31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Neighborhood Graph and Learning Discriminative Distance Functions for Clinical Decision Support

Alexey Tsymbal, Shaohua Kevin Zhou, and Martin Huber

Abstract—**There are two essential reasons for the slow progress in the acceptance of clinical case retrieval and similarity search-based decision support systems; the especial complexity of clinical data making it difficult to define a meaningful and effective distance function on them and the lack of transparency and explanation ability in many existing clinical case retrieval decision support systems. In this paper, we try to address these two problems by introducing a novel technique for visualizing inter-patient similarity based on a node-link representation with neighborhood graphs and by considering two techniques for learning discriminative distance function that help to combine the power of strong "black box" learners with the transparency of case retrieval and nearest neighbor classification.**

I. Introduction

There is growing interest in the use of computer-based clinical decision support systems (DSSs) to reduce medical errors and to increase health care quality and efficiency [1]. Clinical DSSs vary greatly in design, functionality, and use. According to the reasoning method used in clinical DSS, one important subclass is that of Case-Based Reasoning (CBR) systems – systems which have reasoning by similarity as the central element of decision support [1], [2].

One reason for the slow acceptance of CBR systems in biomedical practice is the especial complexity of clinical data and the resulting difficulty in defining a meaningful distance function on them and adapting the final solution [3].

Another commonly reported reason for the relatively slow progress of the field is the lack of transparency and explanation in clinical CBR. Often, similar patients are retrieved and their diagnoses are presented, without specifying why and to what extent the patients are chosen to be similar and why a certain decision is suggested. We believe that, one way to approach this problem is to better visualize the underlying inter-patient similarity, which is the central concept of any clinical CBR.

Our main goal with this paper is to introduce a novel technique for visualizing patient similarity, based on neighborhood graphs, which can be helpful in clinical decision support. Besides, we consider two related techniques for learning discriminative distance function, which when used in combination with the neighborhood graphs can make them a powerful and flexible tool for clinical decision support in different classification contexts.

This paper is organized as follows. In Section II we review two techniques for learning discriminative distance functions; learning from equivalence constraints and the intrinsic Random Forest (dis-)similarity. In Section III a technique for visualizing inter-patient similarity based on neighborhood graphs is introduced and our implementation of it is discussed. We finish in Section IV with a brief summary and discussion of ongoing and future work.

Manuscript received April 7, 2009.

Alexey Tsymbal is with the Corporate Technology Division, Siemens AG, Erlangen 91058, Germany (phone: +49-9131-734624; fax: +49-9131-733190; e-mail: alexey.tsymbal@siemens.com).

Shaohua Kevin Zhou is with Siemens Corporate Research, Princeton, NJ 08540, USA (e-mail: shaohua.zhou@siemens.com).

Martin Huber is with the Corporate Technology Division, Siemens AG, Erlangen 91058, Germany (e-mail: martin.huber@siemens.com).

II. Learning Discriminative Distance Function

There are several reasons that motivate the studies in the area of learning distance functions and their use in practise [4]. First, learning a distance function helps to combine the power of strong learners with the transparency of nearest neighbor classification. Besides, learning a proper distance function was shown to be especially helpful for high-dimensional data with many correlated, weakly relevant and irrelevant features, where most traditional techniques would fail. Also, it is easy to show that choosing an optimal distance function makes classifier learning redundant. Next, learning distance functions breaks the learning process into two sequential steps (distance learning followed by classification or clustering), where each step requires search in a less complex functional space than in the immediate learning. Besides, it fosters the creation of more modular and thus more flexible systems, supporting component reuse. Another important benefit is the opportunity for inductive transfer between similar tasks; this approach is often used in computer vision applications; see e.g. [5].

