MODELING GESTALT LAWS FOR CLASSIFICATION

Guihua Wen, Xingjiang Pan, Lijun Jiang

South China University of Technology
Guangzhou 510641, China
crghwen@scut.edu.cn

Jun Wen

Hubei Institute for Nationalities
Ensi 445400, China
wenj_64@163.com

ABSTRACT

The k-nearest neighbors classifier is simple and often results in good classification performance on problems with unknown and non-normal distributions. However, its selected nearest neighbors on noisy ,sparse,or imbalanced data are often inconsistent with our intuition and in turn leads to the worse performance. This paper applies Gestalt visual perceptual laws to design a new KNN classifier. It applies the neighborhood relation between any two data points to construct the geometry shape of data and then applies the Gestalt laws on this shape to perform the classification.The conducted experiments on challenging benchmark real data validate the proposed approach.

Keywords: *Classification, Gestalt Laws, cognitive geometry,nearest neighbors*

1. INTRODUCTION

Finding nearest neighbors plays an fundamental role in many artificial intelligence tasks, such as similarity searching[1], manifold learning[2, 5], and data mining[3, 4]. The *k* nearest neighbors (KNN) classifier directly applies this idea to perform classification. It is one of most important classification approaches due to its high classification accuracy in problems with unknown and non-normal distributions[3, 6, 7], so that it has wide applications[8]. This approach finds *k* nearest neighbors of the query in the training, and then predicts the class label of the query as the most frequent one occurring in the neighbors[7]. However, on high-dimensional, noisy, or sparse data, its performance is severely influenced [9], as under such cases the formed neighborhood structure is deformed. The selected neighbors can not be able to organize a meaningful and regular group. According to Gestalt theory, grouping is the main process in human visual perception[10]. Whenever previously formed visual objects have one or several characteristics in common, they easily get grouped and form a new, larger visual object, a Gestalt, satisfying properties of vicinity, similarity, continuity of direction, closure, pregnancy, etc [11]. These gestalt grouping laws are directly related to the geometric statistics of the natural world. Like Gestalt theory of perception and topological psychology[16],theory of topological perception also accounts for the amazing ability of human being by defining the global properties as topological invariants[15]. Some computational theories apply these theories to solve the problems such as to find Gestalts in digital images[10]. They used specific properties of images, instead of taking each image as independent objects for classification. The relative transformation is a newly proposed model based on the idea that perception is relative[19], which has been validated in manifold learning[20]. Another novel work is the proposal of principle of homology continuity[12]. It applies the relation of training samples in the same class to design artificial neural network. On the whole, current mathematical models used very little and almost nothing of the Gestalt theory results[10]. Different from these important work, this paper applies Gestalt visual perceptual laws to design a new KNN classifier. It applies the neighborhood relationship between any two data points to construct the geometry shape of data and then applies the Gestalt laws on this shape to perform the classification.

2. GESTALT PSYCHOLOGY

The Gestalt theory explains how we organize mental figure-like images and how we perceive images through various types of sensory input such as visual, auditory, and olfactory stimuli. It is based on the idea that we tend to order our experience in a manner that is regular, orderly, symmetric, and simple. It is also known as the law of the good figures which refers to the tendency to perceive the simplest and most stable figure of all possible perceptual alternatives. Gestalt psychologists organize these experiences as ”gestalt laws”[17, 18], some of which are shown as Fig.1 from ref[17].

–Proximity indicates that elements may be grouped according to their perceived closeness. Elements that appear close together tend to be grouped together. Spatial or temporal proximity of elements may induce the mind to perceive a collective or totality.

–Similarity indicates that similar elements physically tend to be grouped together.If data points are similar physically, they appear to be grouped together to form a Gestalt-like perception. The mind groups similar elements into collective en-

Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)
F. Sun, Y. Wang, J. Lu, B. Zhang, W. Kinsner & L.A. Zadeh (Eds.)
978-1-4244-8040-1/10/$26.00 ©2010 IEEE

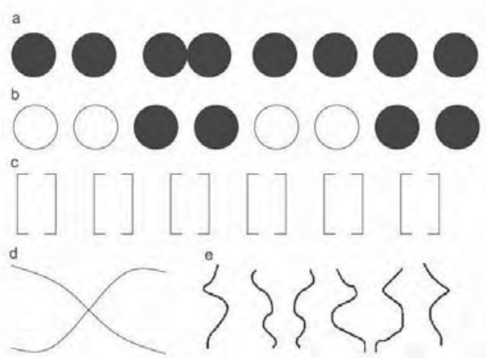

Fig. 1. Some Gestalt properties of grouping: proximity (a), similarity (b), closure (c), continuity (d), symmetry (e). Although segments in (c) are regularly distributed, they are grouped in order to form objects that tend to shape as close ones. Patterns in (d) are seen as two continuous lines intersecting rather than two contiguous curves with cusps.

tities or totalities. This similarity might depend on relationships of form, color, size, or brightness.

