Classifying the unknown: The T-cell Recruitment Mechanism of the Immune System as a Paradigm for Pattern Recognition

M. Köppen, M. Stellmacher, L. Lohmann and B. Nickolay
Department of Pattern Recognition,
Fraunhofer IPK-Berlin
Pascalstr. 8-9, 10587 Berlin, Germany
Email: {mario.koeppen|stellmacher|lohmann|nickolay}@ipk.fhg.de

Abstract

The immune system of animals is considered an adaptive system, capable of solving a classification task: to distinguish microorganisms, which take part in the body metabolism, into self and non-self ones. Recent research results of immunobiologists, esp. the discovery of the Major histocompatibility complex (MHC), are discussed from an information theoretic point of view. One of the most interesting aspects is the training of a two-class problem by the only provision of examples of one class. By adopting the basic processes, which maintain the organisation of the immune system, new algorithm esp. for pattern recognition can be designed. As an example, a framework is proposed for character separation of melted char images, what is an important problem in optical character recognition.

Keywords: Immune Algorithms, T cell Recruitment, T Cell Recognition, Optical Character Recognition, Character Separation

1 Introduction

The study of natural systems, which maintain the ability to classify, has been mostly occupied by the study of the nervous system of higher animals, especially the human brain. However, there are other classification systems in nature, as well. As an example, the carnivorous plant *Dionaea muscipula* (venus fly trap) has to distinguish, whether an crawling insect has triggered its three capillary hairs or a drop of rain. The system here is quite simple: when two hairs are triggered simultaneously or the same hair is triggered twice within a short time intervall, the venus fly trap will close its leave. A more complicated example is the immune system of animals (not only higher ones). The fundamental classification task of this system is to distinguish microorganisms into two classes: self and non-self.

Due to several reasons, the immune system has found particular interest in medical research in the last decade. The rapid spread of the Acquired immunodeficiency syndrome (AIDS), the progress in organic transplantation techniques and the increasing amount of allergical symptoms among the population of industrial countries are problems related to the understanding of the immune system. A lot of fundamental research results have sharpened the picture of the processes involved in the immune system's actions and reactions, and most of them proved to be quite complex.

The most intriguing point is that the immune system is a learning system, capable of adaptation on varying environmental conditions. But, in contrast to the higher nervous system, it is not capable of cognition. Its only task is to eliminate foreign microorganisms, which entered the body and may threat the bodies metabolism and viability.

As for neural networks in the 60s, the immune system was studied as a model for adaptive procedures since the beginning of the 90s. Three purposes of such studies can be noted:

- gain a better understanding of the immune system,
- design of new adaptive algorithms and
- apply the paradigms of the immune system in technical security systems.

The first item is quite important for medical research. By using the information theoretic description of processes, which are involved in the immunoresponse, the justification of the relative importance of all processes can be obtained. Especially genetic algorithms have been considered a

valuable model for the important processes [FJSP93] [HFP95].

The second item follows the line of the development of soft computing algorithms in general: as evolutionary computation was derived from the study of the natural evolution, and neural networks from the nervous system, algorithms can be derived from the study of the immune system processes as well. So far, the most inspiring sources have been the T-cell recruitment mechanism [BV91] and the cooperative behaviour of antibodies [SKS99]. These studies and their relation to the theoretical understanding of soft computing algorithms (as e.g. the schema theorem for genetic algorithms) will give a contribution for a better understanding of the immune system, too.

The third item mentioned above was considered recently. So, in [SHF97] and [FHS97], there the authors start to consider the immune system a model for the design of a computer security system. By monitoring processes, a system becomes able to detect computer viruses. This fundamental idea opens a gate for a completely new research direction in system design. The results could be applied to other complex security systems (as in biometrics or digital watermarking) as well.

Up to now, the study of algorithms and systems inspired by the immune system network are just at the beginning of the development of a prospective fruitful new research direction. Further results in the medical research will enhance that development.

From the pattern recognition point of view, the immune system is able to distinguish two classes of microorganisms: self and non-self ones. The detection of foreign microorganisms is of vital importance. Since a living system is an open system according to its metabolism, there is a continuous exchange of substances with its environment. Among the substances entering the living system, there may be threatening ones like toxic substances or viruses. The immune system acts on different levels to prevent from such situations. On its top, there are the skin and several mucous membranes, doing a very rough prevention. Other basic mechanisms are blood clotting, attack of foreign bacterias by B lymphocytes and, what is in the scope of this paper, the T cell reception of virally infected cells (see fig. 1). Nearly each of those subsystems performs its classification task in its own manner. Especially the lymphocytic systems are adaptive systems, with the ability to become trained. The problem of the underlying information processing is: the immune system has to be trained on a two-class problem, without any examples for one of the two classes! Since the presence of toxics or viruses would cause a danger for the living system, adaption has to be performed by the presence of *self* representing examples alone.

