

IterDE: Một Khung Phân Tách Tri Thức Lặp Đi Lặp Lại cho Đồ Thị Tri Thức Nhúng

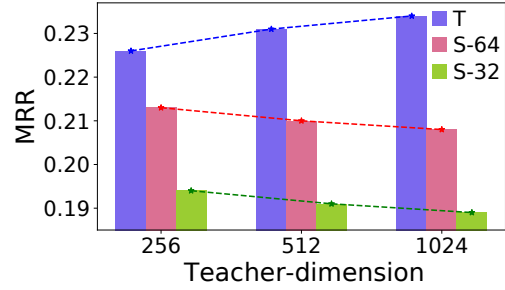
Jiajun Liu, Peng Wang*, Ziyu Shang, Chenxiao Wu
Trường Khoa học và Kỹ thuật Máy tính, Đại học Đông Nam
{jiajliu, pwang, ziyus1999, chenxiaowu}@seu.edu.cn

Tóm tắt

Chúng tôi đề xuất khung phân tách tri thức (KGE) nhằm mục đích giảm kích thước mô hình KGE để giải quyết các thách thức về giới hạn lưu trữ và hiệu quả suy luận tri thức. Tuy nhiên, các công trình hiện tại vẫn gặp phải tình trạng giảm hiệu suất khi nén một mô hình KGE gốc có chiều cao xuống một mô hình KGE chung có chiều thấp. Hơn nữa, hầu hết các công trình tập trung vào việc giảm thời gian suy luận nhưng bỏ qua quá trình đào tạo tốn thời gian của các mô hình KGE chung. Trong bài báo này, chúng tôi đề xuất IterDE, một khung chung phân tách tri thức mới cho KGE. Đầu tiên, IterDE giới thiệu một phương pháp chung phân tách lặp đi lặp lại và cho phép một mô hình KGE luân phiên là một mô hình học sinh và một mô hình giáo viên trong quá trình chung phân tách lặp đi lặp lại. Do đó, tri thức có thể được chuyển giao một cách trơn tru giữa các mô hình giáo viên có chiều cao và các mô hình học sinh có chiều thấp, đồng thời vẫn duy trì hiệu suất KGE tốt. Hơn nữa, để tối ưu hóa quá trình đào tạo, chúng tôi xem xét rằng các đối tượng tối ưu hóa khác nhau giữa mất mát nhân cứng và mất mát nhân mềm có thể ảnh hưởng đến hiệu quả đào tạo, và sau đó chúng tôi đề xuất một cơ chế điều chỉnh động trọng số nhân mềm có thể cân bằng sự không nhất quán của hướng tối ưu hóa giữa mất mát nhân cứng và nhân mềm bằng cách tăng dần trọng số của mất mát nhân mềm. Kết quả thử nghiệm của chúng tôi chứng minh rằng IterDE đạt được hiệu suất chung phân tách hiện đại mới cho KGE so với các đường cơ sở mạnh mẽ trên tác vụ dự đoán liên kết. Đáng chú ý, IterDE có thể giảm thời gian đào tạo trung bình 50%. Cuối cùng, nhiều thử nghiệm thăm dò hơn cho thấy rằng cơ chế điều chỉnh động trọng số nhân mềm và các lần lặp chi tiết hơn có thể cải thiện hiệu suất chung phân tách.

Giới thiệu

Đồ thị tri thức (Knowledge Graphs - KGs) mô tả các khái niệm và sự kiện trong các mô hình đồ thị (Dong et al. 2014), trong đó tri thức được lưu trữ dưới dạng các bộ ba. Với kích thước ngày càng tăng của KGs như Wikipedia (Bizer et al. 2009) và Yago (Suchanek, Kas-neci, and Weikum 2007), việc nhúng đồ thị tri thức (knowledge graph embedding - KGE) hiệu quả, giúp nhúng các bộ ba vào một không gian vector liên tục, đóng một vai trò then chốt trong các ứng dụng hạ nguồn như trả lời câu hỏi (Bordes, Weston, and Usunier 2014), hệ thống đề xuất (Zhang et al. 2016) và tri thức



Hình 1: Hiện tượng *giáo viên giỏi không phải lúc nào cũng dạy được học sinh giỏi* trong KGE dựa trên KD: với sự gia tăng kích thước (từ 256 lên 1024) của giáo viên T, hiệu suất của học sinh (S-64 biểu thị 64 chiều và S-32 biểu thị 32 chiều) giảm. Kết quả thu được trên tập dữ liệu WN18RR với TransE.

hoàn thiện đồ thị (Lin và cộng sự năm 2015). Hầu hết các mô hình KGE như TransE (Bordes và cộng sự năm 2013), ComplEx (Trouillon và cộng sự năm 2016), Simple (Kazemi và Poole năm 2018), RotatE (Sun và cộng sự năm 2019) đã cho thấy hiệu suất tốt hơn với số chiều nhúng cao hơn và kích thước mô hình lớn hơn, tuy nhiên, điều đó cũng dẫn đến hiệu quả suy luận chậm hơn cho các ứng dụng thực tế. Cụ thể, các mô hình KGE 512 chiều có số lượng tham số lớp nhúng nhiều hơn từ 7-15 lần và thời gian suy luận nhiều hơn từ 2-6 lần so với các mô hình KGE 32 chiều (Zhu và cộng sự năm 2022). Do đó, việc nén KGE từ các mô hình giáo viên chiều cao xuống các mô hình học sinh chiều thấp trong khi vẫn duy trì hiệu suất tuyệt vời là một vấn đề không hề nhỏ. Trong các ứng dụng thực tế, các mô hình KGE thường được yêu cầu đồng thời duy trì hiệu suất cao và tốc độ suy luận nhanh. Ví dụ, các nhà đầu tư tài chính cần nhận được các hỗ trợ quyết định thị trường chính xác và nhanh chóng từ KG tài chính thông qua các thiết bị biên. Trong kịch bản này, các mô hình KGE có thể được nén bằng cách chung phân tách kiến thức và sau đó được triển khai tới các thiết bị biên để giúp các nhà đầu tư tài chính đưa ra quyết định nhanh hơn và chính xác hơn.

Chung phân tách tri thức (KD) là một kỹ thuật phổ biến để nén mô hình, trong đó một mô hình lớn hơn đóng vai trò là mô hình giáo viên, và một mô hình nhỏ hơn đóng vai trò là mô hình học sinh cố gắng mô phỏng đầu ra của mô hình giáo viên (Hinton et al. 2014). Gần đây, mặc dù có một số phương pháp nén cho KGES

*Corresponding author.