–Closure means that groupings occur in a way that favors perception of the more enclosed or complete figures,shown as Fig.1(c). It says that, if something is missing in an otherwise complete figure, we will tend to add it. A rectangle, for example, with a small part of its edge missing, will still be seen as a rectangle. We will "close" the gap.

–Continuity means that elements following in the same direction, such as along a straight line or a simple curve, are readily perceived as forming a group. The mind continues visual, auditory, and kinetic patterns when we can see a line, for example, as continuing through another line, rather than stopping and starting, we will do so, as in Fig.1(d), which we see as composed of two lines, not as a combination of two angles.

–Symmetry indicates that priority in perceptual grouping is given to more natural, balanced, and symmetrical figures over the asymmetrical ones. Symmetry is generally considered a critical factor in aesthetics. Symmetrical images are perceived collectively, even in spite of distance. Take a look at this example: [ ][ ][ ] Despite the pressure of proximity to group the brackets nearest each other together, symmetry overwhelms our perception and makes us see them as pairs of symmetrical brackets.

3. CLASSIFICATION BASED ON GESTALT LAWS

On the noisy and sparse data, the classification boundary of KNN may be deformed which in turn threaten the classification results. We apply gestalt laws to solve this problem by selecting *k* nearest neighbors that should organized on the basis of Gestalt grouping principles that can lead to efficient and simpler perceptions. Firstly, the nearest neighbors sequence of point q is constructed by using neighborhood relation as follows:

$$S(q,k) = \{x_{\sigma(0)}, x_{\sigma(1)}, \cdots, x_{\sigma(i)}, \cdots x_{\sigma(k)}\}$$

where $x_{\sigma(i+1)}$ is the nearest neighbor of $x_{\sigma(i)}$ and $x_{\sigma(0)} = q$ be the query sample.

Once the S is constructed, we can take it as nearest neighbors for classification by voting on it. Based on closure law, if there is a closure in the sequence S, the closure will be removed and the first part of S will be taken as nearest neighbors,denoted as S_r, instead of the whole S. Subsequently, we will apply the continuity law to perform the voting on the nearest neighbors in S_r. If the length of the subsequence of S_r having the same class is bigger than the given threshold, we can directly take the class of this subsequence as the class of the query q,or else the usual voting strategy over S_r is applied. Let

$$l_j = arg \max_j |\{x_i : x_i \in S \wedge \xi(x_i) = \xi(x_{i+1}) = \omega_j\}|$$

then classify q into class ω_j if

$$\omega_j = arg \max_{j \in \{1,2,\cdots,N_c\}} \{l_j\}$$

This procedure can be illustrated by Fig.2, where the constructed sequence S of $x0$ is $\{x0, x1, \cdots, x6\}$ and the closure in S is $\{x4, x5, x6\}$. Subsequently the final nearest neighbors set is $\{x1, x2, x3\}$ on which the classification will be performed.

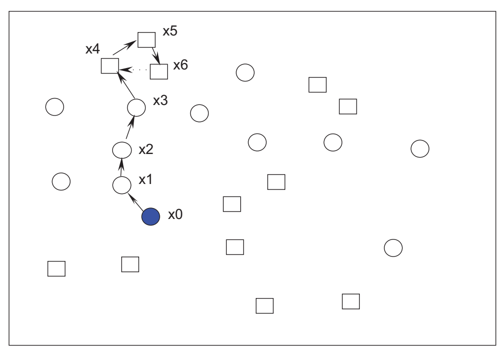

Fig. 2. Illustration of formed shape for $x0$ from data points by neighborhood relation

In some extreme cases, the nearest neighbors can not be organized to the well defined shape so that the above classification approach can not surely result in the better performance. We apply information entropy to measure it. Only the measured value is bigger than the given threshold, the classification approach above can be taken, or else the standard

KNN is applied.

$$E(S) = -\sum_{i=1}^{n_c} \frac{n_i}{|S|} log(\frac{n_i}{|S|})$$

where n_c be the number of classes in S, n_i be the number of samples belong to *i*-th class, and $|S|$ be the total number of samples in S.

Algorithm 1 GL-KNN(q, X, ξ, dc, r, k)

/* *q* is the query sample, X is training samples, $\xi(x_i)$ denotes the class of the sample x_i in *X*, *dc* and *r* are the thresholds, and *k* is the neighborhood size for classification*/

Step 1. Construct the neighborhood sequence *S* with length of *k* for the query *q*,denoted as *S(q,k)*.

