Topic Analysis of Chinese Text Using a Co-word Network LDA Model: A Case Study of Transportation Law Literature (2000-2016) Postprint

Authors: Ma Hong, Cai Yongming

Date: 2017-11-08T00:00:00+00:00

Abstract

Objective: By integrating the strengths of the probabilistic topic extraction method of the traditional LDA model and co-word network analysis for discovering connection structures among terms in literature, this study aims to mitigate the interference of high-frequency terms generated from a small number of documents and enhance topic coherence.

Methods: In the topic analysis of abstracts from traffic law literature, keywords from the documents are incorporated as a compound segmentation dictionary to improve semantic recognition accuracy. We propose the CA-LDA model (Latent Dirichlet Allocation Model with Co-word Analysis), which integrates co-word network analysis into the traditional LDA model. Co-word network topology parameters are employed as weights to control term-topic assignment (using betweenness centrality), prioritizing the extraction of terms with both high co-occurrence (betweenness) and high frequency.

Results: The CA-LDA model can identify high-frequency terms that co-occur across multiple documents simultaneously, making the resulting key term list more meaningful for topic analysis. The algorithm's results not only reflect term frequency probability but also discover hub terms through term associations, enabling a deeper understanding of research hotspots in the field.

Limitations: The number of topics K in the CA-LDA model is determined through perplexity-based standard cross-validation. If the K value is too large in practical analysis, it becomes detrimental to the classification and organization of literature topics. Future research needs to further process these results to consolidate topics.

Conclusion: This paper applies the model to analyze hotspot topics in traffic law research, achieving favorable results in processing large-scale bibliographic

data. Related research can be extended and applied to the automated processing of large-scale bibliographic data in various domains.

Full Text

Abstract

This study aims to improve the coherence of topic extraction while reducing interference from high-frequency terms generated by a small number of documents by combining the probabilistic topic extraction method of the traditional LDA model with the structural discovery capabilities of co-word network analysis. In analyzing topics from transportation law literature abstracts, we incorporated keywords from the literature as a compound dictionary for word segmentation to enhance semantic recognition. We propose the CA-LDA model (Latent Dirichlet Allocation Model with Co-word Analysis), which integrates co-word network analysis into the traditional LDA framework. The model uses co-word network topology parameters (specifically betweenness centrality) as weights to control term-to-topic assignment, prioritizing the extraction of terms that exhibit both high co-occurrence (intermediary) properties and high frequency. The CA-LDA model successfully identifies high-frequency terms that co-occur across multiple documents, producing a more meaningful vocabulary list for topic analysis. The algorithm' s results reflect not only term probability but also reveal hub terms through lexical associations, enabling deeper understanding of research hotspots in the field. A limitation of the CA-LDA model is that the optimal number of topics K is determined through perplexity-based cross-validation, which may yield large K values that complicate document classification. Future research should address this by developing methods to further consolidate topics. Applied to hotspot analysis in transportation law research, the model demonstrates good performance in processing large-scale document data and can be extended to automated processing of large-scale literature data across various domains.

Keywords: Latent Dirichlet Allocation Model with Co-word Analysis; Co-words; Network topology parameters; Stochastic gradient descent; Keywords in transportation law literature

1 Introduction

The continuous accumulation of information has led to increasingly vast text corpora that far exceed human reading capacity. Simultaneously, the growing storage of information in electronic text format facilitates computer-based text analysis. Topic modeling can discover latent semantic relationships (i.e., topics) between documents and terms, where a topic comprises a core event or activity together with all directly related events and activities [1]. Using natural language processing techniques, researchers can perform feature analysis on document content, extract topic concepts, track topics of interest, and rapidly and accurately obtain domain knowledge about hotspots and development trends. Topic analysis technology has become an effective tool for public

opinion analysis and research topic selection.

Topic models primarily use similarity calculations to determine whether new topics belong to known categories. Based on statistical knowledge, these models filter text information and then employ classification strategies to track relevant topics. Currently popular models include the Hierarchical Clustering Algorithm (HCA) [2-3], Language Model (LM) [4-5], Vector Space Model (VSM) [6-7], and Probabilistic Topic Models (PTM). Among these, Latent Dirichlet Allocation (LDA) is recognized as the most successful probabilistic topic model. Improvements to LDA include the fast collapsed Gibbs sampling LDA model [8], distributed learning LDA models [9-10], correlated topic models that break the exchangeability assumption [11], and non-parametric Bayesian HDP models (Hierarchical Dirichlet Processes) [12-13]. These advances have significantly improved topic analysis efficiency and expanded the scope of LDA applications.

