P. Perner,H. Perner, and B. Müller, Similarity Guided Improvement of the System Performance in an Image Classification System, In: S. Craw and A.Preece (Eds.), Advances in Case-Based Reasoning, ECCBR2002, Springer Verlag, Inai 2416, 2002, p. 604-612.

Similarity Guided Learning of the Case Description and Improvement of the System Performance in an Image Classification System

Petra Perner, Horst Perner, and Bernd Müller

Institute of Computer Vision and Applied Computer Sciences
Arno-Nitzsche-Str. 45, 04277 Leipzig
e-mail: ibaiperner@aol.com; http://www.ibai-research.de

Abstract. The development of an automatic image classification system is a hard problem since such a system must imitate the visual strategy of a human expert when interpreting the particular image. Usually it is not easy to make this strategy explicit. Rather than describing the visual strategy and the image features human are able to judge the similarity between the objects. This judgement can be the basis for a guideline of the development process. This guideline can help the developer to understand what kind of case description/features are necessary for a sufficient system performance and can give an idea what system performance can be achieved. In the paper we describe a novel strategy which can support a developer in building image classification systems. The development process as well as the elicitation of the case description is similarity-guided. Based on the similarity between the objects the system developer can provide new image features and improve the system performance until a system performance is reached that fits to the experts understanding about the relationship among the different objects.

Keywords: Case-Based Reasoning, Image Classification, Learning Vocabulary

1 Introduction

The development of an automatic image classification system is a tricky problem since such a system must imitate the visual strategy of a human when interpreting the particular image. Usually it is not easy to make this strategy explicit so that it can be understood by the system developer and copied into the system. Human can not always name the image features and when they are able to do that these features are symbolic in nature which need to be mapped into the numerical features an automatic image analysis system can extract from the images. Rather than describing the visual strategy and image features human are able to judge the similarity between the objects. This judgement can be the basis of a guideline for the development process. This guideline can help the developer to understand what kind of features are necessary for a sufficient system performance and can give an idea what system performance can be achieved. The later can help to avoid not necessary system tuning

which is often the case when not having understood the dependencies between the objects.

In the paper we describe a novel strategy which can support a developer in building image classification systems. The development process as well as the elicitation of the case description is similarity-guided. Based on the similarity between the objects the system developer can provide new image features and improve the system performance until a system performance is reached that fits to the experts understanding about the relationship among the different objects. In Section 2 will describe the problem which usually arise when building image classification system based on a medical application. In Section 3 we describe the architecture of the System. The similarity-based guidance of the system development is described in Section 4. We summarize the recent results in Section 5 and give conclusions in Section 6.

2 The Problem

We will describe the methodological aspects which will usually arise when developing an image classification system based on an application called HEp-2 cell image analysis. However this application will only be used for demonstration purposes. The problems described are general problems and can be observed when doing other application as well.

The task is to develop an automatic system for the classification of HEp-2 cells in medicine. This kind of cells get used for the identification of antinuclear autoantibodies (ANA) [1]. ANA testing for the assessment of systemic and organ specific autoimmune disease has increased progressively since immunofluorescence techniques were first used to demonstrate antinuclear antibodies in 1957. Hep-2 cells allow recognition of over 30 different nuclear and cytoplasmic patterns, which are given by upwards of 100 different autoantibodies. However, not all patterns occur with the same frequency in practice. There are patterns which occur very often there are other pattern which occur very seldom. The identification of the patterns is recently done manually by a human inspecting the slides with the help of a microscope. The less automation of this technique has resulted in the development of alternative techniques based on chemical reactions, which have not the discrimination power of the ANA testing. An automatic image classification system would pave the way for more wide use of ANA testing. At the beginning of the system development process we have the following conditions:

1. A raw idea of how to interpret this kind of cells was obtained from interviewing a human expert see Figure 2. Based on this expert knowledge we can only achieve an classification accuracy of 25% [2]. This leads to the conclusion that the expert has no good conceptual knowledge built up over time.
2. There is a prototype image catalogue available. A set of digital images could only be obtained after having installed an image acquisition unit with storage capacity at the clinical hospital. This allows to collect images as soon as they appear in medical practice. However, this does not ensure to get a large collection of

images with a equally distributed number of samples in each class. There are some classes they did not appear until now although the system is installed since a year. This circumstance is usual for medical applications. Therefore we need a strategy for the system development which allows to develop an automated image classification system under this circumstances.

3. The expert description in Figure 1 allows to select basic image feature descriptors and the necessary image processing facilities which get installed into the system.
4. The experts are able to judge the similarities between the different classes on a scale between 0...1 (0 stands for identity and 1 stands for dissimilar). That gives a similarity matrix for the classes see Figure 2.

