A Global Perspective on Multi-dimensional Knowledge Association Mining in Network Communities (Postprint)

Authors: Xiao Lu, Zhao Zhihui, Chen Guo

Date: 2023-04-01T16:15:51+00:00

Abstract

[Purpose/Significance] Online communities contain multiple knowledge units with complex and intricate relationships among them. It is essential to perform mining of multi-type knowledge associations in a unified and compact manner while retaining the global information of knowledge units. [Method/Process] This paper proposes an implementation scheme for multi-type knowledge association mining in online communities. First, three typical types of knowledge units (users, texts, and words) in online communities and their multiple relationships in knowledge communication are extracted and modeled as a hypernetwork. Second, network representation learning algorithms are utilized to represent nodes in the hypernetwork as low-dimensional dense vectors within a unified feature space. Finally, multi-type knowledge association calculations are performed based on the node vectors. [Results/Conclusion] Taking the DXY Cardiology Forum as a case study, experiments are conducted to validate the effectiveness of the scheme. The scheme preserves complete information of knowledge units, conducts knowledge association mining within a unified low-dimensional feature space, and the resulting knowledge associations meet the demands of diverse knowledge organization scenarios in online communities.

Full Text

Holistic Perspective Multi-knowledge Relations Mining in Network Community

Xiao Lu[1], Zhao Zhihui[1], Chen Guo[2] [1]School of Journalism, Nanjing University of Finance & Economics, Nanjing 210023 [2]School of Economics & Management, Nanjing University of Science and Technology, Nanjing 210094

Abstract: *[Purpose/Significance]* Network communities contain multiple knowledge units with intricate relationships among them. It is necessary to conduct unified and concise multi-knowledge relations mining while preserving the global information of knowledge units. *[Method/Process]* This paper proposes a solution for multi-knowledge relations mining in network communities. First, three typical knowledge units (users, texts, and words) and their multiple relations in knowledge communication are extracted to construct a supernetwork. Second, network representation learning algorithms are employed to represent nodes in the supernetwork as low-dimensional dense vectors in a unified feature space. Finally, multi-knowledge relations are calculated based on these node vectors. *[Result/Conclusion]* Experiments conducted on the DXY cardiovascular forum demonstrate the effectiveness of the proposed solution. This approach not only retains complete information of knowledge units but also enables knowledge relations mining under unified low-dimensional features, with the resulting knowledge relations satisfying the diverse requirements of knowledge organization scenarios in network communities.

Keywords: knowledge relations mining; supernetwork; network representation learning; network community
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2020.06.012

Network communities have rapidly accumulated vast numbers of users and resources, becoming important venues for knowledge exchange and utilization. However, these communities generally suffer from resource fragmentation and coarse-grained knowledge organization, causing valuable knowledge to be submerged in massive datasets and making it difficult for users to effectively acquire and utilize information. A key challenge in network community knowledge organization lies in the complex forms and significant granularity differences among knowledge units—users, posts, comments, and topics are all knowledge units that must be addressed, while both intra-granularity and cross-granularity units exhibit intricate multiple relationships. It is necessary to systematically organize these relationships from a global perspective and conduct multi-relations mining and revelation to promote deeper and finer knowledge organization in network communities.

Current research on knowledge relations discovery in network communities primarily focuses on unary relations, with multi-relations mining mainly employing supernetwork technology to provide multi-descriptions of multiple unary relations. These relation types remain independent, essentially representing a multi-presentation of unary relations. True multi-relations mining from a holistic perspective has not been achieved, and network heterogeneity in nodes and relations also complicates the utilization of cross-granularity multi-relations.

How can we concisely and effectively reveal multi-knowledge relations in network communities while maintaining a global perspective and shielding interference from cross-granularity knowledge units and heterogeneous relations? We propose a solution: constructing a knowledge supernetwork for network communities to ensure globality, then using network representation learning to represent knowledge units as uniform low-dimensional dense vectors, enabling all inter-unit relations to be computed based on these vectors. This scheme can rapidly generate a multi-knowledge relations system for network communities to guide knowledge organization, with feasibility and effectiveness verified using the medical network community DXY cardiovascular forum as a case study.

