

Eliciting Domain Knowledge in Handwritten Digit Recognition

Tuan Trung Nguyen

Polish-Japanese Institute of Computer Technology,
ul. Koszykowa 86, 02-008 Warsaw, Poland
nttrung@pjwstk.edu.pl

Abstract. Pattern recognition methods for complex structured objects such as handwritten characters often have to deal with vast search spaces. Developed techniques, despite significant advancement in the last decade, still face some performance barriers. We believe that additional knowledge about the structure of patterns, elicited from humans perceptions, will help improve the recognition's performance, especially when it comes to classify irregular, outlier cases. We propose a framework for the transfer of such knowledge from human experts and show how to incorporate it into the learning process of a recognition system using methods based on rough mereology. We also demonstrate how this knowledge acquisition can be conducted in an interactive manner, with a large dataset of handwritten digits as an example.

1 Introduction

Most existing pattern recognition systems use a descriptive language and a reasoning scheme, both assumed *a priori* for the learning process and only final quantitative parameters for the model's description are extracted from data. The development of rough set methods, however, have shown that the language, or the scheme itself can, and should, be determined by the data rather than assumed *a priori*. We show that this process can be further refined using background knowledge provided by a human expert.

The main challenge here is that an expert's explanations are usually expressed in his own descriptive language, often heavily based on natural language constructs (a *foreign language* L_f), while classifiers are based on a rather low-level language designed to facilitate the computation of physical traits of the images (a *domestic language* L_d). The knowledge passing process hence can be considered as the approximation of expert's concepts such as '*necks*' or '*circular*' by the classifier construction system. From a broader point of view, the expert's *ontology* on the domain is approximated within the ontology used by the computer system. Not only concepts, but the *reasoning schemes* of the expert about investigated samples are carried onto the recognition system as well.

We propose a method for transferring the expert's reasoning scheme into the recognition system, based on the rough mereology approach to concept's approximation [4], [8]. In particular, we show how complex spatial relations between parts of graphical objects can be approximated.

2 Knowledge Transfer

2.1 Foreign Language

An expert is supposed to provide explanations *why*, and perhaps more importantly, *how* he classifies a certain sample. This is denoted in the form of a rule:

$$[CL(u) = k] \equiv \Im(EFeature_1(u), \dots, EFeature_n(u))$$

where $EFeature_i$ represents the expert's perception of some characteristics of the sample u , while synthesis operator \Im represents his perception of some relations between these characteristics.

For example, the expert may express his perception of digit '5' as:

$$[CL(u) = '5'] \equiv a, b, c, d \text{ are parts of } u; \text{'Above_Right'}(a, b); \text{'HStroke'}(a); b = \text{Compose}(c, d); \text{'VStroke'}(c); \text{'WBelly'}(d); \text{'Above'}(c, d),$$

where *Compose* is an assembling operator that produces a bigger part from smaller components.

The above means if there is a west-open belly below a vertical stroke and the two have a horizontal stroke above-right in the sample's image, then the sample is a '5'. '*Above*', '*WBelly*', '*HStroke*' and the like are expert's concepts that may not be readily comprehensible for the classification system and need to be translated into its language.

2.2 Domestic Language

The classifier construction and sample classification algorithms view image samples as sets of black and white pixels. Basic features are constructed using Enhanced Loci coding, which reflects both local and global relations between regions of black and white pixels in the image. For example, each white pixel is assigned a code that denotes whether it has a black neighbor pixel to one of the four directions: North, East, South, West. For a more detailed description of this coding scheme, see [1]. A classifier may be expressed in this domestic language, for example, as:

$$[CL(u) = '0'] \equiv |\{\text{pixel } p: \text{code}(p) = \text{NESW}\}| > 35.$$

2.3 Approximation of Expert's Concepts

Expert's Features. Knowledge from the expert has to be somehow translated from the foreign language into domestic language. For instance, the presence of a concept '*a belly open to the east*' in a sample might be expressed as a following test: $|\{\text{pixel } p: \text{code}(p) = \text{NSW}\}| > 17\% \times |\{\text{all pixels}\}|$.

Since expert's features are expressed with natural language constructs, the notion of '*concept matching*' here must not be crisp, but should allow for a tolerant matching to a degree. A natural choice of tool for this problem is the rough inclusion measure, established as a basis of the theory of rough mereology

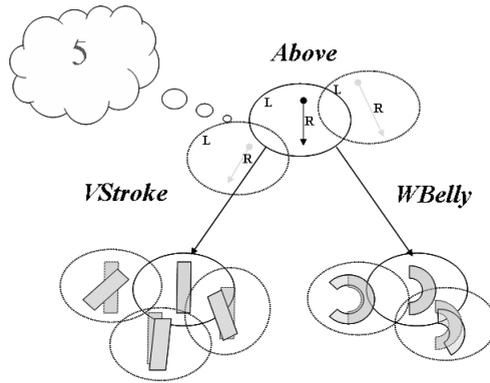


Fig. 1. Tolerant matching by expert

[4,8] to deal with the notion of inclusion to a degree. For instance, a stroke at 85 degree to the horizontal can still be regarded as a vertical stroke, though obviously not a ‘pure’ one (see Fig.1). Instead of just answering in a ‘Yes/No’ fashion, the expert may express his degrees of belief using such terms as ‘Strong’, ‘Fair’, or ‘Weak’.

