

用于高密度图像模式分类的主动学习增强的非线性细胞自动机

P. Kiran Sree¹, Sssn Usha Devi N.²

1 印度 比玛瓦拉姆 Shri Vishnu 女子工程学院计算机科学与工程系

2 印度 卡基纳达 贾瓦哈拉尔·尼赫鲁技术大学大学工程学院计算机科学与工程系

摘要:介绍了一种基于非线性元胞自动机财产的高密度图像分类新方法。我们使用一个状态转换，它由一组不相交的树组成，这些树扎根于单位循环长度的循环状态，从而形成一个自然分类器。所提出的框架通过遗传算法得到加强，以找到作为全局状态函数的所需的建模局部规则。

关键词: 细胞自动机; 主动学习 (DL); 非线性 CA

Non Linear Cellular Automata Enhanced with Active Learning for Pattern Classification in Highly Dense Images

P. Kiran Sree¹, Sssn Usha Devi N.²

1Dept of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India

2Dept of Computer Science and Engineering, University College of Engineering, Jawaharlal Nehru Technological University, Kakinada, India

Abstract: This paper introduces a new approach to classify several high density images based on the properties of Non Linear Cellular Automata. We use a state-transition which consists of a set of disjoint trees rooted at cyclic states of unit cycle length thus forming a natural classifier. The framework proposed is strengthened with genetic algorithm to find the desired local rule of the modeling as a global state function.

Keywords: Cellular automata (CA); Active learning (DL); Non linear CA

1. 引言

在论文的第一部分中,我们开发了一种基于线性 DLM 和非线性主动学习机制的分类器,它可以解决生物信息学中的主要问题,如蛋白质编码区识别、蛋白质结构预测和启动子区识别。我们还提出了人工免疫系统,这是一种新的计算智能技术,用于增强系统的适应性,并将更多的并行性纳入系统。我们还展示了如何使用细胞自动机提高聚类质量。

在论文的第二部分中,我们探索了一种基于启发式的非统一主动学习机制的入侵检测系统,该系统可以监控网络中的恶意活动或策略违规,并向管理站生成报告。我们发现了一种抽象 IDS 模式,它定义了基于行为的 IDS 和基于签名的 IDS 的一般特征和模式,将用于发现网络中的潜在威胁。

蛋白质是一种令人难以置信的高原子量天然强化蛋白,含有由肽键连接的氨基酸。蛋白质是每个活细胞和感染的结构和能力的基础。细胞中的蛋白质决定了该细胞的样子以及该细胞会做什么样的工作。这些品质同样决定了身体中各种

细胞的协调方式。如果我们识别了蛋白质编码区,我们就可以提取大量数据,比如 DNA 如何控制你的手指数量、你的腿放在身体上的位置以及你的眼睛的阴影。DNA 分为内含子和外显子。内含子形成 DNA 链的重要部分,外显子构成 DNA 链的次要部分。尽管如此,外显子只是由蛋白质编码位点组成。识别外显子中的蛋白质编码区是一项真正的测试。所提出的计算 LMADLM、NPCRITDLMDLM 可以处理不同长度的 DNA 序列。试验结果证实了所提出的基于 FDLM 的分类器的多功能性,无论类别、元组和特征的数量如何,都可以处理大量的数据集。已经建立了非常精确的分组。

生物信息学问题

蛋白质是一种复杂的高分子量有机化合物,含有由肽键连接的氨基酸。蛋白质对所有活细胞和病毒的结构和功能至关重要。细胞中的蛋白质决定了细胞的外观和工作。基因也决定了身体中不同细胞的排列方式。如果我们识别蛋白质编码区,我们就能提取很多信息,比如 DNA 如何控制你的手指

数量, 你的腿放在身体上的位置, 以及你眼睛的颜色。DNA 以内含子和外显子的形式组织。内含子构成 DNA 的主要部分链和外显子构成 DNA 链的次要部分。但是, 外显子仅由蛋白质编码区组成。识别外显子中的蛋白质编码区是一个真正的挑战。所提出的算法 LMADLM、NPCRITDLM 可以处理不同长度的 DNA 序列。实验结果证实了所提出的基于 FDLM 的分类器的可扩展性, 无论类、元组和属性的数量如何, 都可以处理大量数据集。已经建立了良好的分类精度。Fickett 和 Tung 数据集用于测量分类器的效率。

在遗传学中, 启动子是启动特定基因转录的 DNA 区域。启动子位于它们转录的基因附近, 位于 DNA 的同一条链和上游。提出了一种利用 DLM 识别启动子区域的算法。真核启动子数据库使用了新的数据集。

蛋白质结构预测是从蛋白质的氨基酸序列预测蛋白质的三维结构, 即从蛋白质的一级结构预测蛋白质的二级、三级和四级结构。结构预测与蛋白质设计的逆问题有本质区别。蛋白质结构预测是生物信息学追求的最重要目标之一。使用的数据集取自 DLMSPP。