Step 2. Find *k* nearest neighbors for the query *q* using the usual way,denoted as *N(q,k)*

Step 3. If $E(S(q,k)) < E(N(q,k))$, the voting strategy of standard KNN is applied to decide the class of the query *q* on *N(q,k)*,return.

Step 4. Remove the closure in *S* and denote the rest as S_r. If the length of S_r is smaller than *dc*,than $S_r = S$.

Step 5. If there is $l_j > r$ in S_r, decide the class of the query *q* by

$$\omega = arg \max_{j \in \{1,2,\cdots,N_c\}} \{l_j\}$$

or else we apply the voting strategy of standard KNN to decide the class of the query *q* on S_r

The parameters in GL-KNN can be optimized for each given data set by the cross validation method.

4. EXPERIMENTAL RESULTS

To validate the idea that gestalt law can be applied to improve the classification performance, we do experiments to make comparison between proposed GL-KNN and standard KNN. The accuracy rate is taken as the measure of performance of all compared classifiers, which usually is used as the most effective measure of the performance of a classifier[7, 9]. In experiments, *k* takes the value over the range of $\{3, 6, 9, \cdots, 30\}$ while r takes the values from $\{0.01, \cdots, 0.09\}$ and dc takes the values from $\{1, \cdots, 9\}$. When classifying, each data set is divided into training set and testing set according to the 'ModApte' split(70% for training samples and the remaining 30% for testing samples)[14]. The parameters are determined by validation on training set and then applied to perform the classification on testing set. Ten such partitions are generated randomly for the experiments, and then the best average performance is reported. Experimental data from UCI Repository of machine learning databases[13], where the records with missing values and non-numeric attribute are all removed. Most of them may be noisy, sparse, and imbalanced. It can be observed from Table.1 that GL-KNN outperforms KNN by average accuracy 1.13%. Most of them are obvious. It also can be observed that GL-KNN has the higher standard deviation of the accuracy than that of KNN. This reenforces the conclusion that GL-KNN takes effective only on the data having local Gestalt structure. These results do indicate the significant value of the proposed idea and the classifier.

Generally every method has its strengths and weaknesses. It is necessary to apply a measure to evaluate the robustness of the different methods. We take the usually used measure to quantify the robustness by computing the ratio b_m of the error rate e_m of method *m* and the smallest error rate over all methods being compared in a particular dataset: $b_m = e_m / \min_{1 \le k \le 10} e_k$[6, 7]. Thus, the best method m^* for that problem has b_m =1, and all other methods have larger values $b_m > 1$. The larger the value of b_m, the worse the performance of the method. This means that the distribution of b_m will be a good indicator reflecting its robustness. For example, if a particular method has an error rate close to the best in every problem, its b_m values should be densely distributed around the value 1. Any method whose b_m value distribution deviates from this ideal distribution reflect its lack of robustness.We calculate e_m for each method in terms of the average error rate of the ten best results with respect to parameters on each data set. Fig.3 shows the distribution of b_m for each method over the ten real data sets. Clearly, the spread for GL-KNN is much narrower and closer to one. This result demonstrates that it obtains the most robust performance over these data sets. Although these results are data-specific and sensitive to how the classifiers were parameterized, they do indicate the value of proposed approaches and related classifiers.

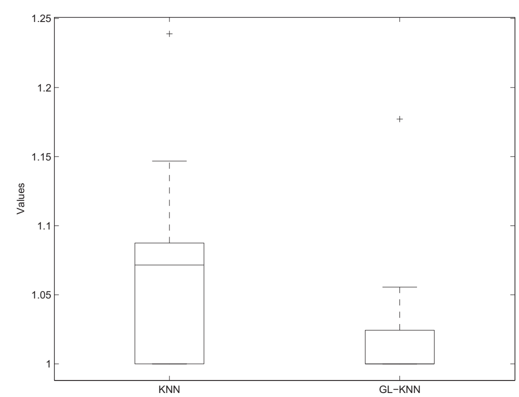

Fig. 3. Average performance distribution of KNN and GL-KNN

The experimental results may differ from the results reported elsewhere due to the particular split of the data used. They may not be enough to reflect many real problems, but at least, GL-KNN is of consistent behaviors on showing better performance. This is consistent with our idea that classifier works well on real data if it fits for the gestalt laws.