A. Learning from Equivalence Constraints

Usually, equivalence constraints are represented using triplets (x_1, x_2, y), where x_1, x_2 are data points in the original space and $y \in \{+1,-1\}$ is a label indicating whether the two points are similar (from the same class) or dissimilar. Learning from these triples is also often called learning in the *product space* (i.e. with pairs of points as input); see [6], [7] for examples. While learning in the product space is perhaps a more popular form of learning from equivalence constraints, yet another common alternative is to learn in the *difference space*, the space of vector differences; see [8], [9] for examples. The difference space is normally used with homogeneous high-dimensional data, such as pixel

intensities or their PCA coefficients in imaging. While both representations demonstrate promising empirical results in different contexts, there is no understanding which representation is better. No comparison was done so far; usually a single representation for the problem is chosen.

There are two essential reasons that motivate the use of equivalence constraints in learning distance functions; their availability in some learning contexts and the fact that they are a natural input for optimal distance function learning [4]. It can be shown that the optimal distance function for classification is of the form $p(y_i \neq y_j \mid x_i, x_j)$. Under the *i.i.d.* assumption the optimal distance measure can be expressed in terms of generative models $p(x \mid y)$ for each class as follows [5]:

$$p(y_i \neq y_j \mid x_i, x_j) = \sum_y p(y \mid x_i)(1 - p(y \mid x_j)) \qquad (1)$$

B. The intrinsic Random Forest distance function

For a Random Forest (RF) learnt for a certain classification problem, the proportion of the trees where two instances appear together in the same leaves can be used as a measure of similarity between them [10]. For a given forest f the similarity between two instances x_1 and x_2 is calculated as follows. The instances are propagated down all K trees within f and their terminal positions $\boldsymbol{z}$ in each of the trees ($\boldsymbol{z_1}=(z_{11},\ldots,z_{1K})$ for x_1, similarly $\boldsymbol{z_2}$ for x_2) are recorded. The similarity between the two instances then equals to (I is the indicator function):

$$S(x_1, x_2) = \frac{1}{K} \sum_{i=1}^{K} I(z_{1i} = z_{2i}) \qquad (2)$$

Similarity (2) can be used for different tasks related to the classification problem. Thus, Shi and Horvath [11] successfully use it for hierarchical clustering of tissue microarray data. First, unlabeled data are expanded with a synthetic class of evenly distributed instances, then a RF is learnt and the intrinsic RF similarities are determined as described above and clustered. The resulting clusters are shown to be clinically more meaningful than the Euclidean distance based clustering with regard to post-operative patient survival.

Interesting is that using this similarity for the most immediate task, nearest neighbor classification, is rather uncommon, comparing to its use for clustering. In one of related works, [12], it is used for protein-protein interaction prediction, and the results compare favourably with all previously suggested methods for this task.

The intrinsic RF distance is rather a "dark horse" with respect to learning from equivalence constraints. The number of known applications for it is still limited; perhaps, the most successful application is clustering genetic data, [11]. Works on learning equivalence constraints never consider it as a possible alternative. In general, we believe that the circle of applications both for distance learning from equivalence constraints (which is currently applied nearly solely to imaging problems) and for the intrinsic RF distance is still, undeservedly, too narrow and may and should be expanded.

III. Visualizing Patient Similarity with Neighborhood Graphs

A. Background and Implementation

Neighborhood graphs provide an intuitive way of patient similarity visualization with an entity-relationship representation. There can be distinguished three basic types of neighborhood graphs that can be used to visualize object proximity in DSSs; (1) relative neighborhood graph (RNG), (2) distance threshold graph, and (3) directed nearest neighbor graph. These graphs are studied and applied in different contexts; in particular, the threshold and neighborhood graphs are often used for analyzing gene expression data in bioinformatics.

In a *relative neighborhood graph*, two vertices corresponding to two instances A and B in a data set are connected with an edge, if there is no other instance C which is closer to both A and B with respect to a certain distance function d [13]:

$$d(A, B) \leq \min_{C \neq A,B} \max\{d(A, C), d(B, C)\} \qquad (3)$$

Originally, relative neighborhood graphs were defined for two-dimensional data (planar sets) with the Euclidean distance metric, but later they were generalized and applied to multiple dimensions and other distance functions [13]-[15].