2. 增强了主动学习的 AIS

描述了一种人工免疫系统 (ARTIS), 它融合了自然免疫系统的许多财产, 包括多样性、分布式计算、容错、动态学习和自适应以及自我监控。ARTIS 是分布式自适应系统的通用框架, 原则上可以应用于许多领域。该 AIS-MDLM 系统用于加强蛋白质编码区识别系统和蛋白质结构预测系统。

人工深度学习机制 (DLM) 的基本单元是具有在离散时间和空间中前进的基本结构的细胞。在 DLM 简单同质结构改善的历史背景下, 最重要的转折点之一是因为 Wolfram。复杂问题的答案要求并行注册环境。大多数并行 PC 包含十几个处理器。DLM 可以在比巨大的并行 PC 更大的规模上实现并行。DLM 通过其单元的可用性来描述。所有通信都发生在附近的场所。手机只能与相邻的手机通话。促进, 互连连接在很大程度上只传递少量数据。这一规则的一个后果是, 没有一个单元对整个框架具有全球视角。

网络入侵检测系统的一个基本问题是熟练的攻击者通过利用监控器所看到的业务流中的模糊性来逃避检测的能力。我们讨论了通过引入称为流量 MADLM 归一化器的新网络转发元素来解决这个问题的可行性。MADLM 标准化器直接位于进入站点的流量路径中, 并在监控器看到流量之前修

补数据包流以消除潜在的歧义, 从而消除规避机会。我们研究了设计 MADLM 规范化器时的一些权衡, 强调了规范化在多大程度上破坏端到端协议语义这一重要问题。

我们讨论了“冷启动”和对 MADLM 归一化器的攻击这两个关键的实际问题, 并开发了一种系统地检查歧义的方法存在于基于漫游协议头部的协议中。然后, 我们提出了规范, 即 MADLM 规范化器的公开可用用户级实现, 该规范化器可以在内存到内存副本中以 100000 pkts/sec 的速度规范化 TCP 流量流, 这表明使用 PC 硬件的内核实现可以与具有足够净空的双向 100Mbps 链路保持同步, 以抵御小数据包的高速洪泛攻击。DARPA 入侵检测数据集用于评估开发的分类器。

我们对 DLM 的关键特性进行了广泛的调查, 这些特性将对模式识别有用。我们已经报告了 DLM 的所有特征及其类别和类别在各个领域的适用性。在这项研究之后, 我们成功地开发了线性和非线性分类器, 以解决生物信息学中的各种问题。然后用人工免疫系统对所提出的算法进行了增强, 具有更好的稳定性和准确性。对所提出的算法进行了轻微修改, 以识别网络中的入侵。

3. DLM 的复杂性

DLM 在空间扩展网络上以分布式方式执行计算。它不同于传统的并行计算方法, 在并行计算中, 问题被分成独立的子问题, 每个子问题由不同的处理器解决; 子问题的解决方案随后被组合以产生最终解决方案。

进化过程由流行的遗传算法 (GA) 指导, 其基本理念是适者生存基因。可以采用该 GA 框架来获得适合于对物理系统建模的期望 CA 规则结构。GA 公式的目标是增强对 DLM 执行计算的方式的理解, 了解 DLM 如何进化以执行特定的计算任务, 并了解进化如何在本地互连的简单单元系统中创建复杂的全局行为。

模式识别的任务在广泛的人类活动中遇到。从更广泛的角度来看, 该术语可以涵盖根据当前可用信息做出某些决定或预测的任何情况。该问题涉及应用于一组输入的程序的结构; 该过程基于观察到的属性或特征将每个新输入分配给一组类中的一个。在输入数据集上构造这样的过程被定义为模式识别

4. 模式识别中的 DLM

模式识别算法有两个阶段, 学习或训练阶段和测试阶段。

在训练阶段,用一些模式对算法进行训练。根据训练的性质,模式分类有两大类

该模型是描述预定义的一组数据类而构建的。数据库中的样本集(每个成员都属于预定义的类之一)用于训练模型。训练阶段称为分类器的监督学习。每个成员可能具有多个功能。分类器基于特定度量进行训练。在训练之后,模型在测试阶段执行预测任务。输入样本类别的预测是基于某种度量(通常是距离度量)进行的。

5.结论

本文可以通过建立基于非线性 DLM 的联想记忆模型的记忆容量来扩展。可以探索基于 FDLM(模糊细胞自动机)的复杂函数模型,该模型涉及具有实数属性的数据集。所提出的算法稍加改动,也可以用作压缩算法。本文可以扩展到提出一种结合非线性 DLM(NLDLM)和模糊集的混合系统。

参考文献

[1] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Investigating an Artificial Immune System to Strengthen the Protein Structure Prediction and Protein Coding Region Identification using Cellular Automata Classifier. International Journal of Bioinformatics Research and Applications, Vol 5, Number 6, pp 647-662, ISSN: 1744-5493. (2009) (Inderscience Journals, UK) Listed & Recognized in US National Library of Medicine National Institutes of Health. National Center for Biotechnology Information (Government of USA) PMID: 19887338 [PubMed-indexed for MEDLINE] H Index (Citation Index): 08 (SCImago, www.scimagojr.com) (Nine Years Old Journal).

