Mário João Gonçalves Antunes

An artificial immune system for anomaly detection based on dynamic tunable activation thresholds

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2011

Mário João Gonçalves Antunes

An artificial immune system for anomaly detection based on dynamic tunable activation thresholds

Tese submetida à Faculdade de Ciências da
Universidade do Porto para obtenção do grau de Doutor
em Ciência de Computadores

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2011

Abstract

Nature has been a bountiful source of inspiration for the development of novel algorithms and computational models. Computer scientists and engineers together with biologists, have for some time been studying nature under the new light provided by the computer simulation of complex biological processes. The level of comprehension and the rhythm of new tested scientific hypothesis made possible with this computer revolution, has accelerated the process of scientific discovery in some application domains to the level where it is now possible to describe some of their complex biological processes fundamental properties in terms of highly detailed concrete abstract models. This has been accomplished in such a way that it is now possible to employ these theoretical functional models as abstract digital information processing entities that can be used to tackle difficult problems in hitherto unrelated problem domains.

These problems are not always related to the original biological domain, but their solution often depends on highly complex cognitive and learning strategies that are usually very difficult to capture and implement within more traditional algorithmic approaches. For example the inspiration taken from the computational study of the brain and on the differentiation of species provided by the theory of evolution are amongst the most relevant studies of biological systems by computers, that lead respectively to the highly successful development of Artificial Neural Networks (ANN) and Genetic Algorithms (GA) computational models.

The development of abstract models explaining the behaviour of the vertebrate immune system is another example of a vibrant area of research and an appealing source of inspiration on the development and deployment of innovative and adaptive computational anomaly detectors in the form Artificial Immune Systems (AIS). These are mainly the result of the development of new immune inspired computational models and novel algorithms implementing and embodying what is now known about the Vertebrate Immune System (IS) intrinsic cognitive and learning mechanisms for self tolerance and non-self discrimination processes. The primary goal of the resulting anomaly detection AIS is the ability to successful distinguish ongoing anomalous and harmful environment activities that may compromise part or the whole system being monitored, from those environment activities that presumably correspond to benign *"normal"* behaviour.

Negative selection and danger theory are well known classical functional models that try to explain some facets of the behaviour of the immune system, which have positively inspired the development of some successful AIS for anomaly detection. In spite of the great contributions and promising results obtained so far in a wide range of application domains, the research in AIS for anomaly detection is still faced with many difficult challenges. First,

there is a need to base the AIS implementations on a more complete understanding of the IS self/non-self discrimination capabilities. In the future, more complete and powerful AIS could be the result of "assemblages" of diverse abstract models of many biological processes, some of which have not yet been sufficiently understood. Second, and this is one of the main recurrent themes of this thesis, is the need to evaluate and integrate into this future "assemblage" other not yet so popular immunological theories that can explain a particular facet on the way the IS copes with environmental changes (*concept drift*) throughout time. Finally, we also touch on the difficulties arising from the transposition and application of the developed abstract computerised IS models into other non biological domains.

The results we infer and the more general observations we can make from the application of our AIS framework into non biological domain problems are also immensely valuable and can be used as an alternative validation feedback mechanism for the original theoretical functional models. The behaviour of these models in the digital domain can also help to shed some light into some of the fundamental behavioural properties of the original biological phenomenons under study, thus providing stronger scientific support and evidence of their validity.

In this thesis we propose a novel AIS framework based on the Tunable Activation Thresholds (TAT) immunological theory to tackle some of these challenges. We start by describing how TAT helps to explain the observed biological discriminative behaviour of a particular immune cell, the T-cell, and then proceed to adapt and evaluate it in the context of two highly specific, non biological, information processing domains: anomaly detection and text classification. The AIS framework components we have developed together with their underlying immune counterparts, were kept as faithful as possible to the original TAT immunological model, in order to keep the digital validation feedback mechanism for the original biological model open and as strong as possible. The generic AIS framework, termed TAT-AIS, is highly modular and its core is composed by a TAT dynamics simulator for a population of abstract T-cells.