While LDA can extract topics from texts, it does not consider term co-occurrence phenomena across multiple documents. Clearly, terms that co-occur in multiple documents form co-word networks that provide valuable guidance for topic coherence. Co-word Analysis, proposed by Callon et al., is another topic analysis technique that examines term co-occurrence frequencies. Through co-word matrices, it clusters thematically close terms to consolidate document topics [14]. Examples include Callon et al.' s analysis of polymer chemistry topics [15], Coulter et al.' s study of software engineering topics [16], and Zhang Xiaodong et al.' s research on computer integrated manufacturing topics [17].

However, co-word analysis is a literature correlation analysis based on existing term frequencies and co-occurrence patterns; it does not generate topics itself. Therefore, this paper combines the strengths of both approaches by proposing a Co-word Network LDA topic model (CA-LDA) that incorporates co-word network feature parameters into the traditional LDA framework to regulate topic generation. To address the computational complexity introduced by new parameters, we employ Stochastic Gradient Descent (SGD) optimization to improve algorithm efficiency, achieving good results in large-scale text processing.

2 Latent Dirichlet Allocation Model

Latent Dirichlet Allocation (LDA), proposed by Blei et al. in 2003, is a probabilistic topic language model that represents any document as a mixture of several latent topics following a Dirichlet distribution [18]. Topics are characterized by term frequency distributions, with topic mixture weights treated as K-dimensional parameter hidden random variables. The topic generation process is illustrated in Figure 1 [Figure 1: see original paper] [18-19].

Traditional LDA algorithms primarily use two methods: Bayesian Variational Inference (VBI) [20] and Hoffman' s Stochastic Variational Inference (SVI) [21]. The Gibbs sampling process in traditional LDA is time-consuming and sometimes produces stochastic gradient noise that affects convergence speed. The

traditional LDA algorithm proceeds as follows: (1) Sample a document-topic vector θ from a Dirichlet distribution with parameter α to determine the probability of each topic being selected; (2) Select a topic z from the topic vector θ; (3) Generate individual terms based on the term probability distribution of topic z. This process repeats, traversing all terms in the document until topics for all documents are generated.

The topic model includes a corpus $\{d\}$ containing a vocabulary set, documents, and terms belonging to K topics. d_{jz} represents the j-th term in document d being assigned to topic z. The joint probability density function of LDA [18] is:

$$P(\theta, Z, W|\alpha, \beta) = P(\theta)P(Z|\theta)P(W|Z, \beta)$$

Parameter α represents the Dirichlet distribution prior over topics in the document set, describing the relative strength of latent topics. β is a $K \times V$ matrix, where β_{ij} denotes the probability of generating the j-th term under the i-th topic, describing the probability that the j-th feature term belongs to the i-th latent topic. θ_d represents the multinomial distribution of document d over T topics, where θ is a document-level topic vector with each value corresponding to the probability of topic z appearing in the document. Both z and w are term-level variables: z is generated from θ, while w is generated from z and β. All terms w belong to K topics z.

3 CA-LDA Topic Model

3.1 Co-word Network Construction for Text Corpus

A co-word network is a special type of scientific knowledge network formed by the co-occurrence relationships of subject terms across multiple articles or paragraphs. In this study of abstract text analysis, the co-word network represents term co-occurrence across different articles. We define the co-word network graph $G(\text{Vertex}, \text{Edge})$, where Vertex represents the set of vocabulary network nodes (the complete vocabulary set from corpus D), and Edge represents connections between co-occurring terms: $\text{Edge} = \{e|(w_i, w_j), w_i, w_j \in \text{Vertex}\}$, meaning terms w_i and w_j co-occur within a text (or paragraph). This network is undirected, with adjacency matrix A of size $N \times N$, forming a large-scale sparse matrix.