Related Research

2.1 Network Community Multi-granularity Knowledge Unit Relations Mining

Network communities mainly contain three granularities of knowledge units: users, texts, and words. Multiple relationships exist among both same-granularity and cross-granularity units. Unary relations mining includes: (1) User-user relations, divided into direct and indirect relations. Direct relations are obtained by analyzing following and replying behaviors, while indirect relations are derived through analyzing direct associations. (2) Text-text relations, measured using text similarity with common expansion sources including domain ontologies and search engines. (3) Word-word relations, where tags are important objects for mining semantic relations, with relation strength primarily derived from co-occurrence frequency using methods like social network analysis and LSA. (4) User-text relations, subdivided into direct and indirect relations. Direct relations are obtained from publishing and sharing behaviors, with relation strengths manually thresholded by behavior type. Indirect relations are derived through text-text or user-user relations, essentially representing collaborative recommendations based on texts or users. (5) Text-word relations, established when extracting tags or feature words from texts, with strength calculation either binary (presence/absence) or using feature weights like TF-IDF and information gain. (6) User-word relations, obtained through transitive user-text and text-word relations.

Additionally, scholars have attempted to fuse multi-relations using supernetwork technology. For instance, Xiao Lu constructed a knowledge supernetwork for multi-granularity knowledge aggregation in network communities, incorporating words, sentences, and texts with co-occurrence, grammatical, inclusion, and affiliation relations. Wang Chuanqing et al. built a digital resource supernetwork for deep aggregation, containing literature knowledge, copyright holders, and material carriers with citation, co-occurrence, and coupling relations.

2.2 Network Representation Learning Applications in Knowledge Organization

Research on network representation learning for knowledge organization focuses on three aspects: (1) Scholar collaboration and paper impact prediction. Zhang Jinzhu et al. used network representation learning on co-authorship networks to predict academic collaboration through vector similarity. Lin Yuan et al. applied representation learning to multiple co-occurrence networks (authors, keywords, institutions), overcoming traditional methods' focus on highly productive scholars when analyzing potential collaborations. Fan Wei et al. mapped papers, authors, and venues into low-dimensional dense vectors to construct heterogeneous academic network representation models that better predict paper impact. (2) Knowledge representation learning. Zhang Xiaokun et al. incorporated external word vectors into text information network representation learning to fuse semantic and structural features. Zhu Guojin et al. built network text representation models integrating named entities and word vectors. Zhu Jingwen et al. applied network representation learning to HowNet for cross-lingual and semantic unit vector representation. (3) Social network user relation analysis. Han Zhongming et al. fused user attributes and network structure through network representation learning for multi-angle user relation mining. Yang Yizhuo et al. fused username and topological information for cross-network user identity matching.

Design of Multi-knowledge Relations Mining Scheme Based on Network Representation Learning

3.1 Basic Idea of Multi-knowledge Relations Mining Based on Network Representation Learning

A multi-knowledge relations system forms the foundation for refined knowledge organization and enhanced knowledge services in network communities. Current mining primarily focuses on unary relations, with multi-relations research mainly describing multi-associations through heterogeneous networks without achieving true holistic multi-relations mining. Network representation learning is a technique that represents network nodes as low-dimensional dense vectors with reasoning capabilities based on the initial network, simultaneously preserving network information and enabling network reconstruction. Our proposed mining approach is: first, use supernetworks to describe multi-granularity knowledge units and their multi-relations in a unified network, which serves as the basis for holistic knowledge relations mining using network representation learning. Then, employ network representation learning to represent knowledge units as uniform vector sets, where one vector represents one knowledge unit and relations are characterized by vector similarity. We term this collection of knowledge units and their relations the network community multi-relations system, which can be rapidly processed and analyzed by computers to support multi-dimensional knowledge organization as domain background knowledge. The specific concept is illustrated in Figure 1 [Figure 1: see original paper].

3.2 Process Flow of Multi-knowledge Relations Mining Based on Network Representation Learning

The network community multi-relations system is obtained by applying network representation learning to the knowledge supernetwork, while multi-granularity knowledge unit identification and multi-relations mining form the foundation for supernetwork construction. Our proposed process comprises three main components, as shown in Figure 2 [Figure 2: see original paper]. Notably, the first part enumerates multi-relations mining methods for multi-granularity knowledge units, with specific approaches determined according to network community characteristics in practice.

(1) Network Community Knowledge Unit Base Construction. This includes multi-granularity knowledge unit identification, multi-relations mining among knowledge units, and structured representation and storage.