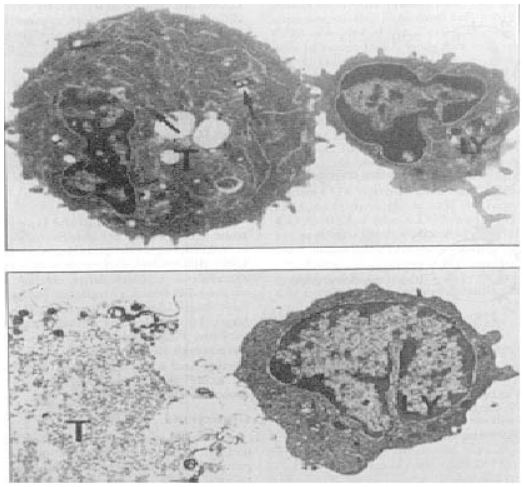

Figure 1: A virally infected cell is recognized by a T cell and destroyed.

A similar problem can be stated for pattern recognition: how is it possible to learn to distinguish objects of class A from objects of class B with just a stock of examples for class A? How is it possible to learn to classify or to become prepared for classifying the unknown?

The problem, formulated this way, is mathematically ill-posed. There is always a chance for deception, as stated by the fundamental "No Free Lunch" theorem [WM95] and the "Ugly Duckling" theorem [Wat85]. Also, the immune system fails in some cases, as can be quickly exemplified by the rapid spread of viral diseases or allergies as immunic reactions on harmless substances (which can be considered pathogenic self-immune responses).

However, non-optimality of a system does not prevent the system to be able to cover a reasonable set of threatening situations, and this is, what the immune system actually does. It is related to the optimal employment of a priori knowledge without any AI concepts, since there is no "intelligence" in the immune system.

In this paper, a framework is presented, which takes the so-called *immune recruitment mechanism* [BV91] as a paradigm for the classification of patterns. The purpose of the framework, which is exemplified for the case of optical character recognition, is to solve the task of distinguishing known character images from character-like patterns, as ligatures (i.e. meltings of two or more character images into one image). By a one-to-one correspondance of the

framework with the processes that establish the T cell recruitment, an immune inspired procedure is given. This procedure is important for the understanding of biological T cell recruitment as well.

The T cell recruitment has been considered in [BV91], too. However, there are two important issues considered different in the following discussion.

The first issue is the difference of the goal of the framework, which was presented in [BV91] and the one, which is proposed here. While in [BV91] the task of T cell recruitment was formulated as the provision of a sufficient amount of different T cells for keeping up the lymphocytic system in an active level, the task here is considered as the provision of enough variablity among the T cell to become prepared for the recognition of possible antigens, which enter the living system. However, for both frameworks, the thymic selection procedure is considered as the basic step in achieving both goals.

The second issue is related to the progress in medical knowledge. In 1991, the MHC-antigen complexes and the dual selection processes in the thymus were not known. However, for a proper understanding of the T cell response this should be considered essential. The framework, which is presented here, acknowledges these new research results.

Section 2 will recall the T cell recruitment and T cell response system from an information theoretic point of view. Since a complete description of the whole system would go far beyond the scope of this paper, the focus is done on the points, which are essential for the framework description. A lack in medical correctness might be the inescapable consequence of that shortening. Among many, consider [Dav97] a comprehensive textbook on animal immune systems.

Then, section 3 will describe the proposed framework and study an example of its application to character recognition. Section 4 will conclude this paper with a short discussion.

2 The T cell recruitment and response system

T cells are vitally responsible for the detection of virally infected cells. Essentially, three components take part in the T cell response (see fig. 2).

