Knowledge Graphs Aided Entity Relation Networks Using Chinese Quality Supervision News

Jie Leng, Zhihua Yan, and Xijin Tang

***Abstract*— This paper, with the aid of knowledge bases and graph neural networks, tries a new way to information extraction and correlation analysis in the Chinese open domain. Chinese quality supervision news are taken as the corpus. We categorize event types based on topics generated via LDA. As for entities in texts, their types and relations are recognized through a combination of local and distant knowledge graphs. Those relations that do not exist in the knowledge graphs can be predicted through graph neural networks, with fully connected dependency syntactic trees as inputs. The correlation analysis on entities and events from quality news provides supports for relevant departments of the government, manufacturers, and consumers.**

***Index Terms*—Open information extration, Knowledge bases, Relation extraction, Graph neural networks.**

I. Introduction

EXTRACTING informative knowledge from massive texts is an important task in the natural language process (NLP). The traditional tasks of information extraction generally refer to the named entity recognition, entity relation classification, and event extraction. Access to open domain information not only enriches and enlarges the corpus but also makes it possible to analyze the semantic content beyond predefined categories and limitations. Open information extraction for English has already achieved much progress, due to its mature and diverse technological methods and tools. Chinese text mining still needs practical methods besides man-made syntax character templates, considering most effective solutions in English corpus cannot be applied to Chinese corpus directly [1]. For the relation extraction task, knowledge bases are useful for distant supervision [2,3]. Chinese scholars have already proposed different kinds of knowledge bases online and maintain continuous enrichment. In the meantime, deep learning methods with attention mechanisms [4-6], and the latest graph neural networks (GCNs) [7-9],have outstanding performances in NLP tasks. Various pre-training models also provide supports for Chinese open information extraction and knowledge discovery from Internet texts.

Inspired by the available methods, we try to explore a new way of entity extraction and correlation analysis for Chinese open-domain text with the aid of knowledge bases and GCNs. Quality supervision news from the Chinese Quality Net are taken as the corpus. Events are derived from topic analysis via the Latent Dirichlet Allocation model (LDA) [10]. We have crawled a series of entities such as manufacturers, government departments, (industrial) products, and quality inspection items from quality inspection records on the official websites of the state, municipalities, and provinces [11]. The results are stored as a local knowledge base,which helps conduct the named entity recognition for the nouns or noun phrases (raw entities) in texts. Entities not exist in the local knowledge base inherit the conceptual definitions in an distant open concepts knowledge graph. Finally, entities are categorized into eight types, i.e. ***Organizations***, ***Products***, ***Materials***, ***Foods***, etc. Another open knowledge graph is employed to check the binary relations of entities in the context. Though a majority of the entity tuples cannot be judged in this way, GCNs can make predictions. The results of correlation analysis on events and entities provide pieces of advice for relevant departments of the government, manufacturers, and consumers. The related entity tuples can also help enrich the knowledge bases. Fig.1 gives an illustration of the architecture.

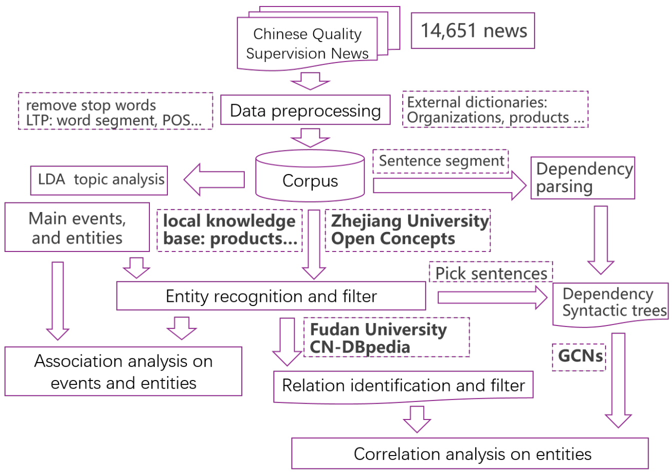


Fig. 1: Overview of the whole analysis process

II. Related Work

Open information extraction on English sentences has achieved excellent performances [12]. Researches on Chinese open relation extraction seem much more difficult in terms of complex grammar and spelling. Zhao et al. indicate that the methods for open information extraction are statistical learning models with knowledge bases mined from large-scale and heterogeneous Web resources [13]. Tseng et al. are considered the first to apply NLP methods, such as word segmentation, part-of-speech (POS) tagging, and dependency parsing, to extract entity-relation triples from Chinese free texts [1]. Qiu et

All authors are with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences, Beijing, e-mail: {lengjie,zhyan,xjtang}@amss.ac.cn.