These conditions usually exist in practice and do not allow a system set up based on conventional pattern recognition methods. In the following we will describe our strategy for the system development under these conditions.

4 Architecture of the System

The cases are image descriptions which are automatically extracted from the images based on the procedures installed in the feature extraction unit and stored into the case base together with the class name. In a separate image data base are kept all images, class names, and image descriptions given by an human operator for later evaluation purposes.

The feature extraction unit contains feature extraction procedures that have been identified based on the interview with the expert. We should note here that a particular application requires special feature descriptors. Therefore not all possible feature extraction procedure can be implemented into such a system from scratch. But we hope that we can come up with a special vocabulary and the associated feature extraction procedures for our cell image application which can be applied to a wider range of cell types.

For each case class are selected a prototype image from the image catalogue. A special unit observes how the collected cases deviate from the prototype image in the class. If the difference between the prototype image and a new collected image is too large the system gives an alert to the system developer. He checks the image together with the medical expert and if necessary he can install a new prototype into the case base.

From the prototypes are calculated the pairwise similarities based on the calculated image features. That gives a second proximity matrix besides the on based on the experts judgement. Now, the task is to tune the system by the system developer so that the two similarity matrix become identical. Identical similarity matrix are certainly the optimal case. A more relaxed version would be to ask for minimal distance between these two matrix.

Besides the system development it should be possible that the recent system can be used by the human expert. Therefore, the cbr process is using the case base with the prototype images for reasoning. We are considering the case base to have a flat structure. A new case is classified by first extracting the case description from the

image based on the recent established image analysis and feature extraction procedures and afterwards using this description for the determination of the closest case among the cases in case base. The resulting answer of the system is observed by a human operator who is an expert in the field of application. He criticizes the result.

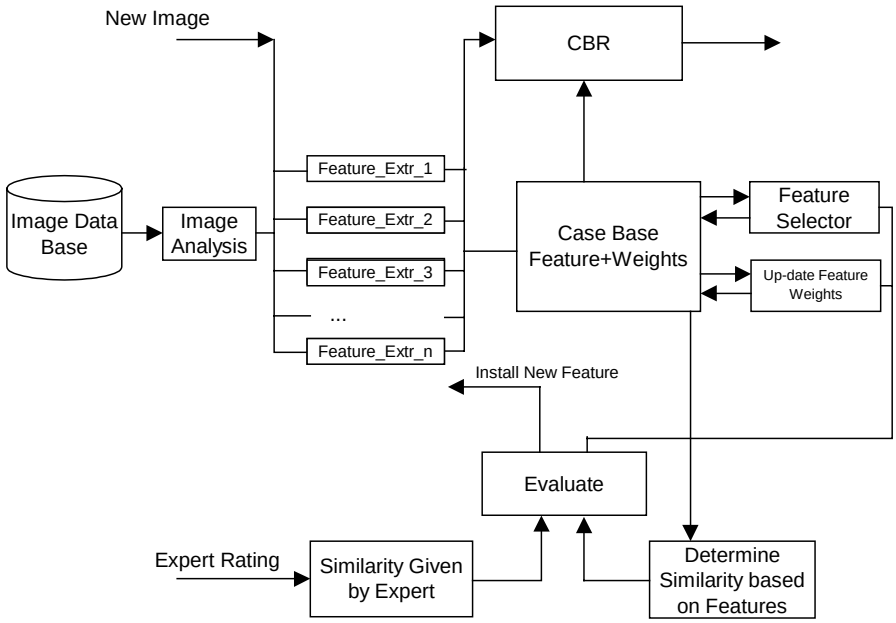


Fig. 1. Architecture of the System

3 Similarity-Guided Improvement of the System Performance

We think that the intuitively given rating of the similarity between the different objects given by an expert can help to improve the development process. This rating gives the developer an idea what system performance can be achieved based on the automatically extracted image features and can guide the development process as well as the elicitation of new features.

Usually for that purposes is used the error rate of the classifier. However, this gives only a general understanding of how good the system performance is but does not give a hint what should be done to improve the system performance.

The aim should be to minimize the difference between the user given similarity rating sim_E of the prototype objects o_i and the similarity sim_S between these prototypes derived based on the calculated image features f_k:

$$sim_E(o_i, o_j) - sim_S(f_{ik} - f_{jk}) = Min!$$

The optimal case would be zero difference between the experts rating and the similarity that can be achieved by the system based on objective calculation of image features from the images. However, this is hardly to achieve. There are three possible solutions to this problem:

1. Put feature weights to the features and up-date the initially feature weights until a sufficient difference between sim_E and sim_S is reached. Since we know from the expert´s similarity where the vast difference is we can update the weight locally according to that.
2. Select among the whole set of features the right subset of features for the classification problem. This is a feature selection problem but has also a link to the feature weight problem. Those features which are not necessary get the feature weight zero and those which are necessary get the feature weight one.
3. If solution 1 and 2 do not give sufficient results then only the identification of new features will bring the expected result.

