (18.7)**	55.8 (18.9)
hist 5D	**60.6 (27.6)**	**54.6 (18.7)**
histeq 25D	45.6 (26.1)	52.7 (21.4)
distnn 10	**68.8 (16.8)**	50.8 (17.2)
disthub 5D	**73.3 (19.4)**	**68.8 (17.1)**

Musk 2
MILES 95.3 (1.5)

repr	LDA	1-NN
spect 10D	**53.5 (21.1)**	**58.6 (19.5)**
hist 50D	**59.1 (28.6)**	**50.8 (21.5)**
histeq 5D	**59.3 (23.3)**	**51.6 (15.5)**
distnn 5	**64.3 (26.1)**	**63.0 (23.3)**
disthub 5D	**55.8 (22.4)**	**63.7 (20.7)**

Corel African
EM-DD 91.5 (0.4)

repr	LDA	1-NN
spect 25D	65.3 (9.1)	**74.7 (11.0)**
hist 100D	81.2 (12.0)	**73.0 (7.7)**
histeq 10D	**87.8 (9.1)**	**76.1 (15.7)**
distnn 5	**87.5 (5.8)**	**78.2 (10.4)**
disthub 5D	59.5 (9.4)	51.6 (12.2)

SIVAL AjaxOrange
MI-SVM 99.6 (0.1)

repr	LDA	1-NN
spect 25D	**87.0 (9.1)**	**70.0 (15.4)**
hist 10D	72.3 (11.6)	60.3 (10.8)
histeq 100D	68.8 (12.5)	61.1 (10.9)
distnn 5	73.6 (10.9)	**66.7 (15.2)**
disthub 20D	64.6 (13.2)	**68.8 (11.1)**

Web recomm. 1
MI-SVM 91.9 (0.0)

repr	LDA	1-NN
spect 5D	48.8 (20.3)	**67.3 (18.6)**
hist 50D	**63.2 (25.0)**	**66.8 (31.1)**
histeq 5D	58.3 (24.9)	**74.7 (29.2)**
distnn 20	**72.8 (22.6)**	**69.5 (21.1)**
disthub 20D	50.8 (21.9)	**53.3 (22.8)**

alt.atheism
Logistic on mean 85.2 (2.2)

repr	LDA	1-NN
spect 5D	**86.8 (10.7)**	**75.2 (14.7)**
hist 100D	76.0 (10.2)	**75.2 (12.9)**
histeq 5D	**76.8 (17.6)**	61.2 (14.0)
distnn 5	**82.8 (10.2)**	**66.4 (16.8)**
disthub 10D	56.8 (18.2)	60.0 (19.9)

comp.graphics
SimpleMIL logistic 73.0 (1.7)

repr	LDA	1-NN
spect 5D	**89.0 (11.4)**	61.3 (20.6)
hist 50D	73.2 (12.8)	**73.5 (14.8)**
histeq 50D	**82.6 (14.0)**	**72.4 (19.1)**
distnn 10	**90.9 (13.2)**	**79.4 (10.7)**
disthub 10D	55.6 (14.2)	**71.4 (10.2)**

Harddrive
P-posterior 98.5 (0.5)

repr	LDA	1-NN
spect 5D	88.5 (7.5)	96.2 (3.8)
hist 5D	95.0 (6.7)	95.3 (5.0)
histeq 25D	**98.7 (2.5)**	**99.1 (1.8)**
distnn 20	76.1 (20.8)	94.1 (3.0)
disthub 10D	89.6 (6.7)	80.5 (6.0)

Brown Creeper
MILES 95.8 (0.3)

repr	LDA	1-NN
spect 50D	56.4 (14.5)	69.1 (8.7)
hist 100D	**87.5 (12.3)**	**76.4 (10.0)**
histeq 100D	**81.9 (23.2)**	**82.2 (15.1)**
distnn 20	64.8 (14.6)	72.0 (10.4)
disthub 10D	76.3 (11.9)	65.7 (9.2)

Biocreative component
MI-SVM 84.0 (0.0)

repr	LDA	1-NN
spect 50D	78.2 (8.6)	81.0 (9.7)
hist 100D	85.9 (7.4)	78.0 (10.7)
histeq 25D	**88.6 (7.2)**	**87.1 (8.3)**
distnn 10	**89.1 (8.8)**	**86.9 (10.4)**
disthub 20D	76.7 (10.1)	67.7 (8.3)

Table 2. AUC mean (standard deviation) ×100% of two classifiers on five representations of MIL bags. Bold = best or not significantly worse than best representation per classifier.

of client processes, commercial guidance of shops, and (4) IT: project assessment, IT processes, IT governance. From the answers a dissimilarity is derived by computing the pairwise Euclidean distances between the answer scores of all the members in a team.