Manuscript received August 24, 2021; revised October 25, 2021.

al. design a syntax-based system to identify relations with semantic patterns (combined with a Chinese taxonomy *Extended Cilin*) through a novel double propagation algorithm [14]. That method depends on verbs to indicate relations. Li et al. make full use of dependency parsing and rely on verb predicated sentences [15]. Qin et al. present an unsupervised Chinese open relation extraction method [16]. The system uses distance constraints to generate candidates of entity triples. Relation words are selected after ranking. Finally, the extracted relation words and sentence-level rules are used to filter candidates. Jia et al. summarize three common phenomena in Chinese linguistics and provide seven kinds of Dependency Semantic Normal Forms for unsupervised Chinese open relation extraction [17]. The model shows its superiority to all previous methods while being heavily limited by the performance of NLP techniques, such as word segmentation. But this method is considered to help label samples automatically to support models with machine learning or deep learning algorithms. Ref. [18] and Ref. [19] focus on dependency syntactic analysis combined with traditional machine learning methods and rules. Wang et al. employ bidirectional long short term memory (Bi-LSTM) neural networks to extract features from dependency syntactic analysis for relation extraction [20].

Dependency syntactic analysis is demonstrated to be useful for extracting accurate relations for words in a sentence. Dependency syntactic trees of adjacent sentences are fit for GCNs as a common practice. We adopt densely connected graph neural networks with a multi-head attention mechanism from Ref.[21] to explore related entity tuples.

III. DATA COLLECTIONS

In this section, we introduce the datasets for the research. They are collected from the China Quality Net (https://www.cqn.com.cn/,CQN) and official websites. The former provides the corpus, and the latter releases official records which are the sources of the local knowledge base.

A. Quality supervision news

CQN is maintained by the State Administration for Market Regulation. Policies and regulations on markets and producers, together with product quality information issued by state and local governments, are released on the CQN. Boards, like Product Quality Investigation, Hot Civic Issues, Complaints, Rights Protection, Recall, and Public Opinion, are selected to be closely related to consumers. Python crawler technology is applied to download news IDs, released dates, board names, URLs, titles, sources, and contents. The news are accumulated from January 2019 to November 2020, to a total of 14,702. Table 1 lists the detailed information of each board. After filtration, 14,651 articles are retained for analysis.

TABLE I: THE DETAILS OF THE CRAWLED NEWS OF SIX BOARDS FROM CHINESE QUALITY NET

Board name	The numbers	Released time range
Product Quality Investigation	170	2010.8.19-2017.9.15
Hot Civic Issues	5,443	2018.6.4-2020.11.26
Complaints	1,996	2012.8.1-2020.11.27
Rights Protection	1,815	2014.9.12-2020.10.15
Recall	3,260	2017.11.24-2020.11.27
Public Opinion	1,995	2016.1.5-2020.11.27
Total	14,702	

B. Records from official websites

The official websites provide much valuable information for research. We select three parts as follows:

Product inspection records. From the official website of the State Administration for Market Regulation (http://www.samr.gov.cn/zljds/zlgg/bsgg/), 5,212 inspection records during 2017 and 2018 are crawled. Each record contains the product, manufacturer, retailer, conducting government department, testing institute, and defective items. There are about 490 thousand inspection records, which are described in Ref.[11], from the corresponding websites of 30 provinces and municipalities.

Production licenses of specific industrial products. Production licenses of specific products are disclosed every year (http://www.samr.gov.cn/zljds/), following the supervision regulations. We download the license information from 2010 to 2019, including product names and manufacturers, about 18,250 records in total.

Enterprise registration records. The website of the State Administration for Market Regulation releases enterprise registration records every day (http://zwfw.samr.gov.cn/gs/). There are a total of 33,279 new enterprise names in 2019.

After combining the records above and removing duplicates and noises, we construct a local entity knowledge base for the product quality area, consisting of 91,458 organizations, 65,376 products, and 7,416 inspection items. It will be helpful for word segmentation and entity recognition as a user-defined lexicon.

IV. ENTITY RECOGNITION AND RELATION IDENTIFICATION VIA KNOWLEDGE GRAPHS

Besides the above local knowledge bases, distant knowledge graphs are applied to recognize entities and relations. Entities are generalized into 8 categories. The entity pairs are simplified to be related or unrelated.

A. Entity recognition and classification

Based on the local entity knowledge base provided from Section III, some entities belonging to the pre-defined types are recognized, but more nouns or noun phrases remain unidentified. Given a word, ***Open Concepts*** (http://openconcepts.openkg.cn) released by Zhejiang University provides conceptual explanations at three levels. Each level expresses a different aggregation extent: level 1 always indicates the essential attribute, level 2 generalizes some possible categories the word may fall in, and level 3 gives some instances for reference. We retain level 2 as a priority and level 3 as an alternative if level 2 is missing. Furthermore, whole entities are grouped into 8 types. The results are listed in Table 2.

Morever, we find that some entity mentions may belong to more than one group. *Benzoic acid*, for example, can be

TABLE II: THE NUMBERS AND TYPES OF RECOGNIZED ENTITIES

Entity types	The numbers	Examples
Organizations	24,637	factory, police, supermarket, court, travel agency, operators
Products & Devices	1,881	telephone, computer, furniture, tool, respirator, clothing, air-condition
Materials	932	methanol, plastics, building material, dyestuff, fiber, glass, grease, benzoic acid
Foods	2,120	health care product, food additive, cake, wine, grape, nuts
Health & Medicines	1,158	medicine, specimen, virus, bacteria, toxin, Chinese traditional medicine, prescription, immunity, detectable rate
Cosmetics	189	hand sanitizer, aloe, hair dye, perfume, sun cream, facial mask, hand cream, facial wash
Inspection Items	991	function, label, performance, packaging, ingredient, structure, accessory, fiber content, color fastness
Laws	116	Evidence, capital, the right to know, Contract Law, deposit, property right, proprietor, liquidated damage
Total	34,367	