Multi-granularity knowledge unit identification. Network community knowledge units mainly include three types: users, texts, and words. Texts can be further divided into full texts and sentences. Considering that most network community texts are short, traditional topic sentence extraction methods have limited effect, so sentence texts are not considered here. All subsequent references to texts denote full texts.

Knowledge unit multi-relations mining. User-user relations are divided into direct and indirect relations. Our multi-relations system only considers direct relations established through replying and following behaviors; indirect relations can be obtained through user-word and user-text relations. Word-word relations include two categories: those containing user tags obtained through statistical co-occurrence frequency, and those lacking user tags where topic extraction techniques (e.g., LDA) or domain dictionary-based methods automatically extract terms from texts. User-text relations are obtained by analyzing user operations on texts, with relation type thresholds manually determined and strength comprehensively decided by type thresholds and behavior intensity. Text-word relations are obtained by analyzing word presence in texts or tags, with strength calculation either binary (presence/absence) or using text feature weights (e.g., information gain, mutual information). User-word relations are derived through transitive user-text and text-word relations.

Structured representation and storage. Knowledge units are stored as: knowledge_{unit} = <entity, type, description>, where entity represents the knowledge unit, type indicates the unit type (text, user, or word), and description provides the unit description. Knowledge unit relations are stored as: knowledge_{relationship} = <entity1, entity2, relation_{type}, weight>, where entity1 and entity2 represent related knowledge units, relation_{type} indicates the relation type, and weight represents relation strength.

(2) Network Community Knowledge Supernetwork Extraction. Multiple relation types and strength calculation methods exist among users, texts, and words in network communities. Representing them with a single network can cause node confusion and unclear network structure. We therefore divide them into two single-node-type but heterogeneous-relation networks: a user relation network and a word relation network, then connect them through user-text and text-word relations to form a heterogeneous connected network. Considering that traditional heterogeneous network technology has limited capability in multi-network connection representation, we adopt supernetwork technology for network community heterogeneous knowledge relation network construction. American scholar A. Nagurney defines supernetworks as networks that are above and beyond existing networks, typically composed of multiple networks where nodes can be viewed as sets of networks and edges represent preferences for network combinations, enabling network structure adjustment through edge addition or deletion. The mathematical model of the network community knowledge supernetwork is similar to that in reference [11], with sentence knowledge subnetworks replaced by user knowledge subnetworks. The graphical model is shown in Figure 3 [Figure 3: see original paper].

(3) Network Community Multi-relations System Generation. Network representation forms the foundation of network analysis. Traditional methods include adjacency matrix-based and network graph-based representations. The former represents network nodes through row vectors, easily leading to high dimensionality, while the latter contains numerous associated edges, making analysis processes tend toward iterative or combinatorial approaches that greatly increase algorithmic time complexity and affect analysis effectiveness. Network representation learning maps network nodes to low-dimensional vectors, using vector distance or similarity to represent node relations. This method largely preserves the overall network structure while achieving network reconstruction, bridging the gap between real-world networks and network analysis, and demonstrating advantages in heterogeneous network analysis. We use network representation learning to represent knowledge units in the network community knowledge supernetwork as low-dimensional dense vectors, enabling holistic knowledge unit relation discovery while reducing adverse effects from relation heterogeneity on knowledge organization.

Network representation learning algorithms are mainly divided into structure-based and external information-combined categories. The former includes matrix eigenvector computation, matrix factorization, shallow neural networks, and deep neural networks, while the latter incorporates text information and edge labels. Based on analysis object characteristics, we select the LINE algorithm (Large-scale Information Network Embedding) based on shallow neural networks for network community knowledge unit representation. LINE addresses the lack of optimized objective functions for network structures in DeepWalk and node2vec algorithms while preserving both first-order and second-order similarities, demonstrating high applicability for large-scale networks. The complementarity of first-order and second-order similarities enables the algorithm to consider both local and global network structures.

Empirical Analysis

DXY is an important medical social media platform in China, with its forum ranking high among academic network communities in user and post numbers. However, current resource organization relies primarily on traditional publication time and pinning operations, lacking multi-dimensional organization oriented toward user and knowledge intrinsic relations. Taking data from DXY's cardiovascular specialty discussion board in the clinical medicine discussion zone as the data source, we construct a cardiovascular domain-oriented multi-relations system to verify feasibility and effectiveness.