Besides the relative neighborhood graphs we focus on, there are known some other similar known node-link (graph-based) visualizations of instance proximity. These include the Minimum spanning tree (MST), the Gabriel graph, and the Delanay tessellation [13]. We believe that out of this family, the relative neighborhood graph is the best candidate to visualize patient proximity in a DSS. The MST has usually too few edges to spot groupings/patterns in the data, while the Gabriel graph and the Delanay tessellation are, vice versa, usually too overcrowded, which becomes a problem with already more than 10 instances.

A threshold graph is simply defined as a graph where two vertices are connected with an edge if the distance between the two corresponding instances is less than a certain threshold. In a *nearest neighbor* graph, each instance is connected with one or a set of its nearest neighbors. This graph is usually directed as the relation of being a nearest neighbor is not necessarily symmetric. An important benefit of RNG comparing to the other two graphs is the fact that it is always connected with nodes having a reasonable small degree; it is often planar or close to planar.

In machine learning, neighborhood graphs find various applications, including clustering, outlier removal, and even supervised discretization [15]. In other areas, other, more

exotic applications may also be found. In [16], for example, the minimum spanning tree of a neighborhood graph was applied to contrast-based hierarchical image segmentation. In [17] a directed relative neighborhood graph and directed minimum spanning tree are successfully applied to topology control, in order to create a power-efficient network topology in wireless multi-hop networks with limited mobility.

One important limitation of RNGs which affects their scalability is the computational complexity of the algorithms that construct them. There are known solutions for speed-up for dimensionalities up to three, but in the general case the computational complexity remains $O(n^3)$, where n is the number of nodes (instances). One possible solution for too large data sets is to cluster them into a number of groupings, for which calculating the RNG would be feasible, and then to create an RNG separately for each of them, also connecting close clusters (the corresponding neighborhood graphs) with each other.

In our toolbox for visualization, navigation and management of the three neighborhood graphs introduced above, which is being developed as a part of similarity search-based clinical DSS CaseReasoner, we implement the following functionality:

- node colouring, to represent numeric and nominal attributes;
- node filtering, according to attribute values in the patient record;
- edge colouring and filtering, according to the underlying distance;
- graph clustering into an arbitrary number of components;
- reconfigurable tooltips displaying clinical data from the patient record and images;
- nearest neighbor classification and regression performance visualization for each node, for a selected attribute and a certain similarity context;
- image visualization within the nodes of the graph (e.g. meshes corresponding to the pulmonary trunk of the patient can be displayed).

For clustering the neighborhood graphs, we use the following two algorithms; (1) the Girvan and Newman's algorithm for graph clustering which is often used for clustering of complex social and biological networks [18], (2) top-down induction of a semantic clustering tree (in the original feature space), the goal of which is to provide every cluster with a semantic description that can be inspected by a clinician and may carry important information.

Perhaps the main competitor to neighborhood graphs as a tool for visualizing patient similarity is heatmaps, which are well known and often used by clinical researchers, in particular by geneticists. In comparison to heatmaps, as follows also from the feedback obtained from partner clinicians in the project, neighborhood graphs possess a number of advantages. In particular, they are easier to read with the more intuitive node-link representation, they allow visualizing additional features or even image thumbnails at nodes, and they have a flexible layout allowing to naturally visualize clusters, enlarge nodes, and filter our a set of nodes and edges.

B. Examples

In Fig. 1 an example RNG for a set of 63 meshes representing aortic valves is shown. Two classes of valves are distinguished; healthy (blue) and diseased (red). The current patient is highlighted in orange and her immediate neighbors in yellow. The underlying distance is learnt using RF in the product space of equivalence constraints, that was proven to be the best technique for this task. Girvan-Newman clustering of the graph into two clusters is presented. One may see that the clustering nearly perfectly corresponds to the two classes. Such a graph is an expressive way of presenting relevant information. From the graph, one may easily comprehend patient distribution according to the studied similarity context and see patient groupings, identify outliers, easy to classify cases and the borderline cases classification for which is likely to be uncertain. More information about this data set, the problem domain and the model for mesh generation can be found in [19]. Leave-one-out accuracy for this problem, using distance learning with RF in the product space, is as high as 93%.

