

Constructing Knowledge Graphs and Their Biomedical Applications

This manuscript ([permalink](#)) was automatically generated from [greenelab/knowledge-graph-review@b44f2c4](#) on August 16, 2019.

Authors

- **David Nicholson**

 [0000-0003-0002-5761](#) ·  [danich1](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552 and T32 HG000046

- **Jane Roe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [janeroe](#)

Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

1. Give high level description of review as it pertains to knowledge graphs (creation and application)

Introduction

1. Talk about knowledge graphs
2. Define what they are
3. Provide a diagram to orient the reader
4. Mention how these graphs are constructed
5. Mention some applications these graphs can be used for
 1. Biomedical setting
 2. Social Media
 3. Question and answering systems
 4. more if it comes to mind
6. Conclude section by mentioning that this review dives deeper into approaches used for construction and applying these graphs to biomedical applications

Building Biomedical Knowledge Graphs

1. Set up the context for relationship extraction
 1. Define the problem
 2. Talk about the importance of the problem (filling knowledge bases -> point researchers to relevant papers)
2. Give overview towards taxonomy of approaches (hand written rules, unsupervised machine learning, supervised machine learning etc.)

Constructing Databases and Manual Curation

1. Talk about papers that construct knowledge graphs without text mining approaches
2. Discuss the positives and negatives for these methods

Text Mining for Relationship Extraction

Rule-Based Natural Language Processing

1. Mention papers on hand written rules and expressions

Unsupervised Machine Learning

1. Mention Clustering Analysis
2. Mention Co-Occurrence approaches

Supervised Machine Learning

1. Mention the availability of publically available data
 1. PPI - 5 datasets
 1. 10.1016/j.artmed.2004.07.016
 2. 10.1186/1471-2105-8-50
 3. Learning language in logic - genic interaction extraction challenge
 4. 10.1093/bioinformatics/btl616
 5. <http://helix-web.stanford.edu/psb02/ding.pdf>
 2. DaG - 3 datasets
 1. 10.1016/j.jbi.2012.04.004
 2. 10.1186/s12859-015-0472-9

3. 10.1186/1471-2105-14-323
4. 10.1186/1471-2105-13-161
3. CiD
4. 10.1093/database/baw068
5. CbG
6. Biocreative VI track 5 - raw citation
7. more if exists talk about deep learning methods
2. Mention the use of Support Vector Machines and other non deep learning classifiers
 1. Will have to mention that field has moved to deep learning.
 2. 10.1186/s13326-017-0168-3
 3. 10.1371/journal.pcbi.1004630
3. Mention deep learning methods
 1. 1901.06103v1
 2. 10.1016/j.knosys.2018.11.020
 3. 10.1177/0165551516673485
 4. 1706.01556v2
 5. ^^ A few papers here but a lot more will be put into place
 6. Mention caveat which is the need for large annotated datasets
 7. Mention a direction the field is moving to which is weak supervision and more that info that will come in time.

Applying Knowledge Graphs to Biomedical Challenges

1. Mention that these graphs can be used for discovery
2. Mention representation learning (aka representing a graph as dense vectors for nodes and/or edges)
- 3.

Unifying Techniques

1. Set up the problem that maps a knowledge graph into a low dimensional space

Matrix Factorization

1. Mention techniques for these with some papers

Deep Learning

1. Define node neighborhoods
2. Talk about random walks
3. Talk about auto encoders random walk independent approaches

Unifying Applications

1. Mention how the previous section is used in a biomedical setting

Disease and Gene Interactions

1. Mention disease gene prioritization
2. Mention Disease gene associations

Protein Protein Interactions

1. Mention predicting genes interacting genes

Drug Interactions

1. Talk about drug side effects
2. Drug repurposing
3. Drug-Disease Interactions

Clinical applications

1. Can mention EHR use and other related applications
2. Mention Tiffany's work on private data embeddings

Conclusion

1. Summarize discussed positives and pitfalls
2. Leave some open ended questions yet to be explored
3. Will come into play as I write this review paper

References