4.1 Data Collection and Domain Dictionary Construction

We retrieved user post information from the DXY cardiovascular forum on March 17, 2019, using a train browser to crawl post texts, obtaining 65,364 texts. Each text was saved as a TXT file containing user, post, and reply information.

Since DXY cardiovascular forum lacks user tag functionality, we used a domain dictionary to identify word-granularity knowledge units. After comprehensive comparison, we selected the "Cardiovascular Internal Medicine" column from "39 Disease Encyclopedia" as the data source for extracting cardiovascular domain terms. "39 Disease Encyclopedia" provides detailed structured annotations for each disease in information boxes, serving as both important term supplements and basis for term category classification. We collected 2,211 terms, categorized into diseases (including conditions, aliases, complications), organs (pathological sites), symptoms, and diagnostics (diagnostic methods), as shown in Table 1 .

4.2 DXY Cardiovascular Forum Knowledge Unit Base and Knowledge Supernetwork Construction

We extracted user and post text information from TXT files, statistically analyzed co-occurrence frequencies to obtain user-user relations and strengths, user-text relations and strengths. Then, using the constructed cardiovascular domain term dictionary for text segmentation, we statistically obtained text-word relations and strengths, word-word relations and strengths, and user-word relations and strengths, as detailed in Table 2 .

4.3 DXY Cardiovascular Forum Multi-relations System Generation

Based on the constructed knowledge supernetwork, we applied the LINE algorithm to represent knowledge units in the DXY cardiovascular forum as low-dimensional dense vectors, with manually set vector dimensionality of 100, obtaining 168,086 knowledge unit vectors (partial results shown in Figure 4 [Figure 4: see original paper]). The first line in Figure 4 indicates 168,086 knowledge units in the DXY cardiovascular forum multi-relations system, each represented by a 100-dimensional vector. The first column (except the header) shows knowledge unit IDs in the system, with other columns representing corresponding vector values.

Knowledge unit relation strength calculation is key to constructing the multi-relations system. After representing knowledge units as low-dimensional dense vectors using LINE, the relation calculation problem transforms into vector similarity computation. Common methods include cosine similarity, correlation coefficients, Euclidean distance, and Mahalanobis distance. Since knowledge units are represented as low-dimensional vectors, we selected Euclidean distance to measure knowledge unit association strength, obtaining the DXY cardiovascular forum multi-relations system. Using the disease "hypertension" as an example, Table 3 shows the multi-relations set for word knowledge units. Notably, term type labeling during dictionary construction enables the multi-relations system to include both relation strength and type for word associations, providing the foundation for advanced knowledge services.

For better visualization of the multi-relations system, we applied PCA (Principal Component Analysis) for dimensionality reduction and used Python's Matplotlib for visualization. Figure 5 [Figure 5: see original paper] visualizes the multi-relations set for the word knowledge unit "hypertension."

To better compare the multi-relations system with supernetwork quality, we measured knowledge unit relevance based on the previous supernetwork extraction. Current network node similarity calculation methods include topology-based, attribute-based, and hybrid approaches. For simplified computation, we adopted the attribute-based method by constructing knowledge unit attribute vectors and using Pearson correlation coefficient to measure association, obtaining the knowledge unit relevance based on supernetworks. For clear comparison, Table 4 lists only the top 10 words highly associated with "hypertension." Comparison between Tables 3 and 4 shows that representation learning-based mining yields more comprehensive association term types (all four categories) and identifies potentially associated terms like "diabetes" and "coronary heart disease." Diabetes and hypertension are homologous diseases, and numerous studies have addressed coronary heart disease combined with hypertension. However, "cervical CT examination" in Table 4 was judged by domain experts as having low association with hypertension, demonstrating our method's effectiveness.

4.4 DXY Cardiovascular Forum Multi-dimensional Knowledge Aggregation Prototype System Design

Currently, DXY cardiovascular forum resource organization relies primarily on pinning operations without sub-forum-specific search functionality. Searching "hypertension" across the entire community yields results shown in Figure 6 [Figure 6: see original paper]. This keyword-matching list-based retrieval model cannot satisfy advanced knowledge service requirements. Therefore, we propose mining resource internal associations based on the multi-relations system to achieve multi-dimensional knowledge services through multi-dimensional aggregation of hit resources.