We have evaluated TAT-AIS in the context of two distinct application domains. In network intrusions detection we have tested and compared it against the widely used and popular `snort` signature-based intrusion detection system. Both systems have been compared using the same real world based network traffic, comprised of real network packets related to normal application and user behaviours. These have all been collected from a local network and we have also populated the testing data sets with small random bursts of previously validated and catalogued malicious activity. The results thus obtained have shown great promise on the TAT-AIS ability to deal with temporal dynamic changes in the nature of the

network traffic. Unlike `snort`, TAT-AIS was able to detect some new unseen attacks and, at the same time, to gradually and gracefully tolerate new previously unseen network packets corresponding to normal network activity.

Regarding text classification, we have developed a two-tier hybrid learning committee based ensemble for classification, composed by both a Support Vector Machine (SVM) classifier and TAT-AIS. The experimental results have been obtained with the popular Reuters-21578 benchmark and have shown that it is possible to produce classification committees based ensembles with TAT-AIS that are able to outperform the results obtained by each classifier individually, on the same data.

The problems addressed in this thesis also deal with one of the most important recurrent challenges in machine learning, that is the detection of behavioural drifting episodes in continuous but dynamically changing streams of data. We believe that our TAT-AIS framework constitutes a versatile tool that provides a valid contribution to the increasing arsenal of algorithms and utilities already in use to cope with this difficult problem. At the end of the thesis we present some new lines of research that resulted from the work conducing to this dissertation and that we intend to pursue in the near future.

Resumo

A natureza tem sido uma generosa fonte de inspiração para o desenvolvimento de novos algoritmos e modelos computacionais. Os cientistas e engenheiros da computação, juntamente com os biólogos, têm há algum tempo estudado a natureza através da simulação em computador de processos biológicos complexos. O nível de compreensão e o ritmo a que podem ser realizados os testes a hipóteses científicas com recurso a computadores, tem acelerado o processo de descoberta científica nalguns domínios de aplicação, tornando possível descrever as propriedades fundamentais de alguns desses processos biológicos complexos através de modelos abstractos detalhados. Tal tem sido concretizado de forma a que se tenha tornado possível aplicar esses modelos funcionais teóricos como entidades abstractas de processamento de informação digital, que possam ser usados para lidar com problemas difíceis e não explorados até ao momento.

Estes problemas não estão sempre relacionados com o domínio biológico original, mas as suas soluções dependem muitas vezes de estratégias complexas de aprendizagem e cognição que são normalmente difíceis de identificar e implementar através das abordagens algorítmicas mais tradicionais. Por exemplo, a inspiração obtida no estudo computacional do cérebro e na diferenciação das espécies fornecida pela teoria da evolução, estão entre os sistemas biológicos mais relevantes implementados em computadores, que levou respectivamente ao enorme sucesso dos modelos computacionais baseados em Redes Neuronais Artificiais e Algoritmos Genéticos.

O desenvolvimento de modelos abstractos para a explicação do comportamento do sistema imune dos vertebrados constitui um outro exemplo de uma área vibrante de investigação e uma fonte de inspiração apelativa para a implantação e desenvolvimento de sistemas inovadores e adaptativos de detecção de anomalias, sob a forma de Sistemas Imunes Artificiais (SIA). Estes sistemas são o resultado do desenvolvimento de novos modelos computacionais e algoritmos inspirados no sistema imunitário, que implementam e incorporam os mecanismos intrínsecos de cognição e aprendizagem do sistema imunitário, aplicados aos processos de tolerância ao *próprio* e discriminação do *não-próprio*. O principal objectivo dos SIA desenvolvidos para a detecção de anomalias consiste em distinguir correctamente os comportamentos anómalos e prejudiciais que poderão comprometer parcial ou totalmente o sistema em monitorização, dos que presumivelmente correspondem a comportamentos benignos e, portanto, "normais".