Antigens are substances, which are processed by the T cell reception system. There are antigens causing immune response (immunogens), antigens causing no response (since not strongly related to threatening situations) and the so-called selfantigens, i.e. harmless substances of the body itself. Antigens in

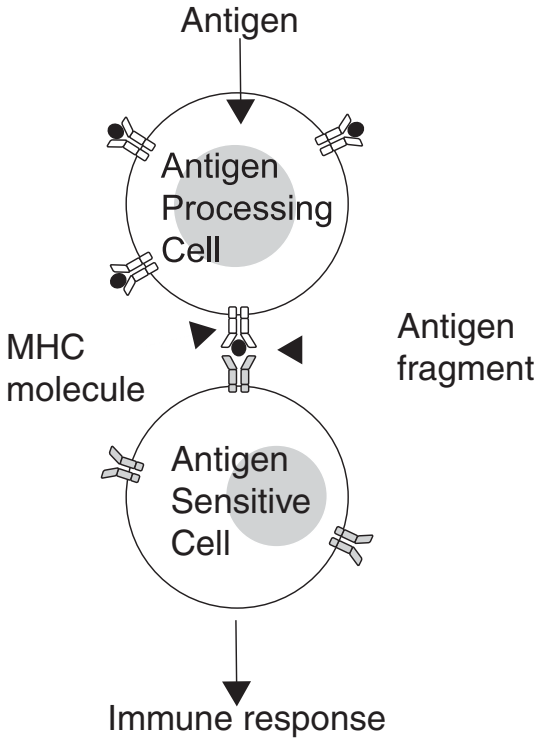

Figure 2: Basic triggering of T cell reception by MHC antigenpeptid presentation.

the cytoplasma of a cell are presented on the surface of cells by means of the *major histocompatibility complex* (MHC) molecules (there are MHC I and MHC II molecules, but the difference is not considered here). For doing so, the antigen is broken up into antigenpeptids, i.e. shorter peptide fragments. Each MHC, located within the endoplasmotic reticulum, is roughly able to recognize a group of such antigenpeptids. Roughly here means that a MHC includes a molecular pocket for embedding one peptid out of a certain group of peptids.

When a MHC carries an antigenpeptid, it moves to the outer surface of the cell. By doing so for several MHC-antigenpeptid complexes, a cell presents something like a "table of contents" of its internal metabolism to the outer world.

T cells, which are passing by a cell, can try to dock on the MHC complexes on the surface of the cell. This will happen, when several conditions are fulfilled (see fig. 3). The most significant one is the fittening of the molecular pocket of the α and β chain, a unique structural property of each T cell, with the antigen presented by a MHC. Also, the T cell must be able to dock on the MHC by means of the so-called co-receptors DC8 (for MHC I molecules) or DC4 (for MHC II molecules). There are some other

costimulatory signals necessary in order to put a T cell into action, which will not be considered here.

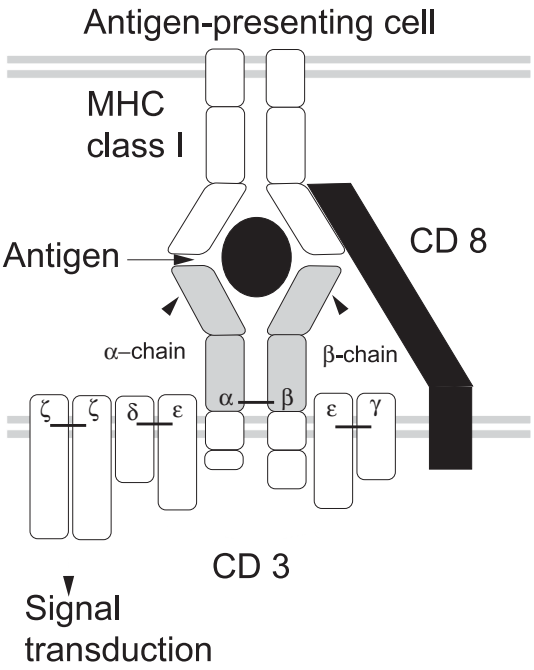

Figure 3: Docking of a T cell to a MHC molecule by matching of antigenpeptid and co-stimulatory responses (MHC I).

After successfull docking on a cell, a class switch occurrs and the T cell is proliferated and initiates a complex chain of reactions, which finally lead to the death of the target cell.

One important point is that the MHC molecules are specific for each living being, but sensitive for a group of antigenpeptids, while the T cells are sensitive for only one antigenpeptid. Hence, the body must provide a large number of different T cells in order to be receptive for a large number of possible antigens.

The design of a T cell is the most interesting point. The genetic coding of T cell generation is different from the general protein production system of the DNA, since it includes ambiguity in the chromosome transcription process. The protein sequences of the α and β chain are generated in a more random fashion by mixing of several chromosome parts. Hence, the diversity of T cells is genetically coded.

T cells are generated in the marrow mark and then transferred to the thymus (as so-called thymocytes), where they will stay for a while. The task of the thymus can be compared to the recruitment of soldiers for a war. Only T cells (i.e. thymocytes), which prove to be able to perform a useful task in the lymphocytic system, are able to pass the thymus and enter the lymph.