Using "hypertension" search results as an example, Figure 7 [Figure 7: see original paper] demonstrates multi-dimensional aggregation results manually adjusted from the mined DXY cardiovascular forum multi-relations system.

The left side shows first-level retrieval results, displaying high-relevance resources from document, term, and user dimensions. The right side shows second-level expansions: (1) document-dimensional expansion of "2010 Hypertension Prevention and Treatment Guidelines"; (2) term-dimensional expansion of "ECG" from the document "Hypertension ECG Question: Why Clockwise Rotation Occurs"; (3) term-dimensional expansion of "coronary heart disease" from term-level retrieval results; (4) document-dimensional expansion of user "heaven197898" from first-level results. Through multi-dimensional knowledge aggregation, users can obtain target documents while conducting related term expansions and user identification, achieving advanced forum knowledge services.

Our proposed supernetwork-based network representation learning scheme addresses challenges in network community multi-knowledge relations mining. DXY cardiovascular forum experiments demonstrate advantages: (1) Knowledge unit association mining comprehensively references relations with users, domain terms, and texts rather than single relation types, yielding more reliable results from a holistic perspective; (2) Transforming different knowledge unit types (users, domain terms, texts) into uniform low-dimensional dense vectors shields subsequent calculations from unit type differences and relation heterogeneity, enabling concise and effective multi-relations mining; (3) Retaining knowledge unit types allows resulting knowledge relations to maintain type differences (e.g., "user-term," "user-user") for effective differentiation in subsequent knowledge organization according to application scenarios.

Limitations include not considering network node text content in supernetwork construction and network representation learning. Current research explores optimizing network representation learning with external node information (e.g., text, tags). Future work will introduce external information-combined network representation learning methods for network community multi-knowledge relations mining.

References

[1] Chen Guo, Zhu Qianling, Xiao Lu. Knowledge aggregation for network community: Development, research foundation, and prospects[J]. Journal of Intelligence, 2017, 36(12): 193-197, 192. [2] Tian Bo, Fan Lingling. Research on online social network community detection method based on interaction behavior[J]. Journal of Intelligence, 2016, 35(11): 183-188. [3] Liu Bingyu, Wang Cuirong, Wang Cong, et al. Microblog community discovery algorithm based on dynamic topic model fusing multi-dimensional data[J]. Journal of Software, 2017, 28(2): 246-261. [4] Ning Yahui, Fan Xinghua, Wu Yu. Short text classification based on domain word ontology[J]. Computer Science, 2009, 36(3): 142-145. [5] Sahami M, Heilman T D. A web-based kernel function for measuring the similarity of short text snippets[C]//Proceedings of the 15th international conference on World Wide Web. New York: ACM, 2006: 377-386. [6] Zelikovitz S, Kogan M. Using web searches on important words