TABLE III: FIVE EVENT TYPES AND CORRESPONDING KEYWORDS

Topic No.	Keywords for each topic	Event types
1	consumer, complaint, service, staff, company, contract, goods, merchant, purchase, work, proprietor, indemnity, quality, furnish, rights protection, rights and interests, price	Consumers' complaints & rights protection
2	product, disqualification, production, quality, inspection, batches, standard, samples, qualified, enterprise, supervision, nominal, random inspection, project, detection, spot check	Products quality inspections
3	food, disposal, exceed permission, Shanghai, qualified, production, cosmetics, food additives, furnish, evidence, sales return, code, certificate, medicine, material, unqualified label, detection, non-conformity	Inspections & quarantines of import & export commodities
4	information, platform, illegal, network, advertisement, company, market, sales, user, release, manage, staff, department, legal case, enterprise, punishment, publicity, crime, fraud	Law enforcements on cybercrimes
5	recall, product, management, market, defect, information, consumer, quality, service, vehicle, enterprise, China, production, supervision, website, nation, hidden risk	Products recall

regarded as a kind of mineral, or one specific medicine. Some companies and products or entities of other types share the same names. In this case, we need to judge whether the mentions are organizations or not. A familiar case: "Apple Inc.", which is often written as "*Apple*" for short, may be confused with fruit "*apple*" without enough contexts. "Xiaomi Tech" and "*millet*" are similar to the case. One simple solution is to compare types of entities in the context. If the confusing mention has a type following other entities, they are likely to share the same topic. The mention may not be judged as a company entity. Otherwise, the mention belongs to ***Organization***.

B. Relation identification

Relations of Entity tuples are firstly checked with ***CN-DBpedia*** (http://kw.fudan.edu.cn/) [22] developed by Fudan University. It claims to be the largest and encyclopedic open Chinese knowledge graph with 223,507,519 relations of entities. Given a mention, relevant entity lists and knowledge triples are provided through APIs. In this research, the relations of two entities are defined as related or unrelated. For one entity in a document, if another entity in the text exists in the graphs, they will combine as a related tuple, otherwise an unrelated one. As a result, entity tuples in 2095 pieces of news are identified, while more tuples from the rest news remain as the test set to explore later.

V. CORRELATION ANALYSIS ON EVENTS AND ENTITIES

Stop words are removed from news at first. Texts are segmented through the ***LTP tools*** (http://www.ltp-cloud.com/) from Harbin Institute of Technology. The whole vocabulary consists of 182,228 words. The number of topics is 5 after perplexity analysis. The LDA model is leveraged to analyze the distributions of topics on the corpus, and extract keywords for each topic. Topics are generalized as event types based on keywords. Table 3 lists the details.

Entities may inherit the topic distributions from texts. Considering that each entity has its categories, we get distributions of events for each entity type through matrix multiplications. Firstly, derived from the LDA model, probability distributions of topic for whole documents (**doc-topic**) can be represented as matrix $D_T^{N\times M}$, here N stands for the number of total news, namely $14,651$, M for topics, equals to 5. Inspired from Bag of Words model, entities distribution in news (**doc-entity**) can also be expressed through a matrix ${D_E}^{K\times N}$, where K is sum of entities recognized in whole news, 34,367 actually. The values in matrix ${D_E}^{K\times N}$ are filled with 0 or 1 which indicates whether k-th $1 \le \mathrm{k} \le \mathrm{K}$) entity exists in the n-th $(1 \le \mathrm{n} \le \mathrm{N})$ news. Similarly, the relations for entities and their categories (**entity-category**) can be described as a matrix ${E_C}^{P\times K}$, in which 1 represents belonging to current types, others are all 0, and $\mathrm{P} = 8$. Eventually, we calculate the portions of each entity type in different events, (**entity_catogory-event**), through the formula:

$$EC_E^{P\times M} = E_C^{P\times K} \times D_E^{K\times N} \times D_T^{N\times M}.$$

For comparison, the same normalization is applied to each row vector in ${EC_E}^{P\times M}$. The radar plot in Fig. 2 depicts the final results. ***Materials*** and ***Inspection Items*** mainly occur in news about quality inspections. A high proportion of ***Foods*** appear in the news focusing on inspections and quarantines of import and export commodities. Entities in ***Laws*** always exist in the news describing complaints, protection of consumers' rights and interests, enforcement of laws on cybercrimes. Other entity types, such as ***Organizations***, ***Products & Devices***, widely occur in whole quality news.

