A Speech-to-Knowledge-Graph Construction System

Xiaoyi Fu[1] , **Jie Zhang**[1] , **Hao Yu**[1] ,
Jiachen Li[1] , **Dong Chen**[1] , **Jie Yuan**[1] and **Xindong Wu**[1,2]
[1]Mininglamp Academy of Sciences, Mininglamp Technology, Beijing, China
[2]Research Institute of Big Knowledge, Hefei University of Technology, China
{fuxiaoyi, zhangjie.c, yuhao, lijiachen, chendong, yuanjie, wuxindong}@mininglamp.com

Abstract

This paper presents a HAO-Graph system that generates and visualizes knowledge graphs from a speech in real-time. When a user speaks to the system, HAO-Graph transforms the voice into knowledge graphs with key phrases from the original speech as nodes and edges. Different from language-to-language systems, such as Chinese-to-English and English-to-English, HAO-Graph converts a speech into graphs, and is the first of its kind. The effectiveness of our HAO-Graph system is verified by a two-hour chairman's talk in front of two thousand participants at an annual meeting in the form of a satisfaction survey.

1 Introduction

Speech interfaces enjoy a huge popularity in recent years. Take smart speakers as an example, an estimated 35 percent of U.S. households are equipped with at least one smart speaker as of 2019.[1] Despite the presence of successful speech recognition toolkits [Povey *et al.*, 2011] and commercial speech transcription systems, people still struggle to focus on the key concepts and relationships between all the concepts during a long talk. Knowledge Graph, which could be traced back to earlier studies of expert systems [Hart, 1986] and semantic networks [Sowa, 1987], provides a methodology for visualizing key ideas the speaker tries to convey.

Different definitions for the concept of Knowledge Graph exist [Ehrlinger and Wöß, 2016]. We follow the definition given in [Wu *et al.*, 2019] that 'Knowledge Graph, as a data representation tool, is to model the entities, attributes, concepts and the relationships between them.' To construct knowledge graphs from a speech, we generate two sets of key components of a knowledge graph, "entity-relation-entity" triplets and "entity-attribute" pairs as visualized in Figure 1. Dominant methodologies for knowledge graph construction from text include information extraction [Tang *et al.*, 2008] and coreference resolution [Soon *et al.*, 2001]. A text-to-knowledge-graph construction system was designed in [Stewart *et al.*, 2019] but cannot handle the Chinese language. In this work we present a prototype called HAO-Graph based on the HAO Intelligence [Wu and Wu, 2019] that integrates Human Intelligence (HI), Artificial Intelligence (AI) and Organizational Intelligence (OI) which fills in the gap and performs both generation and visualization of knowledge graphs in real-time from both texts and speeches in Chinese.

Our contributions are three folds:

- To the best of our knowledge, our system is the first knowledge graph construction system from speech.
- We design and implement an architecture that facilitates the transformation from a speech to knowledge graphs and the switching between graphs according to the speaker's topics.
- Our system also constructs knowledge graphs from open texts in Chinese.

The HAO-Graph system was released during the annual meeting of Mininglamp Technology on January 9, 2020. The system drew the knowledge graphs from a 2-hour chairman's talk in real-time. According to a satisfaction survey, 61.54% of the total respondents thought the HAO-Graph system helped with a clearer understanding of the content of the talk and 41.76% agreed that the system alleviated cognition fatigue. More than 65 percent of respondents gave a 5/5 star rating on whether the system enhanced the communication.

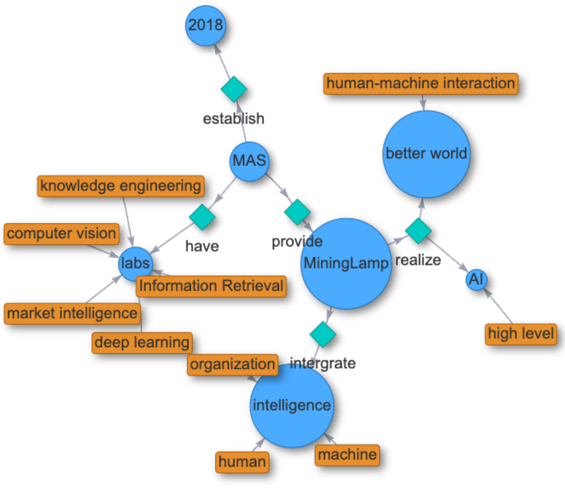

Figure 1: An example of visualization of knowledge graph.