Solution 1 and 2 improve the system performance based on the automatically extracted images features. If the resulting performance is not sufficient enough then only the identification of a new image feature can help to improve the system performance. This process can be guided by the pairwise similarity between the objects as well. When the system gives a pairwise similarity measure sim_s that is less than the experts given similarity sim_E than the question is: What feature makes this two objects different? This is the point where the system developer has to identify a new feature descriptor and install this feature descriptor in the system.

5 Results

From the expert knowledge shown in Figure 1 we can derive that the expert describes the texture of the cells. Therefore we have developed a novel and very flexible texture descriptor based on random sets [3]. This texture descriptor is based on 200 low-level features. Depending on how they are combined they can describe different high-level concepts of texture. This texture descriptor gives a high flexibility which allows to describe different and very complex textures of cell objects. Besides that we have installed feature descriptors such as size of the object and the contour.

Our study has shown that experts are able to say what objects are similar but in order to get unbiased ratings the system has to support the expert. The elicitation of the experts similarity judgement is done by a specific questionnaire. From the whole set of images are randomly selected 3 images and shown on display to the expert. Then the system asks him to select two images which are similar to each other. After having chosen these two images he has to rate the similarity between these images on a scale between 0...1. Table 1 shows the similarity matrix given by the expert.

The similarity matrix obtained based on the extracted image features is shown in Table 2. The similarity measure is the Euclidean distance.
Recently we have installed a feature selection procedure. This procedure tries to identify the features that have the highest dissimilarity to other features. Therefore we calculate the similarities between the features. Feature subset selection was done by hierarchical clustering with single linkage and using Euclidean distance. The resulting dendrogram showed groups of similar features. The dendrogram was then decomposed into groups of similar features by a cut-off value of 10%, 5% and 2.5% similarity. From the remaining groups one feature for each group was selected and a feature weight of one was associated to this feature. All the other features get the feature weight zero. Based on this strategy we could get the improvement shown in Table 3.
However, no further improvement was possible. Now arises the question what makes the objects dissimilar that are very similar based on the calculated features but not on the experts judgement. Together with the expert we discovered new images features based on the spatial relation among the objects inside a cell. Therefore, we are currently going on to develop a new feature extraction procedure for this kind of image description. This will enhance our vocabulary for cell image description. At the end we will come up with a set of basic image descriptors which can be used for other types of cells than HEp-2 cells such as for e.g. pap-smear cells. Then, the expert itself can chose among the whole set of image descriptors the one which fits to his observation and advise the system to test this descriptor. By doing so he can learn the vocabulary for his specific application.
Besides that development of new image descriptors we will install a procedure for learning the feature weights [4].

	homogen	homo. fein gespr.	nukleolär	fein gespr	fein gespr. Nukl.	centromere
	100000	100320	200000	320000	320200	500000
100000	0	0,2	0,3	0,8	0,3	0,6
100320	0,2	0	0,4	0,8	0,2	0,6
200000	0,4	0,3	0	0,8	0,2	0,6
320000	0,2	0,2	0,4	0	0,8	0,6
320200	0,3	0,2	0,2	0,2	0	0,6
500000	0,6	0,6	0,6	0,6	0,6	0

Table 1. Similarity of Classes given by the Expert

	100000	100320	200000	320000	320200	500000
100000	0	0,395	0,174	0,811	0,377	0,42
100320		0	0,475	1	0,524	0,872
200000			0	0,515	0	0,222
320000				0	0,428	0,594
320200					0	0,433
500000						0

Table 2. Similarity between the Objects before Feature Subset Selection

	100000	100320	200000	320000	320200	500000
100000	0	0,403	0,22	0,868	0,391	0,480
100320		0	0,617	0,95	0,622	0,897
200000			0	0,699	0,005	0,274
320000				0	0,540	0,784
320200					0	0,445
500000						0

Table 3. Similarity between the Classes after Feature Subset Selection

6 Related Work

Our work has a link to conversational CBR [5][6] where the user is engaged to incrementally formulate a query duing case retrieval. However, in our appraoch it is used to identify a gap in case description and to encourage the user to think of a new feature for which then a new feature extraction procedure must be developed.

The work described here has also a link to case base maintenance [7] which is the process to refine a CBR system's case base to improve the system's performance. Learning feature weights [8] and feature subset selection are the main topics related to this task in our approach.

7 Conclusion

We have proposed a similarity-guided strategy for the system development of an image classification system. This strategy can help to elicitate new features as well as feature weights and by doing so it can help to improve the system performance. Using the expert´s judgement on similarity and comparing that with the similarity calculated by the system gives the system developer an idea what system performance can be achieved. This can help to avoid not necessary tuning of the system.

