Migration knowledge graph framework and its application

Tongzheng Pu, Ming Huang*, and Jingjing Yang
The School of Information Science and Engineering, Yunnan University, Kunming, Yunnan, 650091, PR China

*Corresponding author’s e-mail: huangming@ynu.edu.cn

Abstract. Comprehensive knowledge of migration involves many fields, such as demographic, economic, social, and political. However, the results of migration theme knowledge retrieval and query are often the knowledge of a single domain, lacking of direct association between different fields. A spatiotemporal migration knowledge graph framework based on migration datasets and Baidu Encyclopedia is presented, and the validity and accuracy of the model are verified by an experiment implemented on the data of 2000. Visualization and Neo4j graph database technologies are used to migration knowledge graph framework which is perfectly adequate for both human beings and machine processing such as graph mining and knowledge reasoning. The migration knowledge graph can quickly and accurately obtain the comprehensive knowledge of migration, significantly improve the ability of human being to apply and analyze the knowledge and data of migration, which has wide theoretical and practical value in the field of migration.

1. Introduction
Knowledge graph is a large-scale semantic network composed of entities, concepts and their semantic relationships, which is used in the knowledge representation. In 2012, it was presented by Google and applied in semantic search. In recent years, with the development of technology, knowledge graph has become an important support resource in many fields, such as information retrieval, question answering system, recommendation system. With the continuously increasing of application requirements, knowledge graph has been used in professional domains. In [1], YAGO, the core of semantic knowledge, is proposed to unify WordNet and Wikipedia. And then, a methodology for enriching large knowledge bases of entity-relationship-oriented facts along the dimensions of time and space was presented [2]. Reference [3] put forward a three-layer framework reference model to solve product problems by using manufacturing knowledge graph, which can help solver making right decision with right knowledge at right time. In [4], a space object situation information service based on the Knowledge Graph was constructed, aiming at different scales of components in space object situation. Reference [5] established a high-quality medical question answering knowledge graph based on relevant professional knowledge in medical question-answering research. Later, a medical knowledge graph was constructed and applied to intelligent medical diagnosis in [6]. As one of the most important research trends in software engineering, software knowledge graph was proposed for the first time in [7]. In [8], a medical knowledge graph is constructed from the electronic medical record text of knee osteoarthritis patients to support intelligent medical applications. In addition, some scholars have explored the application of knowledge graph in the field of radio monitoring, and achieved good results [9-10]. International migration has increased rapidly as a result of the deepening processes of globalization. In 2019, there were around 272 million international migrants in the world,
which equates to 3.5 percent of the global population [11]. It play a dominating role in the social, economic, and demographic dynamics of both sending and receiving countries; however, it also brings problems and challenges for immigrant management department. Scholars have striven to provide general explanations for the phenomenon of human migration, and have constructed models, analytical frameworks, conceptual approaches, empirical generalizations, and related theories researches to explain it [12].

In view of established theories such as push-pull theory, neoclassical approach, economic theories, interdisciplinary research approach and comparative analysis of social network and immigration system has been used to migration analysis [13]-[16]. Furthermore, quantitative analysis and visualization methods were proposed [17].

In this paper, we try to solve the above problems by building migration knowledge graph (MKG) on the grounds of global migration datasets (GMDs) and Baidu Encyclopedia [18]-[20]. The GMDs column fields are divided into entities and attributes, and different relationships are introduced to connect the entity nodes in knowledge graph. At the same time, the general knowledge extracted from Baidu Encyclopedia is added into MKG to provide more background information and make it easier for researchers to obtain the possible information they may need. The rest of this paper is organized as follows: Section 2 introduces the related background information about migration. Section 3 describes the proposed Framework. Section 4 describes the application of migration knowledge graph, and conclusion and future work are described in section 5. The main work of this paper is as follows:

- GMD datasets statistics and visualization.
  Important patterns and trends of migration activities can be found through statistical analysis and visualization of GMD datasets. For example, IOM’s Global Migration Data Analysis Centre describes the station-temporal trends in migration incidents in the world, and examines their characteristics, for example, location, type, number and policy. For visual analysis of GMDs, social network analysis techniques are often used to output human friendly graph results as shown in the figure below.