There are two selections going on in the thymus, a positive one and a negative one.

The positive selection will only keep lymphocytes, which are able to dock on a body MHC by its co-stimulatory receptors (CD4 or CD8). So, this will select only those T cells, which are able to understand the "language" spoken by the cells to present their metabolism.

Then, a negative selection takes place, which eliminates all T cells, which are docking on the complex of a MHC and a selfantigen. This prevents the immune system from autoimmunity.

Only T cells passing both selections are able to take part in the immune response system. But it has to be mentioned that the lymphocytes, which leave the thymus, has still to become activated later on. This is due to the later presentation of self-antigens, which are not available within the thymus. Also, some undisclosured process of T cell elimination takes also place after that.

This is the manner, by which a living system solves the problem to classify the unknown. It uses the advantage of a common language of normal behaving cells and viruses as well (the processing of peptid chains); it maintains a set of schemes (by about a dozen different MHC molecules) of this language; and it generates a vast amount of different "language speaking" T cells, each of which is quite sensitive to exactly one antigenpeptid, which is not found within a proper working cell.

3 The immune based framework for robust character separation

In the following, a framework is presented, which is based on the same paradigms as the T cell recruitment and reception mechanism.

3.1 Character separation

In character recognition applications a correct and robust character separation process is very important. Most of the Optical Character Recognition (OCR) procedures are only able to recognize well-separated character images. Two or more fused characters (e.g. as a result of noise on the carrier medium) can not be classified as a proper element of one character class. Especially in the outdoor area, relevant characters are fused together quite often.

This problems are usually treated by measuring a small number of character stroke pixels or by disturbance of the character objects. For instance, rusty screws between the characters fuse numbers and letters of car number plates.

In approaches to the OCR-processing of recognition systems, separation tasks are running in parallel. Actually, dissectional, recognition based, hybrid and holistic methods are used [CL96] [SP96] [PSH+93]. Very important for the success of the separation task is to decide correctly whether the characters are properly separated or not. Together with the development of a robust licence plate recognition system, a *merge decider network* was proposed in [Loh99].

This neural network structure is able to decide between individual and fused characters in the input image. The disadvantage of the merge decider method is the requirement of a representative training set of individual and fused characters (ligatures). The proposed immune based framework is expected to simplify this approach. Following the traces of the T cell recruitment mechanism, only representative training images of *single character* images should be required.

3.2 Overview of the architecture

The schema of the immune based framework is given in figure 4. Recruitment and reaction stage are separated.

3.2.1 The recruitment stage

The purpose of the recruitment stage is the selection of the randomly initialized templates. This means that the recruitment of the correct templates takes place in order to discriminate coarse versions of the correct individual characters. In its biological meaning the lymphozytes, which are able to respond to self MHCs, are selected from thymozytes. In agreement with the immune system recruitment processes the recruitment stage is separated in the positive and negative selection task.

The positive selection task detects the templates (thymozytes) which positively react on "coarse" features of the correct individual character images. The used matching procedure is based on binary Hamming distance method. As inputs, sample images of individual characters are required.

The result of the negative selection task are the templates (lymphozytes), which are able to detect ligatures (body outside cells). In the matching process, the best matching templates are eliminated.

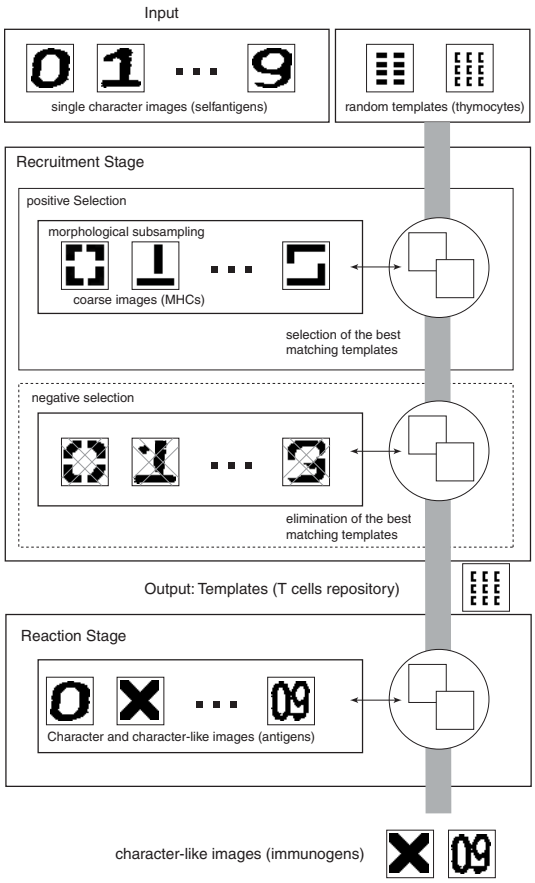

Figure 4: Immune based framework.