From another perspective, similar and frequent entities occur in similar events. As can be seen from Table 4, there

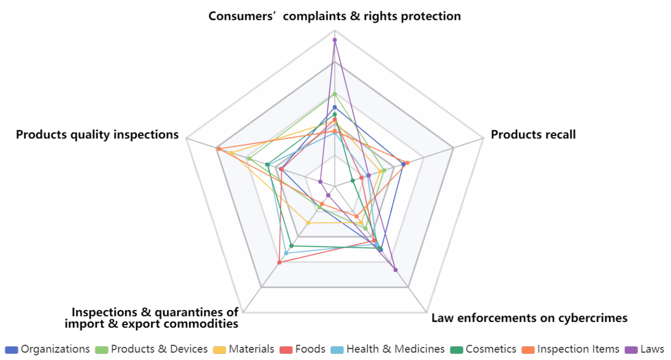


Fig. 2: Distribution of events for entity categories in quality news

are many highly related entities for each event. The key entities enhance that the second and the third events are mainly about quality inspections. There are some related government departments with corresponding duties in the fourth and the five events. These results verify the rationality of the event classification methods for quality news.

TABLE IV: MAIN ENTITIES FOR EACH EVENT

Topic No.	Key entities for each topic	Event types
1	merchant, telephone, law, requirement, content, awareness, evidence, boss, scope, factory, electronic, hospital, amount, computer, court, microblog, tool, suit, medicine, function, air-conditioner, deposit, supplies, capital, healthcare products, address, television, right to know	Consumers' complaints & rights protection
2	sample, requirement, specification, defective item, label, national standard, formaldehyde, purity, results of spot check, range, performance, sign, qualification rate, medicine, telephone, heavy metal, dye, tag, function, plastic, content, law, site, structure	Products quality inspections
3	medicine, sample, requirement, hospital, food additives, illness, patient, symptom, vitamin, range, purity, national standard, law, ingredient, virus, bacteria, protein, telephone, tag, benzoic acid, heavy metal, specification, effect, supermarket, nutrition	Inspections & quarantines of import & export commodities
4	law, content, medicine, awareness, telephone, police, range, requirement, merchant, function, hospital, evidence, illness, microblog, capital, court, electronic, health care products, boss, computer, effect, patient, telecom, virus	Law enforcements on cybercrimes
5	Range, telephone, requirement, specification, awareness, medicine, content, law, attachment, airbag, charger, national standard, function, label, National Medical Products Administration, patient, electronic, site, hospital, respirator, Volkswagen, sample, resource, supplier, tag, kindergarten, Defective Product Administration and Technical Center of Anhui Province	Products recall

VI. CORRELATION ANALYSIS OF ENTITIES

In this part, we try to explore more related entity tuples. Authors in Ref.[23] hold the standpoint that through learning language-agnostic sentence representations, GCNs with universal dependency parses trained on one language can be applied to another language. It is believed that the attention mechanism is effective in distinguishing priorities [24]. We assume that a model combined with GCNs and the attention mechanism will settle the problem. After reviewing many works of literature and comparisons, Guo et al. meet our expectations even though its experimental data sets are all English [21]. In Ref.[21], the accuracies of the experiments on single and multiple sentences are more than 80%. Fig.3 shows the simplified model structure.

As a common practice, sentences are connected with the root nodes in dependency syntactic trees. The linkage is named "next". The initial graph consists of dependency trees and can be represented with an adjacency matrix A. $A_{ij} = 1$ and $A_{ji} = 0$ means that there is an edge going from node i to node j. In GCNs, given the input feature representation h^{l-1}, the output for node i at the l-th layer can be defined as follows:

$$h_i^l = \rho\left(\sum_{j=1}^{n} \mathrm{A}_{ij} W^l h_j^{l-1} + b^l\right)$$

where W^l and b^l are the weight matrix and bias item for the l-th layer, respectively; ρ is one kind of activation function; h_i^0 means the initial state of input word i, which is represented through word embedding vectors added with other features. As for the attention mechanism in [21], the original dependency graph is transformed to be fully connected and edges are weighted by an adjacency matrix $\tilde{A}$ with a multi-head attention mechanism. The t-th adjacency $\tilde{A}^t$, corresponding to the t-th attention head, $t = 1, \ldots, N$,is calculated as:

$$\tilde{A}^t = softmax\left(\frac{QW_{Qi} \times (KW_{Ki})^T}{\sqrt{d}}\right) V$$

where Q, K and V are collective representation $h^{l-1}.W_{Qi}$ and W_{Ki} are parameter matrices to be trained. d is the dimension of input vectors. In addition, dense connections are introduced to capture more structural information. The representations of the node i in all layers are concatenated as:

$$g_i^l = \left[h_i^0; h_i^1; \ldots; h_i^{l-1}\right]$$

Calculation for each node is rewritten as:

$$h_{ti}^l = \rho\left(\sum_{j=1}^{n} \tilde{A}_{ij}^t W_t^l g_j^l + b_t^l\right)$$

The outputs of all dense layers are concatenated as $h_{out} = \left[h^1; \ldots; h^N\right]$ and are fed to a linear combination layer which is expressed as $y = Wx + b$. Finally, sentence representations without entity tokens and representations of each entity are concatenated and applied to a logistic regression classifier for a prediction. Experiments on the cross-sentence n-ary relation extraction and sentence-level relation extraction perform well.