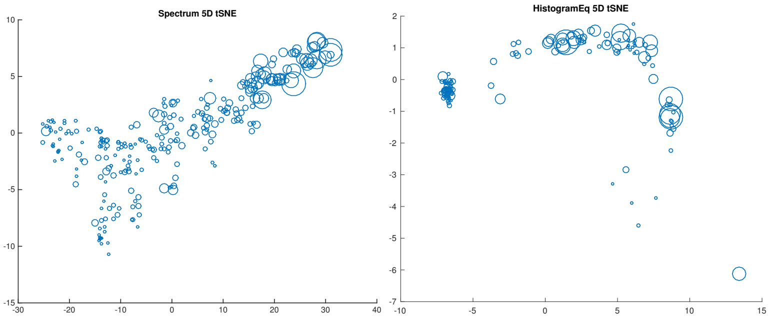

Fig. 4. t-SNE visualisations of 1378 teams of people. Left: spectrum representation of all the teams. The size of the circles indicate the size of the corresponding dissimilarity matrix. Right: equalized histogram representation for teams that got a human resource assessment.

In Figure 4 the resulting embeddings are shown using the spectrum and equalized histogram representations. Both representations are 5-dimensional, and the 2-dimensional embedding of the 5D data is obtained by using t-SNE [17]. In the left subplot the marker size indices the size m_n of the corresponding matrix D_n. It appears that the first important component is the size team. The plot also shows that there is more variation in the smaller teams, suggesting that in smaller team there is more possibility of specialisation. Larger teams tend to become more similar.

When we normalise for team size, and we focus on one type of questionnaires (Human Resource) we obtain the scatterplot on the right. There is one prominent outlier team. This appears to be a team that got a questionnaire with 160 questions, while normally less than 20 questions are used. Furthermore, there is a large cluster on the left, which contains fairly homogeneous team members, and a long tail up to the right where teams get stronger and stronger clusters of subteams. The teams most far in the tail show a clear clustering, while teams more close to the homogeneous cluster only contain a few outliers in a team.

5 Conclusions

We compared several feature vector representations for characterizing (square) dissimilarity matrices that can vary in size, and for which the rows and columns can be arbitrarily permuted. The spectrum representation is very effective, in particular when the sample sizes are small. It can not only characterize the intrinsic dimensionality, it is also able to characterize cluster structure. When a large sample size is available, it is often advantageous to use the more descriptive histograms of distances. These results can be observed in some artificial, and some real-world MIL problems. For MIL, the representations with a linear or nearest neighbor classifier are sometimes competitive to state-of-the-art classifiers.

We then used the representations in an unsupervised manner in order to characterize real-world organizations. Our analysis revealed some clusters of organizations, that could be interpreted by an expert. Given the current dissimilarity scores we suggest further research into the extent to which organizations are similar with respect to issues that affect a multitude of teams (a top management issue), a single team (a middle management issue) or a single employee (a lower management issue), and whether that similarity is particularly present in specific management topics (for example, in Human Resource Management) and/or in specific industries (e.g. in Professional Services).

