MTA 200 ÉVES A MAGYAR TUDOMÁNYOS AKADÉMIA

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI TUDOMÁNYOK OSZTÁLYÁNAK
ÜNNEPI RENDEZVÉNYE

MESTERSÉGES INTELLIGENCIA MATEMATIKAI SZEMMEL

A rendezvény időpontja: 2026. január 5., 13.00 óra

A rendezvény helyszíne: MTA Székház, Díszterem

1051 Budapest, Széchenyi István tér 9.

13.00—16.00 Mesterséges intelligencia matematikai szemmel

Levezető elnök: Páles Zsolt

Miranda Christ: ***The Structure of Relation Decoding Linear Operators in Large Language Models***

We investigate the structure of linear operators that decode specific relational facts in transformer language models. We extend their single-relation findings to a collection of relations and systematically chart their organization. We show that such collections of relation decoders can be highly compressed by simple order-3 tensor networks without significant loss in decoding accuracy. To explain this surprising redundancy, we develop a cross-evaluation protocol, in which we apply each linear decoder operator to the subjects of every other relation. Our results reveal that these linear maps do not encode distinct relations, but extract recurring, coarse-grained semantic properties (e.g., country of capital city and country of food are both in the country-of-X property). This property-centric structure clarifies both the operators' compressibility and highlights why they generalize only to new relations that are semantically close. Our findings thus interpret linear relational decoding in transformer language models as primarily property-based, rather than relation-specific. Joint work with Adrián Csiszárik, Gergely Becsó, Dániel Varga

Jelasity Márk: *Formal Verification of Deployed Neural Networks*

Formal verification seeks to provide mathematical proof that a given neural network behaves nicely in a given environment of an input, for example, that it predicts the same label in the close proximity of the input. However, the known methods of formal verification overlook the fact that both the tools used by the verification algorithm and the neural network itself are implemented in floating point, which introduces a range of practically exploitable vulnerabilities. These vulnerabilities allow an attacker to construct a neural network with undetectable adversarial behavior that can be triggered by, for example, properties of the environment such as floating point precision, or specific orderings of associative arithmetic operations.

Here, we discuss such floating point related vulneraibilities and propose an approach for their mitigation that allows us to properly bound the floating point error even in deployments, where any expression tree is possible. This can be used to fix a number of sound verification approaches including symbolic interval propagation.

Csáji Balázs: *Resampled Median-of-Means for Heavy-Tailed Bandits*

Stochastic multi-armed bandits (MABs) provide a fundamental framework to study sequential decision making in uncertain environments. The upper confidence bounds (UCB) algorithm is a primary choice to solve these problems as it achieves near-optimal regret rates under various moment assumptions. Up until recently most UCB methods relied on concentration inequalities leading to confidence bounds which depend on moment parameters, such as the variance proxy of subgaussian distributions, that are usually unknown in practice. In this talk we present a new distribution-free, data-driven UCB algorithm for symmetric reward distributions which is completely parameter-free, e.g., it needs no moment information. The key idea is to combine a refined, one-sided version of the recently developed resampled median-of-means (RMM) estimator with UCB. The resulting anytime, parameter-free RMM-UCB algorithm achieves near-optimal regret, even for heavy-tailed reward distributions. Experiments also show that RMM-UCB outperforms most state-of-the-art bandit algorithms on difficult MAB problems, i.e., when the suboptimality gap is small and the reward distributions are heavy-tailed. Joint work with Ambrus Tamás and Szabolcs Szentpéteri.

Ifj. Benczúr András: *Trustworthy AI in mobile radio networks: explainability, causality, uncertainty quantification*

Az egyre komplexebb, 5G és azon túli rádiós hálózatokban a gépi tanulás szerepe egyre fontosabbá válik, a modellek döntéseinek megértése és megbízhatósága kulcsfontosságúvá lesz. Először azt vizsgálom, hogyan képes a magyarázható modellezés (XAI) — például az additív lokális jellemző-hozzárendelési módszerek, mint a SHAP — feltárni az oksági kapcsolatokat a hálózati konfiguráció és a teljesítménymutatók (KPI-k) között. Új, az oksági függőségekkel jobban összhangban lévő attribúciós technikákat vezetünk be, amelyek javítják az értelmezhetőséget.

A rádióhálózati előrejelzés motivációjából kiindulva ezután a regressziós problémákban jelentkező bizonytalanság kvantifikálásának kihívását tárgyaljuk. Egy nem-determinisztikus neurális hálózati regressziós keretrendszert javaslunk, amelyet a Continuous Ranked Probability Score (CRPS) mintaalapú közelítésével optimalizálunk. Ez lehetővé teszi az aleatorikus bizonytalanság eloszlásfüggetlen tanulását, jól kalibrált valószínűségi előrejelzéseket biztosítva.

Végül poszthoc, nemparametrikus újrakalibrálási módszereket tárgyalunk, amelyeket a modellkalibrációt vizsgáló új statisztikai tesztek inspiráltak, hogy megbízható döntéshozatalt tegyenek lehetővé összetett, nagy kockázatú hálózati környezetekben.