The training set is constructed based on thousands of related or unrelated entity tuples from 2095 pieces of identified news in Section IV. These entity tuples will be abandoned if one mention contains another. For example, entity tuple *(charger of mobile phone, charger)* will be considered as uninformative and will not be used. An entity tuple will also be discarded if

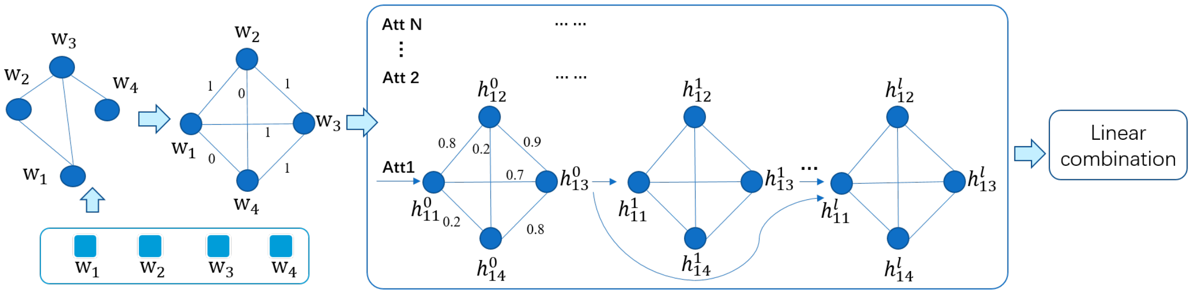


Fig. 3: Densely connected GCNs with the multi-head attention mechanism

the distance between the entities is more than one sentence. The strict distance requirement helps improve the precision of models [25]. Another kind of entity tuples with no relation will be negative samples for training. All of the distances of sentences in negative instances should be less than or equal to 2. We also consider the balance of the numbers of positive samples and negative ones. Finally, the numbers of positive and negative are 2,411 and 2,390 respectively.

Entity tuples with unknown relations form the test set. We only pick out important entities for prediction. Each word in each document has its Term Frequency and Inverse Document Frequency (TF-IDF) values for the whole corpus. Entities in each news with the top five TF-IDF values and form pairs with each other. Finally, the test samples are 57,048 in total. The parameter settings of the experiment are almost the same with Ref. [21].

We combine the related entity tuples from testing sets after prediction with the training sets for analysis. At first, we rank the most popular entities for each type and their related partners. Plots in Fig. 4 illustrate the results. The sizes of nodes in the chord diagrams indicate the numbers of the related entity tuples, and the colors of nodes represent corresponding types. For a tuple, the color of a link is the same as the entity with a leading role. Legends are the same with Fig. 2.

As can be seen from Fig.4, for the type of ***Organizations***, vehicles and Internet companies show their popularities. Most entities related to them are components or devices of cars, products, and services. In the area of ***Products & Devices***, respirators for surgery or dust occupy a large portion, which may reflect the influences of COVID-19 in daily life. Other large nodes are “furniture”, “computers”, “chargers”, “electric cars”, “toys”, “air conditioners”, and other daily supplies. Those entities related to ***Products & Devices*** are various ***Inspection Items***. For the ***Materials***, “glass”, “plastic”, “chemical fertilizer”, “fiber” are closely linked to various products and related ***Inspection Items***. ***Foods*** are always about “food additives”, “wine”, “tea”, “cakes”, “pork”, and “seafood”. They are also highly related to ***Health & Medicines***, especially for “health care products” that link with “health care food”. While focusing on ***Health & Medicines***, “virus”, “illness” and “bacteria” are attractive. The number of entities in ***Cosmetics*** are fewer than that of other types in results, where “facial mask”, “lipstick”, “hair dye” list top 3. Ingredients from ***Materials*** and some ***Inspection Items*** are connected. Due to the specificity of the quality news, ***Inspection Items***, such as “specifications”, “labels”, “user manuals”, “accessories”, “flexible wires” and “fiber contents”, have widely related entities. The links are comparatively homogeneous to ***Products & Devices***, ***Foods*** and ***Health & Medicines*** respectively. Entities concerned in ***Laws*** are mainly “property rights”, “rights to know”, “Contract Law”, “Criminal Law”, “copyright” and many others connected with civil life. ***Organizations*** and ***Products & Devices*** are most involved in this area.

In general, the entity “State Council” and the entity “the police” are active, same with “take-out restaurants” and many Internet companies. Entities closely related to daily life are “respirators”, “household appliances”, “food additives”, “health care products”, “facial masks”, which remind of enhanced supervision. Frequent ***Inspection Items***, such as “specification”, “colorfastness”, “fiber content”, may draw the attention of manufacturers. “Proprietor”, “deposit”, and “property rights” which are outstanding in ***Laws*** may reflect some social issues.

From another point of view, we try to find whether these related entity tuples will help enlarge the knowledge graph we have used before. Take “respirator” as an example, Fig.5 shows its relation network presented on the ***CN-DBpedia*** platform. Knowledge for “respirator” is about product types, characteristics, and explanations. We sum up and filter entities related to the “respirator” with high probabilities and draw the relation networks as shown in Fig.6. More entities around the “respirator” are detected and apprehensible in different circumstances.

