AceKG: Large-scale Knowledge Graph for Academic Data Mining

515021910338 Ruijie Wang 2018.05.27

Content

- Background
- Dataset
- Experimental Benchmark
- Future Work

Background

- Graph:
 - G=(V,E), where V is vertex set and E is edge set;
 - Preserve network topology structure;
- Knowledge Graph:
 - G=(E,R,P), where E is entity, R is relation and P is property.
 - topology structure + vertex content + side information
 - Format: triplets

Background





Experimental benchmark

- Knowledge embedding:
 - Aim: enrich the knowledge base;
 - Method: Embed triplets into vectors;
 - Tasks: Link prediction.....

Knowledge Embedding



Figure 1: Simple illustration of TransE and TransH.

[Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.]

Knowledge Embedding

| | MRR | | | Hits at | | | |
|----------|-------|--------|--|---------|------|------|--|
| Model | Raw | Filter | | 1 | 3 | 10 | |
| TransE | 0.358 | 0.719 | | 62.7 | 82.5 | 89.2 | |
| TransH | 0.315 | 0.701 | | 61.0 | 77.2 | 84.6 | |
| DistMult | 0.432 | 0.749 | | 68.7 | 79.5 | 86.1 | |
| HolE | 0.482 | 0.864 | | 83.8 | 87.1 | 88.2 | |
| ComplEx | 0.440 | 0.817 | | 75.4 | 85.8 | 89.0 | |

Table 4: Results of link prediction task on AK18K.

Experimental benchmark

- Network representation learning:
 - NRL assigns nodes in a network to low-dimensional representations and effectively preserves the network structure.
 - Using the embedding result, we can learn the classification of the academic entities. Then we can predict the academic labels of them.

Network representation learning



[B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710. ACM, 2014.]

Network representation learning

| Metric | Method | FOS_BI | FOS_CS | FOS_EC | FOS_ME | FOS_PH | FOS_5F | Google | | | | |
|----------|---------------|--------|--------|--------|--------|--------|--------|--------|--|--|--|--|
| Micro-F1 | DeepWalk | 0.792 | 0.545 | 0.692 | 0.663 | 0.774 | 0.731 | 0.948 | | | | |
| | LINE(1st+2nd) | 0.722 | 0.633 | 0.717 | 0.701 | 0.779 | 0.755 | 0.955 | | | | |
| | PTE | 0.759 | 0.574 | 0.654 | 0.694 | 0.723 | 0.664 | 0.966 | | | | |
| | metapath2vec | 0.828 | 0.678 | 0.753 | 0.770 | 0.794 | 0.831 | 0.971 | | | | |
| Macro-F1 | DeepWalk | 0.547 | 0.454 | 0.277 | 0.496 | 0.592 | 0.589 | 0.942 | | | | |
| | LINE(1st+2nd) | 0.445 | 0.542 | 0.385 | 0.577 | 0.640 | 0.655 | 0.949 | | | | |
| | PTE | 0.495 | 0.454 | 0.276 | 0.555 | 0.571 | 0.528 | 0.961 | | | | |
| | metapath2vec | 0.637 | 0.570 | 0.485 | 0.659 | 0.635 | 0.682 | 0.968 | | | | |

Table 6: Results of scholar classification.

Network representation learning



THANK YOU