- Migration knowledge graph framework.
  A migration knowledge graph model is presented to resolve the problems of domain specialization, flexibility, and dynamic variation, to enable the adaptation of a general knowledge graph model to the migration domain.

2. Background

In large part, migration related to the broader global economic, social, political and technological transformations that are affecting a wide range of high-priority policy issues [12]. It requires thorough understanding of the history of human civilization, and panoramic view to understand dynamics of population, geography, resources, economy, politics, culture and other aspects of today's world is necessary. All in all, the perspective of the study of international migration must be both synchronic and diachronic.

To put it crudely, international migration research is to interpret migration behaviour reasonably and promote the formulation of migration policies reasonably, so as to achieve the reasonable and effective sharing of global resources and the peaceful coexistence and common prosperity of global people. In particular, international migration study, probes can be summed up in the following two questions:

- Who, where, when, why and how to move across borders?
- What is the likely impact of international migration on individuals, families and related groups, regions, countries and even the world as a whole?

As just described, the knowledge of international migration is the explanation of the spatial distribution law, process and the interaction between causes and phenomena of migration, which can be used to answer the above questions. The phenomenon of international migration has the basic characteristics of time, space and attributes, and there is a certain logical relationship between these characteristics, which constitutes a big data network. Among them, the time and space is the basic condition for the existence of the phenomenon of immigration.
3. Framework of migration knowledge graph

From the perspective of figure, knowledge graph is a semantic network, which is a powerful tool that change the way we do data organization, retrieval and analysis; and knowledge is expressed in the interconnected nodes and edges in knowledge graphs. With that in mind, using reliable data of migration and diversity of Internet data, show migration phenomenon, the nature of the transition of "data - knowledge", and provide new ideas for the research on migration.

Therefore, we build a Migration Knowledge Graph (MKG) from GMDs and Baidu Encyclopedia, which aims at semantically representing the truth of migration information in the form of machine readable graph structure. The whole processing flowchart is shown in figure 1.

3.1. Knowledge acquisition

Migration data can be classified as text and numeric data, which mainly from statistical data of countries and regions around the world, world migration reports, relevant data collected by the International Organization for Migration and relevant literature data. Comparatively speaking, the computer representation of migration knowledge has not made breakthrough progress. The current research of migration is mainly concentrated on theoretical models, development, trends and policies, and so on. So, in this paper, natural language processing technology is used in international migration knowledge processing.

3.1.1. International migration named entity. It refers to extract the named entity from text by computer. For the complexity of artificial construction features and the inaccuracy of migration text segmentation in the traditional methods, a migration named entity recognition method based on deep learning is proposed [21].

3.1.2. Migration spatio-temporal information. According to the International Organization for Migration, migrants are the movement of persons who away from their place of usual residence and across an international border to a country of which they are not nationals, and the temporal and spatial features are obvious.

Based on the representation, time can be divided into implicit and explicit time; explicit time can be extracted by rule-based approach, while for implicit time, maximum entropy, and neural network were used in time extracted. In natural language description, toponym, coordinates and spatial relations and other included in spatial information; however, due to the limitations of international migration information and data, the region, country or region and serial number adopted by the International Organization for Migration are used to represent the spatial information in this paper.
3.1.3. **Attribute.** It is an important part of the international migration entity, and the value is the specific reflection of attribute. Entity attribute extraction aims to extract attributes and corresponding attribute values. Currently, entity attribute extractions mainly use analysis theoretical and methods, for example, syntactic relation, supervised learning, topic model and coordination method. Furthermore, different information for attribute and treatment from different data has been standardized according to information standards, so that the requirements of data integration, processing and usage can be researched.

3.1.4. **Migration relationship extraction.** It aims to extract immigration entities and relationships between entities from text or semi-structured data, for example, spatial relationships, attribute relationships, state relationships. After years of research, relation extractions have made dramatic advances in terms of performance, but it is still an open problem and a great deal of problems remained to solve further finally. In this paper, according to predefined relationships, entity pairs which satisfying the given relationship category are extracted by deep learning model.