[1] https://www.statista.com/topics/4748/smart-speakers/

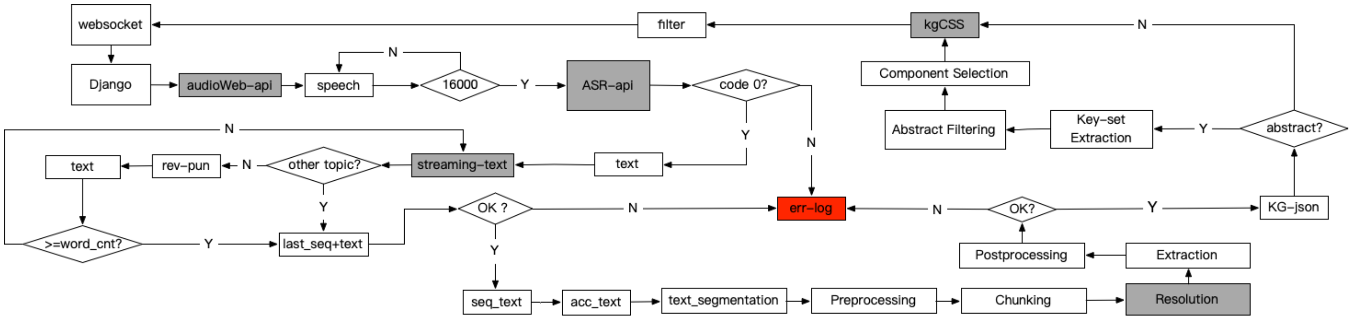

Figure 2: System architecture overview.

2 System Architecture

2.1 Speech to Text Phase

The following three modules are designed for speech to text conversion.

The Monitor. A speech is transmitted from a front-end HTML page according to the WebSocket protocol. This module monitors the binary voice stream signal data sent from the front-end page in real-time through the port and saves the data in a buffer pool. When the buffer pool data is greater than 16000 bytes, the binary voice stream data in the buffer pool is transmitted to the subsequent voice-to-text module.

The ASR Module. A module converts the received binary speech stream data into an un-punctuated text. The text data without punctuation will be sent to the front end in a multi-threaded manner as the result of speech-to-text display. The un-punctuated text in the buffer pool is verified and corrected according to context information, and the corrected results are transferred to the subsequent text punctuation module.

The Punctuation Module. The received un-punctuated text data is converted into text data with punctuation marks by BERT-based model trained on China Daily corpus[2] and saves the conversion results with the text buffer pool. This buffer pool is used to cache the text that has been punctuated because only when a complete sentence is recognized, the sentence will be sent to the subsequent Knowledge Graph construction service, so the module will send the complete sentence to the Knowledge Graph construction phase, and the last part of the text without punctuation is cached. If the punctuated text is all complete sentences, and the period is at the end of the text, then the buffer pool is cleared up.

2.2 Knowledge Graph Construction Phase

The knowledge graph construction from a text consists of the following 5 steps:

Preprocessing. Special characters in the extracted text are removed and BERT-based [Devlin *et al.*, 2019] sequence tagging models are used to perform Chinese word segmentation, part-of-speech analysis, and Head-Driven Phrase Structure Grammar Parsing [Zhou and Zhao, 2019] is reproduced for dependency tree analysis on the extracted text. Models are trained on Penn Chinese Treebank dataset[3].

Chunking. According to the part-of-speech tagging in the preprocessing step and the results of the dependency relationships, the noun parts of the speech such as the proper nouns NR and other nouns NN are grouped and combined by rules. The rules include but are not limited to two consecutive proper nouns (groups), proprietary nouns followed by other nouns, and proper nouns separated by punctuation or conjunctions. It is worth mentioning that this merge process is performed recursively. For example, the phrase ”artificial intelligence, big data, and Internet of Things technology” consists of three proper terms, a punctuation and a conjunction. In this chunking step, these words are recursively merged as ”artificial intelligence, big data and IoT technology” and yield the final chunking results.

Resolution. Based on the combined results of the chunking step, the pronouns in the text to be analyzed are replaced with the results of the coreference resolution model (replacements of the pronouns by the nouns they refer to) by calling natural language processing packages [Manning *et al.*, 2014].

Extraction. Each verb phrase is regarded as a predicate of the candidate triples using the dependency relationship parsed in the preprocessing step, and it is used as the root node to traverse its related noun phrases. Then rule-based methods are adopted to extract the triples. For the subject and object of a group, the extraction rules include, but are not limited to, the subject of the relationship (nsubj) as the triple subject, and the subject of the relationship (dobj) as the triple object.

Postprocessing. Finally, the triple obtained in the extraction step is subject to post-processing operations such as stop word removals and all triples are integrated and output.

2.3 Topic Switching

For vivid visualization, a module that detects topic changes based on the data in the graph database and the results returned by the upstream module is designed. If the current content is on the same topic as the previous content, all entity relationships related to the topic in the graph database are sent to the front end for display. If the current content and the previous content are not on the same topic, only the graph results of the current content are displayed on the front page.