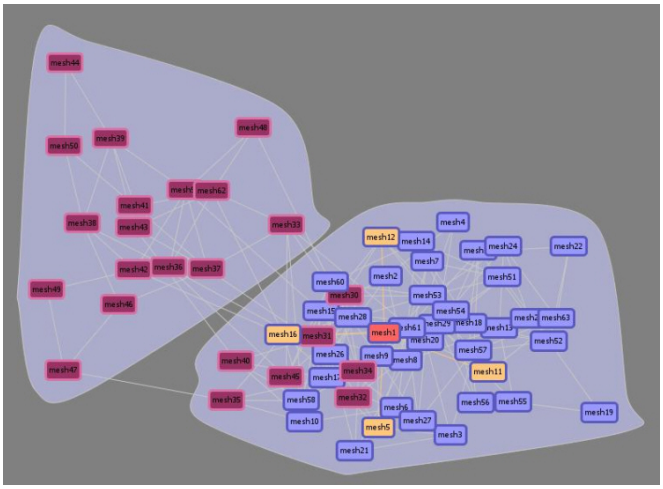

Fig. 1. A relative neighborhood graph for a set of 63 aortic valves.

In Fig. 2 a 1-nearest neighbor graph for the same problem and the same distance function is shown. While this graph is less crowded than RNG and is easier to read and interpret with regards to its elements, it often includes disconnected components, making it often difficult to see and estimate the data distribution as a whole with the component clusters.

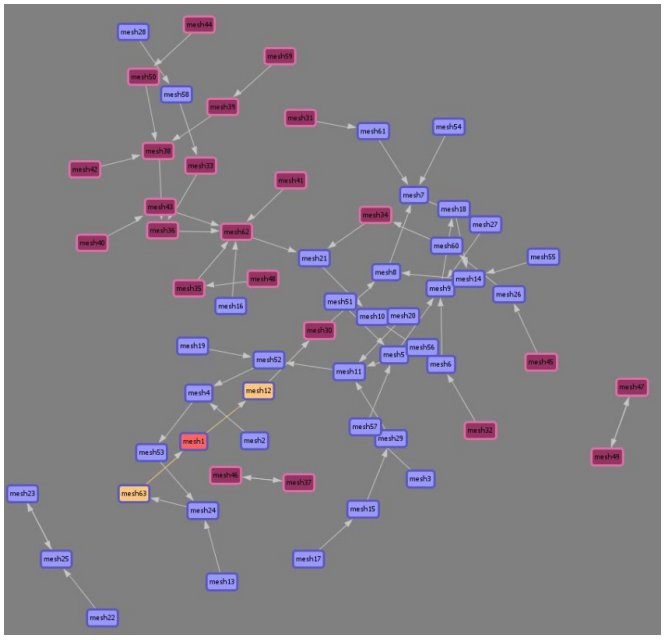

Fig. 2. A 1-nearest neighbor graph for the same set.

IV. Discussion and Conclusion

In this paper two techniques for learning discriminative distance functions were considered; learning from equivalence constraints and the intrinsic RF distance. Our preliminary experiments confirm that both techniques demonstrate competitive performance with respect to the plain learning; they help to combine the power of strong "eager" learners with the transparency of case retrieval and nearest neighbor classification. The intrinsic RF distance is proven to be more robust overall (in terms of better expected accuracy and fewer parameters to be tuned), although finding suitable parameters for learning from equivalence constraints may still be competitive. Future work includes studying various applications of the techniques to different subject domains with complex data, systematic analysis of the techniques and further validation of the findings.

The techniques considered in this paper may be useful for the development of real-world case retrieval and decision support systems. As an example, at present we are developing a system called CaseReasoner intended for the retrieval of patient records that may include complex biomedical data; clinical, imaging and genomic data. The underlying inter-patient similarity (that can be calculated via distance functions learnt using techniques considered in this paper) can be visualized using heatmaps, treemaps and neighborhood graphs, for better knowledge discovery and decision support. An important item of our ongoing work in this context is the acquaintance of partner clinicians with the considered neighborhood graph visualizations in the framework of CaseReasoner, and their evaluation in the context of different data classification and decision support tasks.