3.2. **Migration Knowledge Fusion.**
For different type of volume data, problems of entity ambiguity, relationship conflict and classification can be efficiently solved by knowledge fusion. In this paper, according to the international migration data with the international organization for migration study classification rules matching information, realizing the unified classification system; Secondly, in view of the real problems such as ambiguity and conflict relationship, make full use of the international organization for migration professional vocabulary and Baidu Encyclopaedia, extraction of entity and space relationship and the realization of information entity disambiguation; Finally, Baidu Encyclopedia and text extraction were used for information fusion.

3.3. **Organization and management of knowledge**
Knowledge graph can be stored based on table structure and graph structure. At present the industry recognized graph data model including the attribute, the resource description framework (RDF) and triple hypergraph, among them the first two are widely used in graph database products. Therefore, knowledge graph can use relational database and graph database, such as Neo4j, HugeGraph, TigerGraph. In contrast, the efficiency of relational database is low, while graph database is more conducive to data reading, writing, storage and query. Based on the advantages of graph database, this paper chooses Neo4j database to represent and store relations, such as relations between countries, time and space.

3.4. **Knowledge Graph Visualization**
In the process of implementation, Django which is an open source Web application framework written in Python for web development, and JavaScript are used to add a variety of dynamic functions to the web pages to provide users with a smoother and more beautiful browsing effect, neo4j is a native graph database aims to quickly manage, store, and traverse nodes and relationships, all MKG nodes, edges and relationships are stored in it.

The hierarchical knowledge representation model is adopted to construct the "process-relationship" knowledge graph representation form of migrants. Semantic unit are in the form of triples: $G=\langle E, R, T, S \rangle$, $E$ means the basic characteristics of the migration entities, such as time, space, behaviour and attributes, $R$ said migration entities state, process, and relationship, $T$ is the time; $S$ is the place. In figure 2, we see that China and America are two entities; the edge between two countries means migration relationship which says the conceptual relationship of flow in and out. There is comparability between the two countries in same time, such as GDP, CPI, and population.
4. Application of migration knowledge graph

The practice of the above steps, the migration knowledge graph platform was completed (which contain a total of 6210 entity nodes and 203923 relationships). On this basis, such as migration information intelligent retrieval, spatio-temporal information display and the functions for origin–destination of migration dynamic displaying are done by multiple data processing methods.

An experiment on the application of the migration knowledge graph is used in the analysis of migration network in this paper. The number of migration on any given connection for the year 2010 and the world list of origin-destination pairs connected by direct input were selected in MKG, the resulting migration graph comprises $N=230$ vertices denoting countries (In order to ensure the authenticity of the data, the nodes include unknown countries, regions and other situations) and $E=1833$ edges accounting for the presence of a direct migration connection (No statistics and statistical error data are included). A partial detail view of the knowledge graph is shown in figure 3.

Figure 3. A partial view of the knowledge graph (query results in 2000)
In the case, an edge between vertices $C_i$ and $C_j$ represents the presence of at least one of the two possible migration relationships between the two countries $i$ and $j$, and the weight $W_{ij}$ represents the number of migrants; the degree $k_i$ is the number of connections to other countries, whose elements take the value 1 if an edge connects the vertex $C_i$ to the vertex $C_j$ and 0 otherwise, and the strength $s$ is the total number of migrants handled by any given countries. The average degree of the network is $\langle k \rangle = 2E/N = 15.94$, while the maximal degree is 318. In particular, the average shortest path length, measured as the average number of edges separating any two nodes in the network, shows the value $\langle l \rangle = 2.37$, very small compared with the network size $N$. After data analysis and calculation, the results were presented as follows in figure 4-5.

