Building a Biomedical Knowledge Graph and Predicting Novel Relations

Timo Sztyler, *Carolin Lawrence*, *Brandon Malone*
NEC Laboratories Europe, Heidelberg, Germany

Abstract

Biologists, physicians, clinical researchers, and others have spent countless hours studying the regulatory behavior among genes, the effects of medications on diseases, the correlation of therapy outcomes among patients, and knowledge about many other relationships. These successful discoveries improve our quality of life, but their discovery is time-consuming and expensive. Therefore, we propose a knowledge graph-based approach that predicts novel relationship by automatically augmenting the known relationships with predictions from a graph neural network.

For that purpose, we construct a knowledge graph (KG) which describes known associations, similarities and interactions among diseases, proteins and drugs. These relationships are extracted from publicly-available resources. More precisely, we predict *novel disease-gene associations* as well as *polypharmacy side effects*, that is, unexpected side effects due to drug-drug interactions. For this, we apply the recently-proposed KBLRN to our KG. KBLRN complements a standard neural relational approach with *path features* based on the KG in a mixture-of-experts model. Quantitatively, our approach outperforms a strong baseline (DistMult) on both prediction tasks (see Table 1).

	disease-protein		polypharmacy	
	AuROC	P@50	AuROC	P@50
DistMult	71%	69%	53%	39%
KBLRN	86%	81%	83%	77%

Table 1: Results of our proposed approach using our approach (KBLRN) versus a competitive baseline (DistMult) for disease–protein and polypharmacy (drug–drug) prediction, reporting AuROC and average precision at 50.

The path features offer explanations for our predictions. As a concrete example: It is known that the proteins LPAR1 and MMP2 interact with each other and this information is captured in our KG. Similarly, the KG is aware that the drug paliperidone affects LPAR2, whereas the drug calcium affects MMP2. These relationships are captured as path features and used during learning. Our model predicts that taking both drugs in combination leads to the previously unknown side effect “inflammation”, while explainable AI techniques reveal that the path feature is important for that prediction. Indeed, a manual literature review confirms this effect. Thus, our approach can also facilitate scientific hypothesis generation due to the explainability of the path features.

Our results show that modern neural, yet explainable, approaches can improve on standard approaches to automatically augment knowledge bases represented as KGs. While our KG is constructed from public data sources based on the primary literature, such as PubMed abstracts, we do not yet directly incorporate such resources into our model. Thus, a next step in our work is to incorporate natural language processing models into our mixture of experts.