Table 1. Average values and standard deviations of the accuracies of KNN and GL-KNN on real data sets(%)

Data	Attributes	Classes	Size	KNN	GL-KNN
wine	13	3	178	69.42±4.92	**71.54±6.00**
dermatology	34	6	358	85.95±2.65	**87.08±2.09**
diabetes	8	2	768	**73.35±2.10**	72.70±2.18
ionosphere	34	2	351	83.56±2.38	**86.73±2.21**
glass	9	7	214	66.72±5.13	**70.98±5.47**
optdigits	64	10	1797	98.58±0.95	**98.67±0.57**
segmentation	19	7	210	76.03±4.88	**77.78±3.26**
yeast	8	10	1484	**58.46±2.88**	56.15±3.55
yaleface	4135	15	165	62.00±4.12	**64.44±5.93**
iris	4	3	150	**96.05±1.57**	95.35±1.90
			avg	77.01±3.15	**78.14±3.31**

5. CONCLUSION AND FUTURE WORK

Although KNN has remarkable generalization performance, it can not nicely deal with noisy, sparse, and imbalanced data, as where the true topological structure of data is deformed.This paper proposes a different approach to find nearest neighbors for classification. To the best of our knowledge, this approach is the first one to find nearest neighbors by modeling perceptual Gestalt laws. The most importance of our work seems that it opens a new direction as a fundamental methodology to develop a quantitative Gestalt theory which is then applied to design all kinds of classifiers. We are therefore encouraged to model all kinds of Gestalt laws for classification in the future.

6. ACKNOWLEDGEMENTS

This work was supported by China National Science Foundation under Grants 60973083, ”the Fundamental Research Funds for the Central Universities,SCUT”, Guangdong Science and Technology project under Grants 2007B030803006.

7. REFERENCES

[1] H.Samet,K-Nearest Neighbor Finding Using MaxNearestDist,IEEE Trans. Pattern Anal. Mach. Intell.30,243(2008)

[2] J.B. Tenenbaum and V. de Silva and J.C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction,Science 290,2319(2000)

[3] M.H Tristan and R.Stephane,Tailored Aggregation for Classification,IEEE Trans. Pattern Anal. Mach. Intell.31,2098(2009)

[4] B. J. Frey and D. Dueck, Clustering by Passing Messages Between Data Points,Science 315,972(2007)

[5] G.Wen,L.Jiang, and J.Wen, Using Locally Estimated Geodesic Distance to Optimize Neighborhood Graph for Isometric Data Embedding,Pattern Recognition 41,2226(2008)

[6] C. Domeniconi, J. Peng and D. Gunopulos, Locally adaptive metric nearest-neighbor classification,IEEE Trans. Pattern Anal. Machine Intell.24, 1281(2002)

[7] B.Li, Y.W.Chen, and Y.Q.Chen, The Nearest Neighbor Algorithm of Local Probability Centers, IEEE Trans. Syst., Man, Cybern.38, 141(2008)

[8] S. N.Karl and Q. N.Truong,An Adaptable k-Nearest Neighbors Algorithm for MMSE Image Interpolation,IEEE Trans. Image Processing 18, 1976(2009)

[9] Y.Mitani and Y. Hamamoto, A local mean-based nonparametric classifier, Pattern Recognition Letters 27, 1151(2006)

[10] A.Desolneux, L.Moisan, and J.Morel,Computational gestalts and perception thresholds,Journal of Physiology - Paris 97, 311(2003)

[11] K.Koffka , Principles of Gestalt Psychology, Routledge, London,1955

[12] S.Wang and X.Chen,Biomimetic (Topological) Pattern Recognition- A new Model of Pattern Recognition Theory and Its Application,Proceedings of the international joint conference on neural network,2258(2003)

[13] A. Asuncion and D.J.Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/ mlearn /MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science.(2007).

[14] Wai Lam and Yiqiu Han,Automatic Textual Document Categorization Based on Generalized Instance Sets and a Metamodel,IEEE Trans. Pattern Anal. Mach. Intell.,2003,25(5):628-633

[15] L.Chen,The topological approach to perceptual organization,Visual Cognit 12,553(2005)

[16] K.Lewin, Principles of topological pyschology, (McGraw-Hill Book Company,Inc.,1936)

[17] Nicolas Zlatoff, Bruno Tellez, Atilla Baskurt, Combining local belief from low-level primitives for perceptual grouping,Pattern Recognition 41 (2008) 1215-1229

[18] Shih-Wen Hsiao, Jyh-Rong Chou,A Gestalt-like perceptual measure for home page design using a fuzzy entropy approach,Int. J. Human-Computer Studies 64 (2006) 137-156

[19] Guihua Wen et al.,Local relative transformation with application to isometric embedding,Pattern Recognition Letters, 2009,30(3):203-211

[20] Guihua Wen et al.,Locally linear embedding based on relative manifold, Journal of Software,2009,20(9):2376-2386.