The mask character images combine coarse character features with refined character features. A cross mask was used in the proposed framework.

3.2.2 The reaction stage

The reaction stage includes a classification process of individual or merged character images. In the biological sense this is the detection of immunogens. Based on positive and negative selection, the generated templates detect ligatures and other character-like images.

4 Discussion

The presented framework is different from other character separation approaches in the sense that it learns to react on patterns which are different from the patterns from which it was trained. The recently presented *merge decider network* is trained from a set of single character images for one class, and with a set of merged character images for the second class.

The problem is the large variability of merged character images, which are necessary for training the network. The immune based approach will help to reduce the number of necessary examples and generate the templates, which can be either used for template matching, or for initialising the merge decider network.

An important issue is the relation of the proposed framework to an evolutionary approach. The procedure could simply be made re-entrant, hence changing to a population based approach. However, from first experiments there seems to be no need for this procedure, especially for the chosen application example.

Another issue is related to the high number of templates, which are needed in the online phase of the framework. A large number of different T cells, each of which being sensitive to a different virus, has to be handled by the animal immune system as well. Of course, this is strongly related to the dynamics, the triggering and the spatio-temporal organization of the distribution of the T cells within the body. The proposed framework has to be extended in order to provide T cells where and whenever they are needed.

References

[BV91] Hugues Bersini and Francisco J. Varela. The immune recruitment mechanism: A selective evolutionary strategy. In Richard K. Belew and Lashon B. Booker, editors, *Proc. of the Fourth Int. Conf. on Genetic Algorithms*, pages 520–526, San Mateo, California, 1991. Morgen Kaufmann.

[CL96] G.R. Casey and E. Lecolinet. A survey of methods and strategies in character segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:690–706, 1996.

[Dav97] Huw Davies. *Introductory Immunobiology*. Chapman & Hall, 1997.

[FHS97] Stephanie Forrest, Steven A. Hofmeyr, and Anil Somayaji. Computer immunology. *Communications of the ACM*, 40(10):88–96, 1997.

[FJSP93] Stephanie Forrest, Brenda Javornik, Robert E. Smith, and Alan S. Perelson. Using generic algorithms to explore pattern recognition in the immune system. *Evolutionary Computation*, 1(3):191–211, 1993.

[HFP95] Ron R. Hightower, Stephanie Forrest, and Alan S. Perelson. The evolution of emergent organization in immune system gene libraries. In L. J. Eshelman, editor, *Proc. of the Sixth Int. Conf. on Genetic Algorithms*, pages 344–350, San Francisco, CA, 1995. Morgen Kaufmann.

[Loh99] Lutz Lohmann. *Bildanalysesystem zur robusten Erkennung von Kennzeichen an Fahrzeugen*. PhD thesis, TU Berlin, 1999. (in German).

[PSH+93] B. Plessis, A. Sicsu, L. Heutte, E. Menu, E. Lecolinet, O. Debon, and J.-V. Moreau. A multi-classifier combination strategy for the recognition of handwritten cursive words. In *Proceedings of the 2nd Intl. Conference on Document Analysis and Recognition, ICDAR'93, Tsukuba City, Japan*, pages 415–427, 1993.

[SHF97] Anil Somayaji, Steven Hofmeyr, and Stephanie Forrest. Principles of a computer immune system. In *Proc. of the New Security Paradigms'97*, pages 75–82, Great Langdale, Cumbria, United Kingdom, 1997.

[SKS99] Toshiyuki Shimooka, Yasufumi Kikuchi, and Kouichi Shimizu. Study on the dynamics of the idiotypic network for the feature extraction of patterns. In *Proc. of Int. Workshop on Soft Computing in Industry'99 (IWSCI'99)*, pages 313–318, Muroran, Hokkaido, Japan, 1999.

[SP96] H. Shi and T. Pavlidis. A system for text recognition based on graph embedding matching. In *Proceedings of the Workshop on Document Analysis Systems, DAS'96, Philadelphia, PA*, pages 415–427, 1996.

[Wat85] Satosi Watanabe. *Pattern Recognition: Human and Mechanical*. John Wiley & Sons, New York u.a., 1985.

[WM95] David H. Wolpert and William G. Macready. No free lunch theorems for search. Technical Report SFI-TR-95-02-010, Santa Fe Institute, 6. Februar 1995.