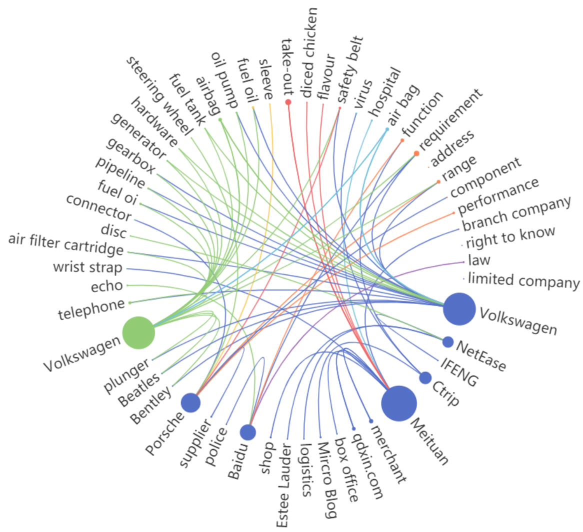
VII. Conclusion

In this paper, we propose a new method for Chinese open information extraction. Unlike designing complicated syntax or grammar patterns, the method makes full use of knowledge graphs and GCNs. The event types are defined based on the LDA model, and the entity types are identified through knowledge graphs. More related entities are discovered through the GCNs. The results drawn from the quality supervision news provide supports for people in need.

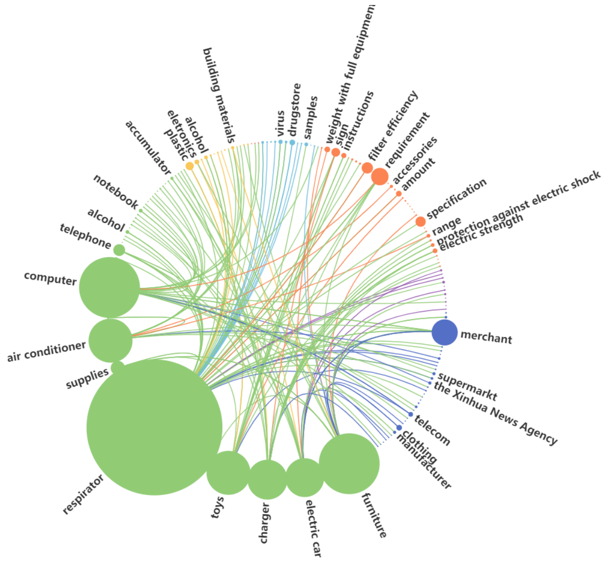
In future work, we will try to expand the types of entities' relations and discover more informative knowledge for the entity relation networks.

Acknowledgment

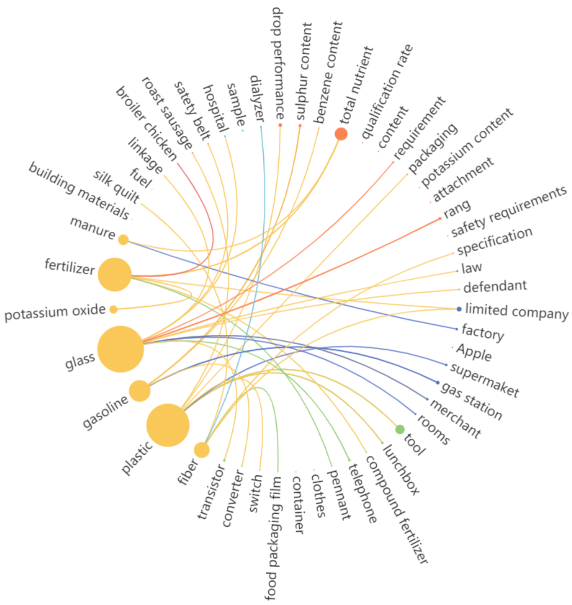
This work is supported by the National Natural Science Foundation of China (No. 71731002,71971190). The computations were done on the high performance computers of State



