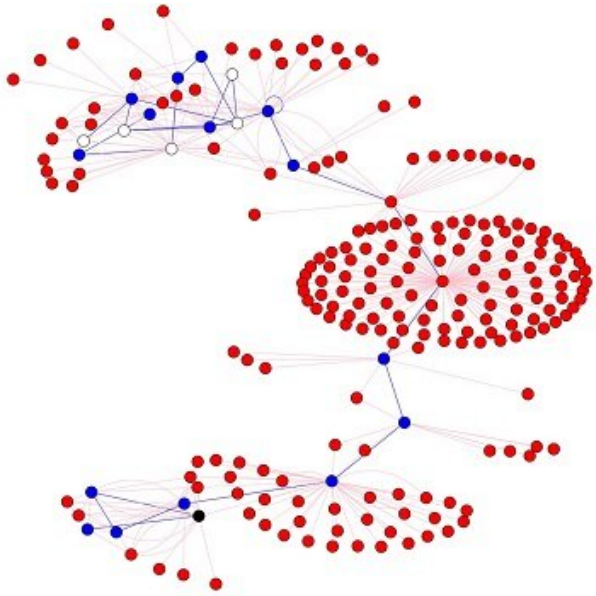
Training agents to walk with purpose: Improving machine learning and relational data classification

June 1 2020



Using reinforcement learning, a search agent can be trained to walk on a relational data set more efficiently than a conventional "random" walk can do. For example, if a search agent needs to learn about a particular individual, it will be positively rewarded for interacting with entities that know the individual (blue dots), negatively rewarded for interacting with entities that do not know the individual (red dots) and unrewarded for interacting with entities that may or may not know the individual (white dots). Note that the agent may smartly interact with a few red dots to reach more blue dots. Credit: Akujuobi et al.

A classification algorithm for relational data that is more accurate, as well as orders of magnitude more efficient than previous schemes, has been developed through a research collaboration between KAUST and Nortonlifelock Research Group in France.

The new algorithm, which uses an approach called reinforcement learning, demonstrates the power of machining learning techniques in even tried-and-true tasks like classifying relational data.

One of the most common classes of data is relational data, where discrete data points or nodes are connected in some way to others. A social network is a good example, where each user is connected by friend relationships to others, and also by shared interests, geography, or other traits or labels.

Classifying relational data involves a search agent taking an exploratory "walk" following the connections among nodes. A simple agent does this randomly, but such an approach is wildly inefficient and computationally intensive; it can also result in suboptimal classification accuracy if the agent finds itself in a relational cul-de-sac.

Uchenna Akujuobi, in collaboration with KAUST colleagues, and Han Yufei from Nortonlifelock, has now successfully developed a more robust approach by introducing machine learning techniques.

"Most real-world relational data can be put in a graph-structured format comprising data nodes connected by edges denoting the relationships," explains Akujuobi. "We set out to build a graph-based classification model that trains the agent using a reinforcement method in order to achieve a better classification result."

"The intuition behind our method is that instead of randomly selecting the path to explore, we see if we can make the agent smarter," says

Yufei. "To do this, we label some of the nodes in the dataset so that we can train the graph exploration policy."

In reinforcement learning, the agent walks from node to node and is rewarded or penalized as the agent encounters labeled data, resulting in progressive refinement of the decision policy. This training effectively reduces the "randomness" of the walk, making the classification more efficient and also less prone to inaccuracy.

"The trained agent essentially decides which node to move to at each walk step based on the relevance of neighboring nodes to the current node," says Akujuobi.

"Our method reduces the computational complexity of graph exploration by orders of magnitude, while presenting better node classification accuracy than state-of-the-art graph-structure encoding algorithms," says Yufei. "It is also generally applicable to any kind of graph-structured data, such as social-network recommendation systems and classification of biomolecules, as well as cybersecurity."

More information: Uchenna Akujuobi et al. Recurrent Attention Walk for Semi-supervised Classification, *Proceedings of the 13th International Conference on Web Search and Data Mining* (2020). DOI: 10.1145/3336191.3371853

Provided by King Abdullah University of Science and Technology

Citation: Training agents to walk with purpose: Improving machine learning and relational data classification (2020, June 1) retrieved 3 May 2024 from https://techxplore.com/news/2020-06-agents-purpose-machine-classification.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.