to create background sets for LSI classification[C]//Proceedings of the 19th international FLAIRS conference. Florida: AAAI Press, 2006: 598-603. [7] Huang Wei, Gao Junfeng, Li Rui, et al. Research on tag semantic distance measurement and visualization in Folksonomy[J]. New Technology of Library and Information Service, 2014(7/8): 64-70. [8] Tsui E, Wang W M, Cheung C F, et al. A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags[J]. Information processing & management, 2010, 46(1): 44-57. [9] Chen Guo. Research on network community knowledge aggregation based on domain concept association[D]. Wuhan: Wuhan University, 2015. [10] Shi Weijie, Xu Yabin. Research on microblog user interest discovery[J]. New Technology of Library and Information Service, 2015(1): 52-58. [11] Xiao Lu. Research on multi-granularity aggregation of network community academic resources based on knowledge supernetwork[J]. Journal of Intelligence, 2018, 37(12): 182-187, 194. [12] Wang Chuanqing, Bi Qiang. Research on deep aggregation of digital resources from supernetwork perspective[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(1): 4-13. [13] Zhang Jinzhu, Yu Wenqian, Liu Jingjie, et al. Research on scientific collaboration prediction based on network representation learning[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(2): 132-139. [14] Lin Yuan, Liu Haifeng, Wang Hailong, et al. Mining potential cooperation opportunities between scholars based on representation learning[J]. Journal of Intelligence, 2019, 38(5): 65-70. [15] Fan Wei, Han Jianing, Zhang Yuxiang. Paper impact prediction algorithm based on network representation learning[J]. Computer Engineering, 2019, 45(12): 160-165, 170. [16] Zhang Xiaokun, Liu Yan, Chen Jing. Text information network representation learning with external word vectors[J]. CAAI Transactions on Intelligent Systems, 2019, 14(5): 1056-1063. [17] Zhu Guojin, Li Chengqian. Network knowledge resource representation learning model[J]. Intelligent Computer and Applications, 2016, 6(3): 5-10. [18] Zhu Jingwen, Yang Yuji, Xu Bin, et al. Semantic representation learning based on HowNet[J]. Journal of Chinese Information Processing, 2019, 33(3): 33-41. [19] Han Zhongming, Zheng Chenye, Duan Dagao, et al. Correlation user mining algorithm based on multi-information fusion representation learning[J]. Computer Science, 2019, 46(4): 77-82. [20] Yang Yizhuo, Yu Hongtao, Huang Ruiyang, et al. Cross social network user identity matching based on fusion representation learning[J]. Computer Engineering, 2018, 44(9): 45-51. [21] Yin Ying, Ji Lixin, Huang Ruiyang, et al. Research and development of network representation learning[J]. Chinese Journal of Network and Information Security, 2019, 5(2): 77-87. [22] Tu Cunchao, Yang Cheng, Liu Zhiyuan, et al. Network representation learning: A survey[J]. Scientia Sinica Informationis, 2017, 47(8): 980-996. [23] Nagurney A, Dong J. Supernetworks: decision-making for the information age[M]. Cheltenham: Edward Elgar Publishers, 2002. [24] Qi Yuhu, Guo Jinli. Supernetwork research[J]. Journal of University of Shanghai for Science and Technology, 2013, 35(3): 227-239. [25] Zhou Hui, Zhao Zhongying, Li Chao. Research on heterogeneous information network representation learning methods[J]. Computer Science and Exploration, 2019, 13(7): 1081-1093. [26] DXY Forum. Cardiovascular specialty discussion board[EB/OL]. [2019-06-08]. http://cardiovascular.dxy.cn/bbs/board/47. [27] 39 Disease Encyclopedia. Cardiovascular internal medicine[EB/OL]. [2019-06-08]. http://jbk.39.net/bw/xinxueguanneike. [28] Zhang Liangfu, Li Cuiping, Chen Hong. Survey on SimRank computation on large graphs[J]. Chinese Journal of Computers, 2019, 42(12): 2665-2682. [29] Qiu Shaoming, Yu Tao, Du Xiuli, et al. Community detection algorithm based on multi-attribute similarity aggregation of nodes[J/OL]. [2019-10-13]. http://kns.cnki.net/kcms/detail/31.1289.tp.20190808.1701.023.html. [30] Baidu Baike. Diabetes hypertension[EB/OL]. [2019-10-13]. https://baike.baidu.com/item/%E7%B3%96%E5% [31] Zhang Wantong, Hu Yuanhui, Zhu Baochen, et al. Study on correlation between blood pressure level and thrombosis dynamics in coronary heart disease patients with hypertension[J]. Modern Journal of Integrated Traditional Chinese and Western Medicine, 2016, 25(2): 129-131, 140. [32] Zhang Xin, Li Rong, Huang Yuxiao, et al. Effect of OSAHS on ambulatory blood pressure and heart rate variability in coronary heart disease patients with hypertension[J]. Chinese Journal of Integrative Medicine on Cardio-/Cerebrovascular Disease, 2014, 12(1): 35-37. [33] Yang C, Liu Z Y, Zhao D, et al. Network representation learning with rich text information[C]//Proceedings of the 24th international joint conference on artificial intelligence. Buenos Aires: AAAI Press, 2015: 2111-2117. [34] Tu C C, Liu H, Liu Z Y, et al. CANE: context-aware network embedding for relation modeling[C]//Proceedings of the 55th annual meeting of the association for computational linguistics. Vancouver: ACL, 2017: 1722-1731.

Author Contributions

Xiao Lu: Drafted and revised the manuscript, finalized the paper.
Zhao Zhihui: Participated in paper discussions and literature organization.
Chen Guo: Revised the paper and provided the domain dictionary.

Note: Figure translations are in progress. See original paper for figures.