Relations Between Features. Relations between expert’s features may include concepts such as ‘Above’, ‘Below’ or simply ‘Near’. They express not only expert’s perceptions about particular concepts, but also the interdependencies among them. Similarly to the stand-alone features, these relations can also be described by the expert with a degree of tolerance (See Fig. 1).

The approximation of these relations has been formalized within the framework of perception structures recently developed by Skowron [7]. A *perception structure* S , in a simpler form, is defined as:

$$S = (U, M, F, \models, p)$$

where U is a set of samples, F is a family of formulas expressed in domestic language that describe certain features of the samples and M is a family of relational structures in which these formulas can be evaluated, while $p: U \rightarrow M \times F$ is a *perception function* such that $\forall u \in U : p_1(u) \models p_2(u)$ (p_1 and p_2 are the first and second component projections of p) which means that $p_2(u)$ is satisfied (is true) in the relational structure $p_1(u)$. This may express that some relations among features within samples are observed.

Perception structures, following natural constructs in the expert’s foreign language, can involve tolerant matching. Two relational structures might be considered approximately *the same* if they allow for similar formulas to yield similar results in majority of cases when these formulas are applicable.

Quality Constraints. The knowledge transfer process aims to find sets of patterns Pat, Pat_1, \dots, Pat_n and a relation \mathfrak{S}_d in domestic language such that:

if $(\forall i : Match(Pat_i, EFeature_i) \geq p_i) \wedge (Pat = \mathfrak{S}_d(Pat_1, \dots, Pat_n))$ **then**
 $Quality(Pat) > \alpha$

Quality criteria are introduced so we can generalize the expert’s reasoning to the largest possible number of investigated samples. The requirements on inclusion degrees ensure the *robustness* of the target reasoning scheme, as the target pattern Pat retains its quality regardless of deviations of input patterns. Another important aspect of this process is its *global stability*, which guarantees that if we have some input patterns Pat_i equally “close” or “similar” to $EFeature_i$, then the target pattern $Pat' = \mathfrak{S}_d(Pat'_1, \dots, Pat'_n)$ will meet the same quality requirements as Pat to a satisfactory degree. This leads to an approximation of $EFeature_i$ which is *independent* from particular patterns Pat_i , allowing for approximation schemes that focus on inclusion degrees p_i rather than on a specific input patterns Pat_i .

3 Implementation

3.1 Learning Expert’s Features

In this step, we have to approximate expert’s concept such as ‘Circle’ or ‘WBelly’. The expert will explain what he means when he says, e.g. ‘Circle’, by providing a decision table (U, d) with reference samples, where d is the expert decision to which degree he considers that ‘Circle’ appears in samples $u \in U$. (See Tab. 1). The samples in U may be provided by the expert, or may be picked up by him among samples explicitly submitted by the system, e.g. those that had been misclassified in previous attempts.

Table 1. Perceived features

	<i>Circle</i>
u_1	<i>Strong</i>
u_2	<i>Weak</i>
...	...
u_n	<i>Fair</i>

Table 2. Translated features

	#NESW	<i>Circle</i>
u_1	252	<i>Strong</i>
u_2	4	<i>Weak</i>
...
u_n	90	<i>Fair</i>

We then attempt to find domestic feature(s) that approximates these degrees of belief. This can be conducted using, among other means, genetic algorithms (For example, [3] reported using this kind of tools and methods for a similar problem). In this example, such feature may be the number of pixels that have black neighbors in all four directions (See Tab. 2).

3.2 Learning Expert’s Relations

Having approximated the expert’s features $EFeature_i$, we can try to translate his relation \mathfrak{S} into our \mathfrak{S}_d by asking the expert to go through U and provide

Table 3. Perceived relations

	<i>VStroke</i>	<i>WBelly</i>	<i>Above</i>
u_1	<i>Strong</i>	<i>Strong</i>	<i>Strong</i>
u_2	<i>Fair</i>	<i>Weak</i>	<i>Weak</i>
...
u_n	<i>Fair</i>	<i>Fair</i>	<i>Weak</i>

Table 4. Translated relations

	#V_S	#NES	$S_y < B_y$	<i>Above</i>
u_1	0.8	0.9	(<i>Strong</i> ,1.0)	(<i>Strong</i> , 0.9)
u_2	0.9	1.0	(<i>Weak</i> , 0.1)	(<i>Weak</i> , 0.1)
...
u_n	0.9	0.6	(<i>Fair</i> , 0.3)	(<i>Weak</i> , 0.2)

us with the additional attributes of how strongly he considers the presence of $EFeature_i$ and to what degree he believes the relation \mathfrak{S} holds (See Tab. 3).

