Verificação Formal Aplicada a Redes Neurais Profundas

Fabio Baldissera fabio.baldissera@ufsc.br
Max Hering de Queiroz max.queiroz@ufsc.br
Rodrigo Tacla Saad rtsaad@das.ufsc.br

Motivação

As Redes Neurais foram por muito tempo objeto de pesquisa nas universidade e centros de pesquisa sem muita importância em aplicações comerciais devido à necessidade de grandes recursos computacionais para realizar o seu aprendizado. Entretanto, o advento das redes neurais profundas e a redução significativa no custo associado ao processamento computacional possibilitou, em alguns segmentos, soluções com resultados superiores aos algoritmos baseados em regras. Um grande exemplo desta evolução são as redes neurais convolucionais que são hoje o estado da arte para tarefas de reconhecimento de imagem [1, 2].

Atualmente, as redes neurais começaram a ser empregadas em sistemas classificados como críticos, ou seja, sistemas em que uma falha pode resultar em lesões pessoais graves, mortes, danos ao meio ambiente ou perda financeira significativa. Um exemplo de sistema crítico que emprega redes neurais são os carros autônomos [6]. Este novo cenário de utilização impõe um problema de segurança visto que o aprendizado de máquina não fornece nenhuma explicação satisfatória das regras aprendidas na rede neural. Somando-se a este desconhecimento de como se comportam, trabalhos recentes demonstram que redes neurais podem ser facilmente enganadas [5].

Neste contexto, a verificação formal se torna importante para aumentar a confiança dando previsibilidade ao comportamento da rede neural. O grupo de pesquisa em verificação formal busca um mestrando com conhecimentos de diferentes áreas como Estatística, Inteligência Artificial, Verificação Formal e Computação (estruturas de dados e algoritmos) para integrar a sua equipe. O objetivo deste mestrado é estudar técnicas de extração de conhecimento e verificação para analisar a segurança de redes neurais aplicadas a sistemas críticos.

Objetivos

Os objetivos esperados para o desenvolvimento deste tema de mestrado são apresentados abaixo:

1. Estudo bibliográfico detalhado das técnicas de verificação formal aplicadas para redes neurais disponíveis na literatura;
2. Realizar um estudo comparativo das principais técnicas de verificação formal para redes neurais;
3. Exemplo de caso: aplicar as técnicas de verificação formal nas plataformas de robótica móvel disponíveis no Departamento de Automação e Sistemas da UFSC;

Referências

[1] K. He, X. Zhang, S. Ren, e J. Sun, “Deep Residual Learning for Image Recognition”, arXiv:1512.03385 [cs], dez. 2015.

[2] C. Szegedy et al., “Going deeper with convolutions”, in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, p. 1–9.

[3] Q. Wang, K. Zhang, X. Liu, e C. L. Giles, “Verification of Recurrent Neural Networks Through Rule Extraction”, arXiv:1811.06029 [cs, stat], nov. 2018.

[4] C. Liu, T. Arnon, C. Lazarus, C. Barrett, e M. J. Kochenderfer, “Algorithms for Verifying Deep Neural Networks”, arXiv:1903.06758 [cs, stat], mar. 2019.

[5] T. B. Brown, D. Mané, A. Roy, M. Abadi, e J. Gilmer, “Adversarial Patch”, arXiv:1712.09665 [cs], dez. 2017.

[6] S. Thrun et al., “Stanley: The robot that won the DARPA Grand Challenge”, Journal of Field Robotics, vol. 23, nº 9, p. 661–692, set. 2006.

[7] T. Dreossi et al., “VERIFAI: A Toolkit for the Formal Design and Analysis of Artificial Intelligence-Based Systems”, p. 10.

[8] S. A. Seshia et al., “Formal Specification for Deep Neural Networks”, in Automated Technology for Verification and Analysis, vol. 11138, S. K. Lahiri e C. Wang, Orgs. Cham: Springer International Publishing, 2018, p. 20–34.