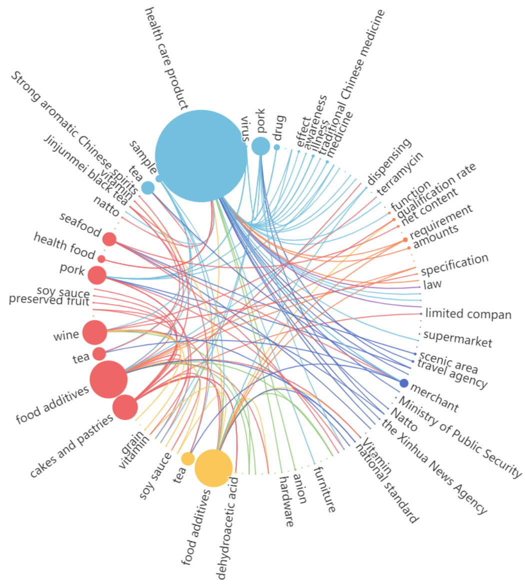
(a) Organizations



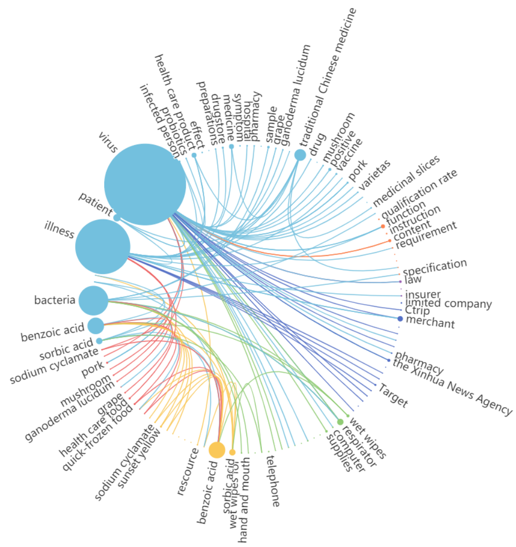
(b) Products & Devices



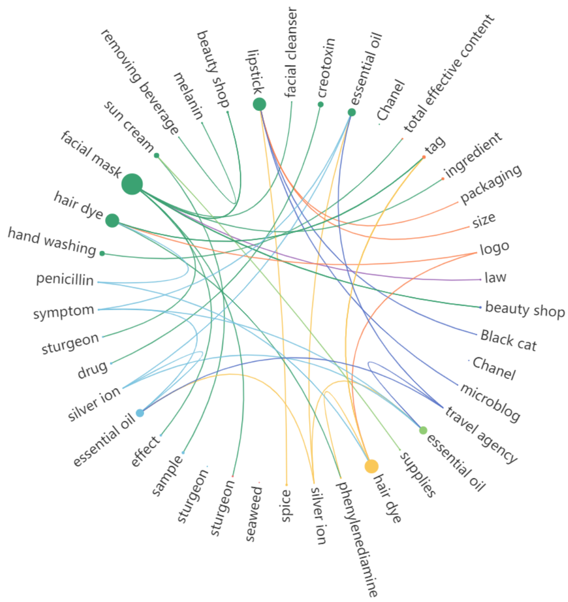
(c) Materials



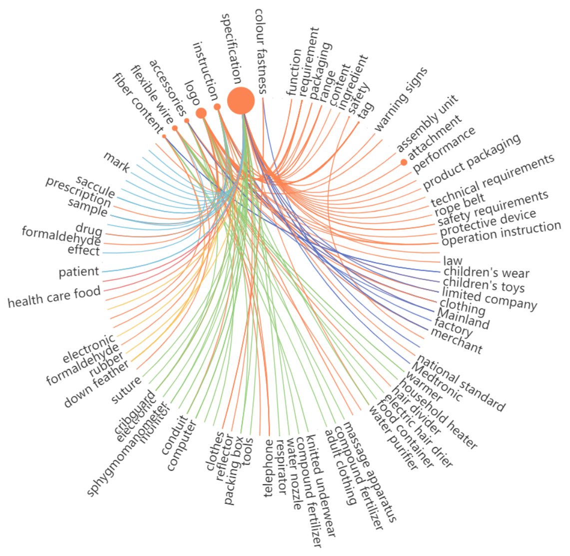
(d) Foods



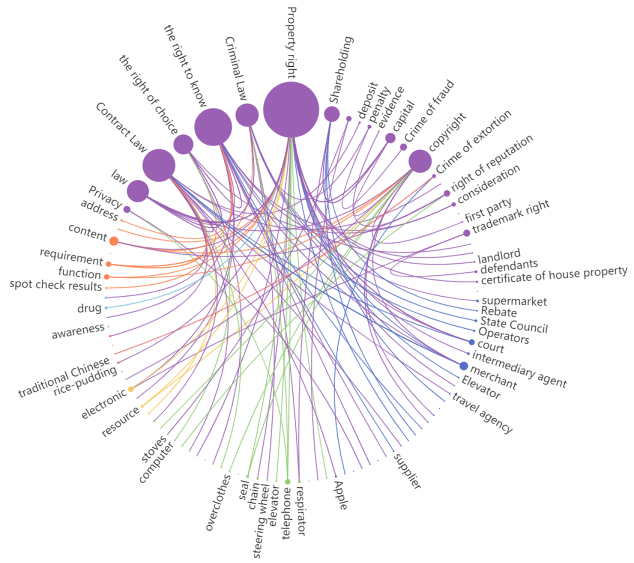
(e) Health & Medicines



(f) Cosmetics

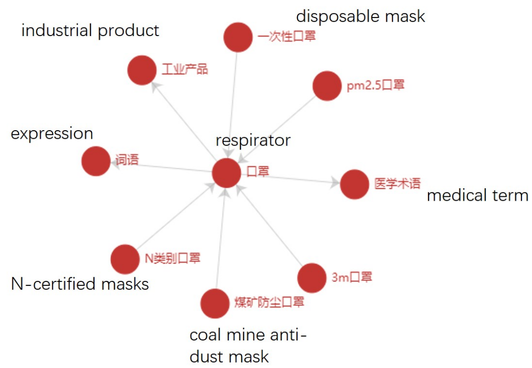


(g) Inspection Items

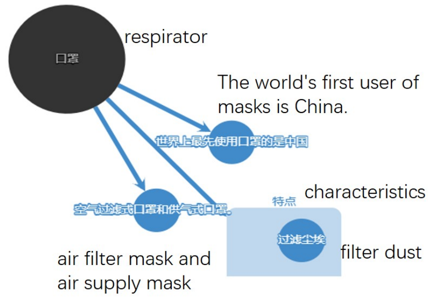


(h) Laws

Fig. 4: Main related entities for each kind of entity type



(a) entity networks for mention respirator



(b) relation triples for resistor

Fig. 5: Knowledge graphs for respirator in CN-DBpedia

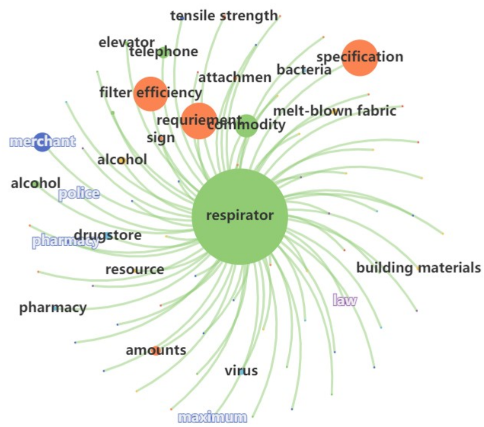


Fig. 6: New related entity networks for respirator

Key Laboratory of Scientific and Engineering Computing, Chinese Academy of Sciences, and took approximately 60 hours to run. The authors would like to thank the anonymous reviewers for their comments.

REFERENCES

[1] TSENG Y H, LEE L H, LIN S Y and et al. Chinese open relation extraction for knowledge acquisition. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*,2014. pp:12-16.

[2] MINTZ M, BILLS S, SNOW R, and et al. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 2009:1003-1011.

[3] JI G L, LIU K, HE S Z, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2017:3060-3066.

[4] LIN Y K, SHEN S Q, LIU Z Y. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016:2124-2133.

[5] LI W J, LI T, QI F, Chinese entity relation extraction based on multi-features self-attention Bi-LSTM. *Journal of Chinese Information Processing*. 2019,33(10):47-56, 72. (in Chinese)

[6] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. 2016:207-212.

[7] THOMAS N K and MAX W. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*. 2017.

[8] SONG L F, ZHANG Y, WANG Z G, and et al. N-ary relation extraction using graph state LSTM. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*,2018:2226-2235.

[9] FU T J, LI P H, MA W Y. GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual*

Meeting of the Association for Computational Linguistics,2019:1409-1418.
[10] BLEI D M, Ng A Y, Jordan M I. Latent dirichlet allocation. *Journal of Machine Learning Research*,2003:993-1022.
[11] TANG X J, YAN Z H, LENG J, and et al. New Trial toward Products Quality Supervision by Notices and News released on the Internet. In Proceedings of the 41st Conference of Japan Creativity Society, International Session Volume, pp2-6, 2019. ISSN 1883-0749
[12] YANG B, CAI D F, YANG H. Progress in open information extraction. *Journal of Chinese Information Processing*,2014, 28(04):1-11+36. (in Chinese)
[13] ZHAO J, LIU K, ZHOU G Y, et al. Open information extraction. *Journal of Chinese Information Processing*,2011,25(06):98-110. (in Chinese)