We then replace the attributes corresponding to $EFeature_i$ with the rough inclusion measures of the domestic feature sets that approximate those concepts (computed in the previous step). In the next stage, we try to add other features, possibly induced from original domestic primitives, in order to approximate the decision d . Such a feature may be expressed by $S_y < B_y$, which tells whether the median center of the stroke is placed closer to the upper edge of the image than the median center of the belly. (See Tab. 4) Again, this task should be resolved by means of adaptive or evolutionary search strategies without too much computing burden, although it is more time-expensive.

The expert’s perception “A ‘6’ is something that has a ‘vertical stroke’ ‘above’ a ‘belly open to the west’” is eventually approximated by a classifier in the form of a rule:

if $S(\#BL_SL > 23)$ **AND** $B(\#NESW > 12\%)$ **AND** $S_y < B_y$ **then** CL=‘6’,

where S and B are designations of pixel collections, $\#BL_SL$ and $\#NESW$ are numbers of pixels with particular Loci codes, and $S_y < B_y$ reasons about centers of gravity of the two collections.

3.3 Experiments

In order to illustrate the developed methods, we conducted a series of experiments on the NIST Handwritten Segmented Character Special Database 3. We compared the performances gained by a standard learning approach with and without the aid of the domain knowledge. The additional knowledge, passed by a human expert on popular classes as well as some atypical samples allowed to reduce the time needed by the learning phase from 205 minutes to 168 minutes, which means an improvement of about 22 percent without loss in classification quality. In case of screening classifiers, i.e. those that decide a sample *does not* belong to given classes, the improvement is around 60 percent. The representational samples found are also slightly simpler than those computed without using the background knowledge.

4 Conclusion

We presented a general yet powerful framework for the transfer of human domain knowledge to adaptive learning systems. Approximate reasoning schemes

based on rough mereology can be effectively used to incorporate human ontology, expressed in natural language, into processes constructing computable features. We demonstrated that this communication and knowledge transfer process can be conducted in a convenient, interactive manner. We have shown how granular computing, equipped with rough mereology concepts can be effectively applied to highly practical fields such as OCR and handwritten digit recognition. Preliminary experiments conducted showed that presented methods can help improve the learning process and provide a better understanding of the dataset investigated.

Acknowledgment. This work has been partly supported by Grant 3 T11C 002 26 funded by the Ministry of Scientific Research and Information Technology of the Republic of Poland.

References

1. K. Komori, T. Kawatani, K. Ishii, and Y. Iida. A feature concentrated method for character recognition. In Bruce Gilchrist, editor, *Information Processing 77, Proceedings of the International Federation for Information Processing Congress 77*, pages 29–34, Toronto, Canada, August 8-12, 1977. North Holland.
2. T. T. Nguyen and A. Skowron. Rough set approach to domain knowledge approximation. In G. Wang, Q. Liu, Y. Yao, and A. Skowron, editors, *Proceedings of the 9th International Conference: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, RSFDGRC'03. Lecture Notes in Computer Science Vol. 2639*, pages 221–228, Chongqing, China, Oct 19-22, 2003. Springer Verlag.
3. L.S. Oliveira, R. Sabourin, F. Bortolozzi, and C.Y. Suen. Feature selection using multi-objective genetic algorithms for handwritten digit recognition. In *International Conference on Pattern Recognition (ICPR02)*, pages I: 568–571, 2002.
4. L. Polkowski and A. Skowron. Rough mereology: A new paradigm for approximate reasoning. *Journal of Approximate Reasoning*, 15(4):333–365, 1996.
5. L. Polkowski and A. Skowron. Constructing rough mereological granules of classifying rules and classifying algorithms. In B. Bouchon-Meunier, J.Rios-Gutierrez, L. Magdalena, and R.R. Yager, editors, *Technologies for Constructing Intelligent Systems I*, pages 57–70, Heidelberg, 2002. Physica-Verlag.
6. R. J. Schalkoff. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons, Inc., 1992.
7. A. Skowron. Rough sets in perception-based computing. In *First International Conference on Pattern Recognition and Machine Intelligence (PReMI'05)*. Springer-Verlag, 2005.
8. A. Skowron and L. Polkowski. Rough mereological foundations for design, analysis, synthesis, and control in distributed systems. *Information Sciences*, 104(1-2):129–156, 1998.