

Głębokie modele sieci neuronowych dla nienadzorowanego uczenia reprezentacji danych grafowych

Streszczenie popularnonaukowe

Motywacja

Korzystając z różnego rodzaju portali społecznościowych, takich jak Facebook albo Twitter, stajemy się nieświadomie częścią dużej sieci powiązań. Formalnie taka sieć jest opisywana za pomocą grafu, który składają się z wierzchołków połączonych krawędziami. W przypadku sieci społecznej wierzchołkami są osoby korzystające z tych portali, natomiast krawędzie określają różne relacje między wierzchołkami, np. relacje rodzinne, zatrudnienie, przeglądane treści czy zainteresowania.

Na podstawie tak zbieranych danych sieciowych można wykonywać różne zadania, w tym rekomendować nowych znajomych lub wyświetlać spersonalizowane reklamy. Uczenie maszynowe dostarcza narzędzia pozwalające wykonywać te zadania w zautomatyzowany i stosunkowo prosty sposób. Modele uczenia maszynowego wymagają odpowiednich reprezentacji danych wejściowych. O ile część danych nie wymaga dodatkowego przetwarzania, tak w przypadku grafów należy wpierw utworzyć numeryczne opisy wierzchołków, krawędzi czy pod-grafów. Dziedzina naukowa zajmująca się tym zagadnieniem jest nazywana uczeniem reprezentacji dla grafów i dostarcza metody i algorytmy, które na podstawie grafu są w stanie wyliczyć reprezentację wektorową (np. dla każdego wierzchołka w grafie). Większość obecnie dostępnych algorytmów jest przeznaczona dla grafów z pojedynczym typem krawędzi, natomiast wiele rzeczywistych zbiorów danych posiada wiele różnych relacji (typów krawędzi).

Cel projektu

Celem projektu jest wypełnienie luki badawczej związanej z uczeniem reprezentacji grafów z wieloma typami relacji. Opracowane algorytmy mają posłużyć jako efektywne metody wyznaczania reprezentacji wektorowych w przytoczonych grafach i dzięki temu ułatwić uczenie modeli w typowych zadaniach przetwarzania danych grafowych.

Opis badań

W ramach projektu zostaną opracowane dwie metody nienadzorowanego uczenia reprezentacji dla grafów z wieloma różnymi typami relacji. W przypadku pierwszej metody będzie wymagane opracowanie funkcji kosztu, która będzie uwzględniać istnienie różnych typów relacji oraz będzie wykorzystywać dostępne atrybuty w grafie (na poziomie wierzchołków, krawędzi oraz całego grafu). Dodatkowo będzie zbadana możliwość wykorzystania istniejących metod uczenia reprezentacji grafów w proponowanym modelu. W drugim rozwiązaniu zostanie sprawdzona hipoteza badawcza zakładająca, że można automatycznie wyznaczyć liczbę relacji w grafie (na podstawie struktury oraz dostępnych atrybutów). Ostatnim elementem projektu będzie dokładna analiza zaproponowanych modeli i badania jakości ich działania w typowych zadaniach uczenia reprezentacji grafów (np. klasyfikacja wierzchołków, predykcja krawędzi czy regresja atrybutów).

Spodziewane efekty

Spodziewanym efektem projektu będą dwie publikacje przedstawiające przytoczone metody, opublikowane na najlepszych konferencjach światowych (np. ICLR, NeurIPS, ICML, ECML).