Figure 4. Top 50 countries by number of immigrants in 2010 word cloud map (inflow)

Figure 5. Top 100 countries by number of emigrants in 2010 rectangular treemap (outflow)

5. Conclusions
Knowledge graph technology has been introduced into migration network based on migration facts along the dimensions of time and space in this paper. The construction process can be dynamically updated through data, which can be supplemented to the migration knowledge graph to help realize the application of migration intelligent Q&A and analysis of migration network. The example reflects the benefits of analysing dynamic monitoring data in a generalized network model within a single knowledge graph. The future works will combine artificial intelligence and big data technology, and provide visual interface to help experts using MKG. Graph embedding technique will be soon applied to MKG for migration predication.

Acknowledgments
This work was supported by the National Natural Science Foundation of China under Grant Nos.
61963037, 61863035 and 11564044. (Corresponding author: Ming Huang.)

References

[1] Suchanek F.M., Kasneci G., and Weikum G. (2007) Yago: a core of semantic knowledge. In: Proceedings of the 16th international conference on World Wide Web. Canada. pp. 697-706.

[2] Hoffart J., Suchanek F.M., Berberich K., Weikum G. (2013) YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence, 194: 28-61.

[3] He L., Jiang P. (2019) Manufacturing Knowledge Graph: A Connectivism to Answer Production Problems Query with Knowledge Reuse. IEEE Access, 7: 101231-101244.

[4] Lan C., Lu W., Xu Q., Zhou Y., Shi Q., Lyu L. (2020) Construction of Space Object Situation Information Service Based on Knowledge Graph. IEEE Access, 8: 22625-22640.

[5] Jiang Z., Chi C., Zhan Y. (2021) Research on Medical Question Answering System Based on Knowledge Graph. IEEE Access, 9: 21094-21101.

[6] Chai X. (2020) Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning. IEEE Access, 8: 149787-149795.

[7] Lin ZQ., Xie B., Zou YZ., et al. (2017) Intelligent Development Environment and Software Knowledge Graph. Journal of Computer Science and Technology, 32: 242–249.

[8] Li X., Liu H., Zhao X., Zhang G., Xing C. (2020) Automatic approach for constructing a knowledge graph of knee osteoarthritis in Chinese. Health Information Science and Systems, 8:1-8.

[9] Li C. X., Lu Q. N., Huang M., Yang J. J., Chen L. Q. (2021) Construction of Radio Transmission Equipment Type Approval Knowledge Graph and its Application. In: 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). Chongqing. pp. 956-961.

[10] Lu Q., Li C., Yang J., Li P., Huang M., Li L. (2020) Radio Regulatory Knowledge Graph Framework and Its Application. In: 2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE). Xi’an. pp. 61-65.

[11] Migration I, Nations U. (2020) World migration report 2020. International Organization for Migration, Switzerland.

[12] Willekens F., Massey D., Raymer J., Beauchemin,C. (2016) International migration under the microscope. Science, 352: 897.

[13] Abel G. J., Sander N. (2014) Quantifying Global International Migration Flows. Science, 343:1520-1522.

[14] Tranos E., Gheasi M., Nijkamp P. (2014) International Migration: A Global Complex Network. Environment and Planning B Planning and Design, 42:4-22.

[15] Fawcett James T. (1989) Networks, linkages, and migration systems. International Migration Review, 23:671–88.

[16] Ghatak S., Levine P., Price S.W. (2010) Migration theories and evidence: an assessment. Journal of Economic Surveys, 10: 159-198.

[17] Sander N., Abel G. J., Bauer R., Schmidt J. (2014) Visualizing Migration Flow Data with Circular Plots. Vienna Institute of Demography: Working Papers.

[18] IOM OIM. (2020) Migration data portal. https://migrationdataportal.org/.

[19] Oxford. (2015) DEMIG C2C, version 1.2, Limited Online Edition. http://www.migrationdeterminants.eu

[20] United Nations Population Division. (2019) World Population Prospects: 2019 Revision. https://data.worldbank.org/indicator/SM.POP.NETM

[21] Yadav V. and Bethard S. (2018) A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In: Proceedings of the 27th International Conference on Computational Linguistics. New Mexico. pp. 2145-2158.