[14] QIU L K and Zhang Y. 2014. ZORE : A syntax-based system for Chinese open relation extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*,2014:1870-1880.
[15] LI M YYANG J. Open Chinese entity relation extraction method based on dependency parsing. *Computer Engineering*,2016, 42(6):201-207. (in Chinese)
[16] Qin B, Liu A A, and Liu T. Unsupervised Chinese open entity relation extraction. *Journal of Computer Research and Development*,2015, 52(5):1029-1035. (in Chinese)
[17] JIA S B, EE S J, LI M Z, et al. 2018. Chinese open relation extraction and knowledge base establishment. *ACM Transactions on Asian Language Information Processing*,2018, 17(3):1-22.
[18] ZHENG L M, QI S S, TIAN L J, et al. Research on entity relation extraction of news texts for food safety events. *Transactions of the Chinese Society for Agricultural Machinery*,2020, 51(07):244-253. (in Chinese)
[19] YAO X M, GAN J H, and XU J, Chinese open domain oriented n-ary entity relation extraction. *CAAI Transactions on Intelligent Systems*,2019, 14(3):597-604. (in Chinese)
[20] WANG B, WANG K, WANG C and et al. Chinese entity relationship extraction for journalism. *Telecommunication Engineering*,2021, 61(1):50-57. (in Chinese)
[21] GUO Z J, ZHNAG Y and LU W. Attention Guided Graph Convolutional Networks for Relation Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019:241-251.
[22] XU B, XU Y, LIANG J Q and et al. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System. In *Proceedings of International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. 2017:428-438.
[23] WASHI U A, PENG N Y, CHANG K W. GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction. arXiv preprint arXiv:2010.03009, 2021
[24] ASHISH V, NOAM S, NIKI P and et al. Attention is all you need. In *Proceedings of 31st Conference on Neural Information Processing Systems*. 2017:6000-6010.
[25] WANG D F, HU W, CAO E and et al. Global-to-local neural networks for document-level relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 2020:3711-3721.

Jie Leng is a Ph.D. student in Academy of Mathematics and Systems Science, Chinese Academy of Sciences and University of Chinese Academy of Sciences,Beijing. Her research interests are text mining and knowledge management.

Zhihua Yan is a post-doctoral in Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His research interests are data mining and decision support.

Xijin Tang is a professor in Academy of Mathematics and Systems Science, Chinese Academy of Sciences and University of Chinese Academy of Sciences,Beijing. Her research interests are meta synthesis, decision support system and knowledge science.