

Analiza interdependențelor dintr-o arhitectură de microservicii folosind rețele neuronale Long Short-Term Memory

Emanuel Achirei

Rezumat

În ziua de astăzi, majoritatea aplicațiilor software utilizează o infrastructură cloud, într-un fel sau altul. Una din cele mai utilizate tehnici de dezvoltare ale acestora este arhitectura bazată pe microservicii – entități ce satisfac un scop cât mai simplu. Însă, extinderea lor a dus la amplificarea nevoii de monitorizare și claritate a interdependențelor aferente.

Lucrarea de față propune o nouă abordare cu privire la identificarea dependențelor dintr-o arhitectură de servicii, utilizând rețele neuronale de tip Long Short-Term Memory. Se încearcă obținerea unui graf al co-utilizării diferitelor resurse, folosind ca date de intrare rezultatele unui sistem de monitorizare. Aceste date sunt codificate binar – 0 pentru status funcțional și 1 pentru nefuncțional – obținându-se cuvinte de forma „1001100”.

Aplicația conține două rețele neuronale:

- o rețea LSTM (Long Short-Term Memory) ce este antrenată în scopul de a evidenția precedența stărilor arhitecturii;
- o rețea Fully Connected ce primește ca input rezultatele testării rețelei anterioare pe toate stările posibile și are ca țel asimilarea conexiunilor dintre entități.

Proiectarea rețelelor neuronale a fost făcută cu ajutorul framework-ului TensorFlow în limbajul de programare Python.

În prima fază s-au utilizat date generate aleatoriu de către un algoritm propriu, mai apoi utilizându-se și date dintr-un sistem real (infrastructura cloud BitDefender).

S-au obținut rezultate satisfăcătoare pentru datele generate, dar datele reale au prezentat câteva dificultăți:

- datele au trebuit rafinate, deoarece setul era „poluat” de anumite stări – fie starea în care toate entitățile erau funcționale (stare preponderentă), fie stări eronate cu fals-positiv din cauza funcționării precare a sistemului de monitorizare;
- în urma rafinării, setul de date a fost micșorat, iar rețeaua LSTM nu a mai fost capabilă să asimileze eficient modelele repetitive;
- în esență rețeaua LSTM se bazează pe o cantitate mare de date, variate și totodată acestea trebuie să prezinte stări de nefuncționare ale sistemului – lucru nedorit într-un mediu real.

Una din direcțiile de dezvoltare ar fi proiectarea a unei rețele ce poate observa modele în date mai complexe utilizând un set de date de antrenare minimal.