We are currently implementing our strategy into a system for HEp2-cell image classification. The first results are promising and support our strategy. However, there are a lot of open problems which have not necessarily to do with CBR. The expert needs better support for the determination of the similarity between the objects. Experts are able to say what objects are similar but they are not trained to express the similarity on a scale between 0...1.

We have also to find new feature descriptors that can describe the image features. This is a hard problem since there are not so many work been done in finding feature extractors for cell image classification.

Besides that we want to implement a methods for feature weight learning in our system that can learn global and local feature weights.

Acknowledgement

This research has partially been founded by the German Mininstry of Economy and Technology within the project LERNBILDZELL.

References

1. K. Conrad, R.-L. Humbel, M. Meurer, Y. Shoenfeld, E. M. Tan, eds., Autoantigens and Autoantibodies: Diagnostic Tools and Clues to Understanding Autoimmunity (Pabst Science Publisher, Lengerich, Berlin, Riga, Rom, Wien, Zagreb, 2000).
2. P. Perner, H. Perner, and B. Müller, Mining Knowledge for Hep-2 Cell Image Classification, Journal Artificial Intelligence in Medicine 2002, (to appear)
3. P. Perner, Texture Classification based on the Random Sets and its Application to Hep-2 Cells, IEEE Proc. International Conference on Patter Recognition, ICPR´2002, to appear
4. D. Aha and R.L. Goldstone, Concept learning and flexible weighting, In Proc. of the Fourteenth Annual Conference of Cognitive Science Society, Bloomington, IN: Lawrence Erlbaum 1992, pp. 534-539.
5. D.W. Aha, T. Maney, and L.A. Breslow, Supporting Dialogue Inferencing in Conversational Case-Based Reasoning, In: B. Smyth and P. Cunnigham (Eds.), Advances in Case-Based Reasoning, Springer Verlag, lnai1488, p. 262-272, 1998.
6. D. McSherry, Interactive Case-Based Reasoning in Sequential Diagnosis, Applied Intelligence 14 (2001), 65-76
7. G. Cao, S. Shiu, and X. Wang, A Fuzzy-Rough Approach for Case Base Maintenance, In: D. Aha and I Watson (Eds.), Case-Based Reasoning Research and Development, Springer Verlag, lnai 2080, pp. 118-130, 2001
8. S. Craw, J. Jarmulak, and R. Rowe. Maintaining Retrieval Knowledge in a Case-Based Reasoning System, Computational Intelligence, vol. 17, No. 2, 2001, pp. 346-363.

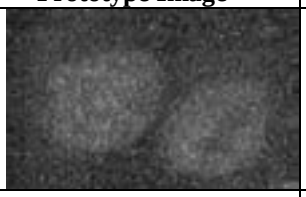
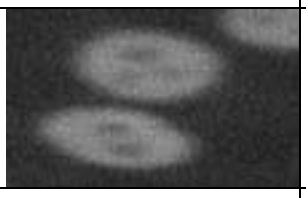
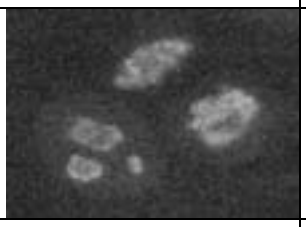
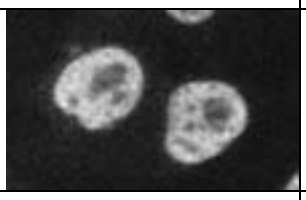
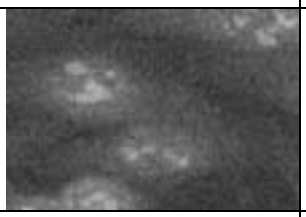
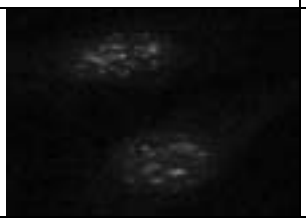
Class_Name	Nomenclature	Prototype Image	Description
Homogen	100 000		A uniform diffuse fluorescence of the entire nucleus of interphase cells. The surrounding cytoplasm is negative.
Homogen Fine Speckled	100 320		A uniform fine speckled fluorescence of the entire nucleus of interphase cells.
Nuclear	200 000		Smooth and uniform fluorescence of the nuclei Nuclei sometimes dark Chromosomes fluoresced weak up to extreme intensive
Fine Speckled	320 000		Fine to discrete speckled staining in a uniform distribution.
Fine Speckled Nuclear	320 200		Dense fine speckled fluorescence Background diffuse fluorescent
Centromere	500 000		Nuclei weak uniform or fine granular, poor distinction from background

Fig. 2. Class Names, Prototype Images and Expert Description