Acknowledgment

This work has been done in the framework of the EU project Health-e-Child (IST 2004-027749). The authors wish to acknowledge support provided by all the members of the Health-e-Child consortium in the preparation of this paper.

References

[1] A. Berlin, M. Sorani and I. Sim, "A taxonomic description of computer-based clinical decision support systems," *J. of Biomedical Informatics*, vol. 39, no. 6, 2006, pp. 656-667.

[2] M. Nilsson and M. Sollenborn, "Advancements and trends in medical case-based reasoning: an overview of systems and system development," in *Proc. 17th Int. FLAIRS Conf. on AI, Special Track on CBR*, AAAI Press, 2004, pp. 178-183.

[3] R. Schmidt and O. Vorobieva, "Adaptation and medical case-based reasoning focusing on endocrine therapy support," in *Proc. Int. Conf. on AI in Medicine*, LNCS, vol. 3581, Springer, 2005, pp. 300-309.

[4] A. Bar-Hillel, "*Learning from weak representations using distance functions and generative models*," Ph.D. dissertation, Dept Comp. Sci., Hebrew Univ. of Jerusalem, 2006.

[5] S. Mahamud and M. Hebert, "The optimal distance measure for object detection", in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 2003.

[6] T. Hertz, A. Bar-Hillel and D. Weinshall, "Learning distance functions for image retrieval," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 2004.

[7] S. K. Zhou, J. Shao, B. Georgescu and D. Comaniciu, "Boostmotion: boosting a discriminative similarity function for motion estimation," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 2006.

[8] J. Amores, N. Sebe and P. Radeva, "Boosting the distance estimation. Application to the k-nearest neighbor classifier," *Pattern Recognition Letters*, vol. 27, 2006, pp. 201-209.

[9] J. Yu, J. Amores, N. Sebe and Q. Tian., "Toward robust distance metric analysis for similarity estimation," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition, CVPR*, 2006.

[10] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, 2001, pp. 5-32.

[11] T. Shi and S. Horvath, "Unsupervised learning with Random Forest predictors," *Computational and Graphical Statistics*, vol. 15, no. 1, 2006, pp. 118-138.

[12] Y. Qi, J. Klein-Seetharaman and Z. Bar-Joseph, "Random Forest similarity for protein-protein interaction prediction from multiple sources," in *Proc. Pacific Symp. on Biocomputing*, 2005.

[13] J. W. Jaromczyk and G. T. Toussaint, "Relative neighborhood graphs and their relatives," in *Proc. IEEE*, vol. 80, no. 9, 1992, pp. 1502-1517.

[14] G. T. Toussaint, "The relative neighborhood graph of a finite planar set," *Pattern Recognition*, vol. 12, no. 4, 1980, pp. 261-268.

[15] F. Muhlenbach and R. Rakotomalala, "Multivariate supervised discretization, a neighborhood graph approach," in *Proc. IEEE Int. Conf. on Data Mining*, IEEE Computer Society, 2002, pp. 314-321.

[16] B. Marcotegui and S. Beucher, "Fast implementation of waterfall based on graphs," in *Proc. 7th Int. Symp. on Mathematical Morphology, Computational Imaging and Vision*, vol. 30, Springer, 2005, pp. 177-186.

[17] N. Li and J. C. Hou, "Topology control in heterogeneous wireless networks: problems and solutions," in: *Proc. 23rd Annual Joint Conf. of IEEE Computer and Communications Societies, Infocom'04*, IEEE, 2004.

[18] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. USA*, vol. 99, 2002, pp. 7821-7826.

[19] R. I. Ionasec, A. Tsymbal, D. Vitanovski, B. Georgescu, S.K. Zhou, N. Navab and D. Comaniciu, "Shape-based diagnosis of the aortic valve," in Proc. SPIE Medical Imaging, vol. 7259, 2009.