2.4 Knowledge Graph Abstraction

In the process of constructing knowledge graphs from the speech, the numbers of nodes and edges increase in a very

[2] http://www.icl.pku.edu.cn/icl_res/

[3] https://catalog.ldc.upenn.edu/LDC2013T21

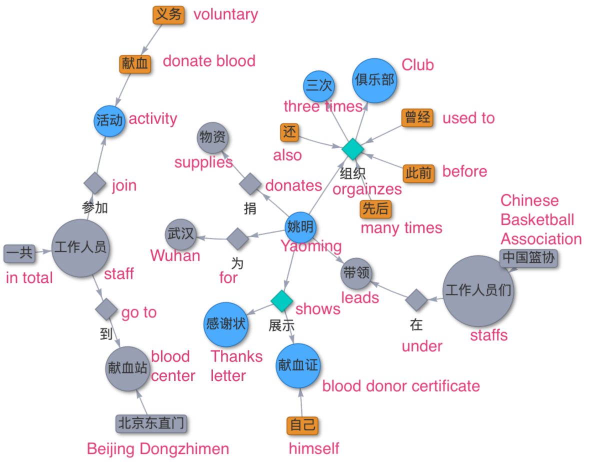

Figure 3: Knowledge graphs constructed from a long speech and their abstract (highlighted in color) before Component Selection.

fast pace as speech streams flow in. As a result, the visualization of knowledge graphs from the complete speech may become very complex and even harder to understand when comparing to raw text. Such a phenomenon goes against our initial intention which is to capture the key concepts and their relationships so that the users can be clear at a glance. To this end we perform the following three steps to obtain an abstract of the knowledge graphs:

Key-set Extraction. First, a set of keywords are obtained from the complete speech transcription by selecting words with the highest TF-IDF calculated across the set of all documents in the NLPCC 2017 corpus[4]. In addition, nodes with high degree centrality are picked into the set of key nodes.

Abstract Filtering. Second, we apply a rule to obtain an abstract of the knowledge graph construction from the speech system. The entity-relation triplets and entity-attribute pairs are filtered by any intersection between the sets of keywords and key nodes.

Component Selection. Finally, the largest connected component is selected from the knowledge graph. We find this step particularly effective because small components usually do not have a clear meaning as shown in the upper left corner of Figure 3.

3 Conclusion

To the best of our knowledge, the HAO-Graph system we presented above is the first automatic knowledge graph construction system from a speech. The system renders knowledge graphs by topics from voice streams in real-time and a summarized knowledge graph can be highlighted after a long speech. Feel free to register in our system[5] and a video demonstration can be found at https://drive.google.com/file/d/1oz0suuCf9Ab2VLlolfMoW_g3YZeSKAKw/view?usp=sharing.

[4] http://tcci.ccf.org.cn/conference/2017/taskdata.php

[5] http://kelab.mininglamp.com

References

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, 2019.

[Ehrlinger and Wöß, 2016] Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. In *SEMANTiCS*, 2016.

[Hart, 1986] Anna Hart. Knowledge acquisition for expert systems. Technical report, School of Computing, Lancashire Polytechnic, Preston, 1986.

[Manning *et al.*, 2014] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

[Povey *et al.*, 2011] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, December 2011.

[Soon *et al.*, 2001] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. *Computational linguistics*, 27(4):521–544, 2001.

[Sowa, 1987] John F Sowa. Semantic networks. 1987.

[Stewart *et al.*, 2019] Michael Stewart, Majigsuren Enkhsaikhan, and Wei Liu. Icdm 2019 knowledge graph contest: Team uwa. *arXiv preprint arXiv:1909.01807*, 2019.

[Tang *et al.*, 2008] Jie Tang, Mingcai Hong, Duo Liang Zhang, and Juanzi Li. Information extraction: Methodologies and applications. In *Emerging Technologies of Text Mining: Techniques and Applications*, pages 1–33. IGI Global, 2008.

[Wu and Wu, 2019] Minghui Wu and Xindong Wu. On big wisdom. *Knowl. Inf. Syst.*, 58(1):1–8, January 2019.

[Wu *et al.*, 2019] Xindong Wu, Jia Wu, Xiaoyi Fu, Jiachen Li, Peng Zhou, and Xu Jiang. Automatic knowledge graph construction: A report on the 2019 icdm/icbk contest. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1540–1545. IEEE, 2019.

[Zhou and Zhao, 2019] Junru Zhou and Hai Zhao. Head-driven phrase structure grammar parsing on Penn treebank. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2396–2408, Florence, Italy, July 2019. Association for Computational Linguistics.