Anu Question Answering System

Balaji Ganesan[1], Avirup Saha[2], Jaydeep Sen[1], Matheen Ahmed Pasha[3], Sumit Bhatia[1], and Arvind Agarwal[1]

[1] IBM Research, {bganesa1, jaydesen, sumitbhatia, arvagarw}@in.ibm.com
[2] IIT Kharagpur, India, avirupsaha@iitkgp.ac.in
[3] IBM Data and AI, matpasha@in.ibm.com

AnuQA is a question answering system built on top of a search index and an enterprise knowledge graph. In this work, we describe five semantic technologies that have helped us address real world challenges in deploying this system. These challenges include bias in knowledge base population, entity re-resolution on streaming data, ontology alignment across data sources, explaining relationships, and providing a single unified query interface for business analytics.

Anu

[2] introduced the Anu Cognitive Compliance platform. It has enabled research in a number of fields including Search Index Optimization, Answer Sentence Selection, Document Similarity, Hypernym Discovery, Fine Grained Entity Classification, Ontology Creation and Link Prediction. We now present five semantic technologies that we have implemented to enable real world deployments of AnuQA system.

Data Augmentation for Knowledge Base Population

Data Augmentation is the process of increasing the diversity in the training data without necessarily having to acquire more data. In the context of Knowledge Bases, we have found data augmentation using IBM's rule based *SystemT* to be effective in increasing the diversity of the populated knowledge graphs and also in making downstream tasks like Link Prediction less dependent on gender, ethnicity, religion and other protected attributes.

Entity Re-resolution using Temporal Point Processes

We define Entity Re-resolution as the localized creation, updation and elimination of entities in a Knowledge Graph based on streaming updates. We have used Dirichlet Hawkes Processes (DHPs) to model both textual similarity and temporal closeness of the updates to the graph. Scaled using IBM's *Master Data Management* platform, we find DHP to be a suitable substitute to neural model predictions which are harder to explain to end users of the system.

Unified Hierarchical Label Set model for Ontology Alignment

AnuQA requires fusing information from different data sources to enable natural language querying. This is typically handled by manual processes which become cumbersome as the number of sources increases. We use the Unified Hierarchical Label Set (UHLS) model [1], based on collective learning of entity types, to integrate labels from different data sources and standard ontologies.

Explainable Link Prediction

While a number of interpretability solutions have been proposed for link prediction by graph neural networks, human understandable explanations are desirable in real world applications. Based on [3], we extract supporting text from unstructured documents, logs, lineage data and relational tables. We also look at existing paths to explain new links predicted between nodes in our knowledge graph.

Reasoning for Natural Language Interpretation

Natural Language Query [4] interfaces allow end-users to ask questions without knowing any specialized query language or data storage and schema details. We use logical reasoning over domain semantics and knowledge to support a wide variety of domain-specific queries in natural language. Domain reasoning helps us to make better interpretation of implicit intents in natural language queries, especially analytic queries typically posed to information access systems.

Deployments

Different parts of this question answering system have been deployed in various customer engagements and product offerings of IBM, especially in the financial services domain. http://covid19-india-qa.mybluemix.net is a sample instance of the AnuQA system for answering questions on COVID19.

References

1. Abhishek, A., Azad, A.P., Ganesan, B., Anand, A., Awekar, A.: Collective learning from diverse datasets for entity typing in the wild. In: Proceedings of the 2nd International Workshop on EntitY REtrieval. pp. 16–23. CEUR-WS (2019)
2. Agarwal, A., Ganesan, B., Gupta, A., Jain, N., Karanam, H.P., Kumar, A., Madaan, N., Munigala, V., Tamilselvam, S.G.: Cognitive compliance for financial regulations. IT Professional **19**(4), 28–35 (2017)
3. Bhatia, S., Dwivedi, P., Kaur, A.: That's interesting, tell me more! finding descriptive support passages for knowledge graph relationships. In: International Semantic Web Conference. pp. 250–267. Springer (2018)
4. Sen, J., Ozcan, F., Quamar, A., Stager, G., Mittal, A., Jammi, M., Lei, C., Saha, D., Sankaranarayanan, K.: Natural language querying of complex business intelligence queries. In: Proceedings of the 2019 International Conference on Management of Data. pp. 1997–2000 (2019)