As teorias imunológicas assentes na *selecção negativa* e na *teoria do perigo* correspondem a modelos funcionais clássicos bem conhecidos, que tentam explicar alguns aspectos do comportamento do sistema imunitário e que têm inspirado positivamente o desenvolvimento

de SIA para a detecção de anomalias. Apesar das excelentes contribuições que realizaram e dos resultados promissores obtidos até ao momento em diversos domínios, a investigação em SIA enfrenta ainda vários desafios. Em primeiro lugar, existe a necessidade em basear a implementação dos SIA numa maior compreensão das capacidades de distinção entre *próprio* e *não próprio* levada a cabo pelo sistema imunitário. No futuro, os SIA poderão ser o resultado da "assemblagem" de vários modelos abstractos de processos biológicos, alguns dos quais ainda não suficientemente estudados. Em segundo lugar, e este é um dos temas recorrentes desta tese, a necessidade de avaliar e integrar nessa "assemblagem" outras teorias imunológicas ainda não estudadas, que expliquem a forma como o sistema imunitário lida com as alterações provocadas pelo ambiente ao longo do tempo. Finalmente, também realçamos as dificuldades que advém da transposição e aplicação dos modelos computacionais abstractos do sistema imunitário em domínios não biológicos.

Os resultados que inferimos e as observações mais gerais que podemos fazer da aplicação dos SIA desenvolvidos em ambientes não biológicos são igualmente valiosos, uma vez que podem ser usados como um mecanismo alternativo de validação dos modelos teóricos originais. A validação destes modelos e o seu comportamento no domínio digital pode ajudar na explicação de algumas das propriedades comportamentais do fenómeno biológico em estudo, fornecendo assim um maior suporte científico e evidência da sua validade.

Nesta tese propomos um SIA original baseado numa teoria imunológica relacionada com o ajuste dinâmico do limiar de activação das células, denominada *Tunable Activation Threshold (TAT)*, com vista a lidar com alguns dos desafios descritos. Começamos por descrever como o modelo baseado no TAT ajuda a explicar o comportamento biológico observado nas células T, avaliando de seguida aquele em dois contextos aplicacionais específicos: a detecção de anomalias e a classificação de texto. Os componentes aplicacionais desenvolvidos no SIA foram mantidos tão parecidos quanto possível com os seus homólogos imunológicos, por forma a manter o mecanismo de validação digital tão próximo quanto possível do modelo biológico original. O SIA genérico que desenvolvemos, denominado TAT-AIS, é altamente modular, tendo como componente principal um simulador da dinâmica do TAT aplicada a uma população de detectores (células).

Avaliámos o funcionamento do TAT-AIS em dois domínios de aplicação distintos. Na detecção de intrusões numa rede de computadores, testámos o funcionamento do TAT e comparámos os resultados obtidos com um sistema muito popular baseado em assinaturas de ataques, o `snort`. Ambos os sistemas foram comparados usando o mesmo conjunto de dados, correspondente a tráfego recolhido em redes operacionais, composto por pacotes relativos a comportamentos normais de aplicações e utilizadores. Estes dados foram recolhidos em redes locais e no conjunto de dados de teste introduzimos de forma aleatória fluxos

de tráfego de rede correspondentes a anomalias previamente catalogadas e validadas. Os resultados obtidos demonstraram a capacidade de um modelo computacional baseado no TAT para lidar com as mudanças temporais na dinâmica da natureza do tráfego da rede. Ao contrário do `snort`, o TAT foi capaz de detectar alguns ataques desconhecidos e, ao mesmo tempo, ajustar gradualmente a sua sensibilidade para tolerar novas formas de actividade normal da rede.

Relativamente à classificação de texto, desenvolvemos uma aplicação híbrida baseada na aprendizagem por comités, composta por dois tipos diferentes de classificadores: máquinas de vector de suporte (SVM) e o TAT-AIS. Os resultados experimentais foram obtidos com o conjunto de dados Reuters-21578, tendo mostrado que o classificador híbrido conseguiu obter melhor desempenho que cada um dos membros do comité individualmente, no processamento do mesmo conjunto de dados.

Os problemas estudados nesta tese estão relacionados com um dos maiores desafios actuais nos sistemas de aprendizagem computacional, designadamente a detecção de episódios de alteração comportamental em ambientes contínuos e dinâmicos. Acreditamos que o sistema TAT-AIS apresentado é uma aplicação versátil e constitui uma contribuição útil para aumentar o arsenal de algoritmos e técnicas actualmente disponíveis para lidar com este problema difícil. No final da tese apresentamos algumas novas linhas de investigação que resultaram do trabalho em que assenta esta dissertação.