

Értelmezhető mélytanulási modellek a számítógépes látásban

Szenkovits Annamária, Portik Ábel

Babeş-Bolyai Tudományegyetem, Kolozsvár

annamaria.szenkovits@ubbcluj.ro, abel.portik@stud.ubbcluj.ro

Az összetett gépi tanulási modellek esetében gyakran nehezen látunk bele ezek működésébe, döntéshozási folyamatába. Ez különösen igaz a mélytanulási modellekre, melyek – bár egyre nagyobb pontossággal oldanak meg sok feladatot – komplexitásuk miatt sokszor fekete dobozként használjuk őket.

Az említett kihívásokat figyelembe véve, az elmúlt néhány évben a mesterséges intelligenciával, gépi tanulással foglalkozó kutatói közösség egyre nagyobb hangsúlyt fektet az értelmezhető mélytanulási modellek tanulmányozására. Olyan modellek kidolgozására törekednek, melyek az emberi felhasználó számára nagyobb transzparenciát, döntéshozási folyamatukba nagyobb beelátást biztosítanak.

Bemutatónkban a számítógépes látás területén alkalmazható értelmezhető modellekre összpontosítunk. Nevesetesen a PIP-Net, illetve a ProtoSeg modellekre térünk ki. Az előbbit képosztályozási, míg az utóbbit szemantikus szegmentációs feladatokra alkalmazhatjuk. Mindkét modell úgynevezett prototípusokat hoz létre betanításuk során, melyek egy adott osztályhoz tartozó képekre jellemző mintázatokat tartalmaznak. A modellek a predikcióikat ezen prototípusok előfordulása alapján számítják ki. Előadásunkban felvázoljuk a modellek architektúráját és a prototípusok kialakításának folyamatát, emellett kitérünk arra, hogy milyen kihívásokba ütköztünk a modellek kapcsán a kísérleteink során, illetve hogy milyen megoldásokkal próbálkoztunk ezek áthidalására.

A modellek működését esettanulmányokon keresztül szemléltetjük, a szakirodalomban sokak által használt CUB-200, valamint Cityscapes adathalmazokon.

Hivatkozások

- [1] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), pages 80–89, 2018.
- [2] M. Nauta, J. Schlotterer, M. van Keulen, and C. Seifert. PIP-Net: Patch-Based Intuitive Prototypes for Interpretable Image Classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2744–2753, 2023.
- [3] M. Sacha, D. Rymarczyk, L. Struski, J. Tabor, and B. Zielinski. ProtoSeg: Interpretable Semantic Segmentation With Prototypical Parts. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1481–1492, 2023.
- [4] Welinder P., Branson S., Mita T., Wah C., Schroff F., Belongie S., Perona, P. Caltech-UCSD Birds 200. California Institute of Technology. CNS-TR-2010-001. 2010.
- [5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding, 2016.