References

1. Ahrens, T., Chapman, C.S.: Doing qualitative field research in management accounting: positioning data to contribute to theory. Accounting, Organizations and Society 31(8), 819–841 (2006)
2. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Advances in Neural Information Processing Systems. pp. 561–568 (2002)
3. Bunke, H., Riesen, K.: Recent advances in graph-based pattern recognition with applications in document analysis. Pattern Recognition 44(5), 1057–1067 (2011)
4. Chen, Y., Bi, J., Wang, J.: MILES: Multiple-instance learning via embedded instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1931–1947 (2006)
5. Cheplygina, V., Tax, D.M.J., Loog, M.: Multiple instance learning with bag dissimilarities. Pattern Recognition 48(1), 264–275 (2015)
6. Cvetkovic, D., Doob, M., Sachs, H.: Spectra of Graphs. Johann Ambrosius Barth Verlag, third edition edn. (1995)
7. De Jong, M.G., Steenkamp, J.B.E., Fox, J.P., Baumgartner, H.: Using item response theory to measure extreme response style in marketing research: A global investigation. Journal of marketing research 45(1), 104–115 (2008)
8. Diamond, I.D., McDonald, J., Shah, I.: Proportional hazards models for current status data: application to the study of age at weaning differentials in pakistan. Demography 23(4), 607–620 (1986)
9. Edelen, M.O., Reeve, B.B.: Applying item response theory (irt) modeling to questionnaire development, evaluation, and refinement. Quality of Life Research 5(5-18) (2007)
10. Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., Borgwardt, K.: Scalable kernels for graphs with continuous attributes. In: Advances in Neural Information Processing Systems. pp. 216–224 (2013)
11. Gärtner, T., Flach, P.A., Kowalczyk, A., Smola, A.J.: Multi-instance kernels. In: International Conference on Machine Learning. pp. 179–186 (2002)
12. Gärtner, T.: Predictive Graph Mining with Kernel Methods. Advanced methods for knowledge discovery from complex data pp. 95–121 (2005)
13. Hopkins, L., Ferguson, K.E.: Looking forward: The role of multiple regression in family business research. Journal of Family Business Strategy 5(1), 52–62 (2014)

14. Hubert, L., Arabie, P., Meulman, J.: 9. Anti-Robinson Matrices for Symmetric Proximity Data, chap. 11, pp. 115–141. ASA-SIAM Series on Statistics and Applied Probability (Book 19) (2006)
15. Lau, L., Yang-Turner, F., Karacapilidis, N.: Mastering Data-Intensive Collaboration and Decision Making: Research and practical applications in the Dicode project, chap. Requirements for Big Data Analytics Supporting Decision Making: A Sensemaking Perspective, pp. 49–70. Springer Int. Publishing, Cham (2014)
16. Lee, W.J., Cheplygina, V., Tax, D.M.J., Loog, M., Duin, R.P.W.: Bridging structure and feature representations in graph matching. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) 26(05) (2012)
17. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research 9(2579-2605), 85 (2008)
18. Papadopoulos A., M.Y.: Structure-based similarity search with graph histograms. In: Proc. of the Int. Workshop on Similarity Search. pp. 174–178 (1999)
19. Piterenko, K.: Business and impact alignment of questionnaire. Master's thesis, Gjovik University College (2013)
20. Plewis, I., Mason, P.: What works and why: combining quantitative and qualitative approaches in large-scale evaluations. International Journal of Social Research Methodology 8(3), 185–194 (2007)
21. Riesen, K., Fankhauser, S., Bunke, H., Dickinson, P.J.: Efficient suboptimal graph isomorphism. In: Graph-based Representations in Pattern Recognition. pp. 124–133 (2009)
22. Roulston, K., Shelton, S.A.: Reconceptualizing bias in teaching qualitative research methods. Qualitative Inquiry 21(4), 332–342 (2015)
23. Schnitzer, D., Flexer, A., Schedl, M., Widmer, G.: Local and global scaling reduce hubs in space. Journal of Machine Learning Research 13, 2871–2902 (2012)
24. Schnitzer, D., Flexer, A., Tomasev, N.: Choosing the metric in high-dimensional spaces based on hub analysis. In: ESANN (2014)
25. Tax, D.M.J., Loog, M., Duin, R.P.W., Cheplygina, V., Lee, W.J.: Bag dissimilarities for multiple instance learning. In: Similarity-Based Pattern Recognition. pp. 222–234. Springer (2011)
26. Wang, H.Y., Yang, Q., Zha, H.: Adaptive p-posterior mixture-model kernels for multiple instance learning. In: Proc. 25th Int'l Conf. Machine learning. pp. 1136–1143 (2008)
27. Zhang, Q., Goldman, S.A., et al.: EM-DD: an improved multiple-instance learning technique. In: Advances in Neural Information Processing Systems. pp. 1073–1080 (2001)