[2] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. The International Arab Journal of Information Technology (IAJIT), Volume 7, No 1, 2010, pp 75-78. ISSN: 1683-3198 H Index (Citation Index): 05 (SCImago, www.scimagojr.com)(Eleven Years Old Journal)(SCI Indexed Journal).

[3] Dr P. KiranSree&DrInampudi Ramesh Babu et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier for Coding Region Prediction in Human Genome, Journal of Bioinformatics and Intelligent Control, Vol. 3, 2014, pp 1-6. DOI: 10.1166/jbic. 2014. 1077 (American Scientific

Publications, USA).

[4] Dr P. KiranSree&DrInampudi Ramesh Babu et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier Promoter Region Prediction in Eukaryotes. Journal of Bioinformatics and Intelligent Control, Vol. 3, 1-6, 2014. DOI: 10.1166/jbic. 2014. 1077 (American Scientific Publications, USA).

[5] Dr P. KiranSree&DrInampudi Ramesh Babu et al, 5. MACAMCC-DA: A Fast MACA with Modified Clonal Classifier Promoter Region Prediction in Drosophila and Arabidopsis. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 22-26, Impact Factor: 1.74.

[6] Dr P. KiranSree & DrInampudi Ramesh Babu et al, Cellular Automata in Splice Site Prediction. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 36-39, Impact Factor: 1.74.

[7] Dr P. KiranSree&DrInampudi Ramesh Babu et al, AIXMACA-Y Multiple Attractor Cellular Automata Based Clonal Classifier for Promoter and Protein Coding Region Prediction. Journal of Bioinformatics and Intelligent Control 3, no. 1 (2014): 23-30. DOI: 10.1166/jbic. 2014. 1071, (American Scientific Publications, USA).

[8] Dr P. KiranSree&DrInampudi Ramesh Babu et al, PSMACA: An Automated Protein Structure Prediction Using MACA (Multiple Attractor Cellular Automata). Journal of Bioinformatics and Intelligent Control 2, no. 3 (2013): 211-215. DOI:10.1166/jbic. 2013. 1052 (American Scientific Publications, USA).

[9] Dr P. KiranSree&DrInampudi Ramesh Babu et al, An extensive report on Cellular Automata based Artificial Immune System for strengthening Automated Protein Prediction. Advances in Biomedical Engineering Research (ABER) Volume 1 Issue 3, September 2013, pp 45-51. Science Publications (USA).

[10] Dr P. KiranSree&DrInampudi Ramesh Babu et al, A Novel Protein Coding Region Identifying Tool using Cellular Automata Classifier with Trust-Region Method and Parallel Scan Algorithm (NPCRITCACA). International Journal of Biotechnology & Biochemistry (IJBB) Volume 4, 177-189 Number 2 (December 2008). (Eight Years Old Journal) Listed in Indian Science Abstracts, ISSN: 0019-6339, Volume 45, Number 22, November 2009.

[11] Dr P. KiranSree&DrInampudi Ramesh Babu et al, HMAACA: Towards proposing Cellular Automata based tool for protein coding, promoter region identification and protein structure prediction. International Journal of Research in Computer Applications & Information Technology, Volume 1 Number 1, pp 26-31, 2013.

[12] Dr P. KiranSree&DrInampudi Ramesh Babu et al, PRMACA: A Promoter Region identification using Multiple Attractor Cellular Automata (MACA) in the proceedings CT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I Advances in Intelligent Systems and Computing Volume 248, 2014, pp 393-399 (Springer-AISC series).

[13] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Towards Proposing an Artificial Immune System for strengthening PSMACA: An Automated Protein Structure Prediction using Multiple Attractor Cellular Automata

proceedings of International Conference on Advances in electrical, electronics, mechanical and Computer Science (ICAEEMCS)-2013, ISBN: 978-93-81693-66-04 on September 2nd 2013, Hyderabad.

[14] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Multiple Attractor Cellular Automata (MACA) for Addressing Major Problems in Bioinformatics in Review of Bioinformatics and Biometrics (RBB) Volume 2 Issue 3, September 2013, pp70-76.

[15] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Protein coding region Identification, in proceedings of 2nd International Conference on Proteomics Bioinformatics, July 2-4, 2012 Embassy Suites Las Vegas, USA “,(Special Issue of Journal of Proteomics & Bioinformatics. (USA), Volume 5 Issue 6 – 123, ISSN: 0974-276X, H Index (Citation Index): 06 (SCImago, www.scimagojr.com)Impact Factor: 2.2, (Five Years Old Journal).