Complex network topology parameters include node connectivity metrics (e.g., degree), centrality metrics (e.g., degree centrality, betweenness centrality, closeness centrality), and inter-node closeness metrics (e.g., clustering coefficient, cliques, community). These parameters indicate a term' s importance in the co-word network and its relationship closeness with other terms, serving as references for calculating term importance during topic generation. The proposed CA-LDA model uses betweenness centrality as a regulating variable for term classification, modifying LDA' s term generation probability and establishing a co-word network to improve topic coherence (degree centrality or closeness

centrality could also be used as regulating variables, with experimental results showing similar effects to betweenness centrality).

Betweenness centrality, derived from social network analysis, measures a node' s importance. In the co-word network, it effectively describes the intermediary relationships between terms, improving intra-topic cohesion when classifying terms around this central term. If σ_{ij} denotes the number of shortest paths between terms w_i and w_j, and $\sigma_{ij}(w_l)$ denotes the number of those paths passing through node l, then the proportion of shortest paths passing through node l between w_i and w_j is $\sigma_{ij}(w_l)/\sigma_{ij}$. According to the Faster Algorithm for Betweenness Centrality [22], the betweenness centrality of node l is defined as:

$$\mathrm{BC}(w_l) = \sum_{i \neq l \neq j} \frac{\sigma_{ij}(w_l)}{\sigma_{ij}}$$

Traditional LDA first selects a topic z for a document and then generates the document, with all terms in the document coming from one topic. The probability of generating document W from topics $z_1, z_2, \ldots, z_K$ [18] is:

$$P(W) = \prod_{i=1}^{K} P(z_i) \prod_{j=1}^{N_i} P(w_j|z_i)$$

The core of CA-LDA considers betweenness centrality when determining term classification. In complex network theory, a node with higher betweenness centrality is more important in the network [23]. Similarly, in the co-word network $G(\mathrm{Vertex}, \mathrm{Edge})$, terms with higher betweenness centrality are more important for topic partitioning. Based on this principle, CA-LDA adds a weight $\mathrm{BC}(w_j)$ to the probability of generating terms to control term classification, modifying the traditional LDA probability formula (5) to formula (6). Thus, terms with high betweenness centrality tend to be assigned to different term bags, while terms associated with these nodes tend to be assigned to the same topic.

$$P_{\text{CA-LDA}}(W) = \prod_{i=1}^{K} P(z_i) \prod_{j=1}^{N_i} \mathrm{BC}(w_j) P(w_j|z_i)$$

3.2 Stochastic Gradient Descent Optimization

The specific implementation can be represented in pseudocode. According to the Gibbs sampling algorithm [8,18], for posterior estimation $P(\theta, z|w)$, if given independent priors α and β, the topic distribution $P(w|\alpha, \beta)$ can be computed. Iteratively solving for α and β that maximize this expression becomes computationally complex when considering relationships between term vectors.

To address this, we improved the sample partitioning and sampling process of traditional Gibbs sampling using stochastic gradient descent to reduce iteration

count. We designed a stochastic gradient function storing CA-LDA model parameters: a topic vocabulary list $\{n_k\}_{k=1,v=1}^{K,V}$ recording the frequency of term v assigned to topic k, with vocabulary length V and topic count K. At each Gibbs sampling point, model parameters α and β can be obtained most rapidly along the gradient descent direction. Using the Gamma function from Gibbs sampling [18], $\gamma = \alpha + \phi \sum$, where $\phi \propto \beta\Psi(\gamma) - \Psi(\gamma \sum)$. The document-topic distribution prior parameters can be solved using gradient descent: $\nabla = \Psi(\alpha) - \Psi(\alpha + \sum)$. For each document' s initial γ and ϕ parameters, iteratively update the topic vocabulary list $\{n_k\}_{k=1,v=1}^{K,V}$ until convergence to obtain all topics z_{ij} and final generated terms w_{ij}.

3.3 CA-LDA Topic Model Algorithm Implementation

The CA-LDA algorithm with stochastic gradient descent optimization is illustrated in Figure 2 [Figure 2: see original paper]. Traditional LDA vocabulary is derived from probability distributions, prioritizing high-frequency terms. In contrast, CA-LDA adjusts using co-word network topology parameters (betweenness centrality in this paper) to prioritize terms with both high connectivity and frequency. This adjustment reduces interference from high-frequency terms generated by few documents, yielding high-frequency terms that co-occur across multiple documents, producing a more meaningful vocabulary list for topic analysis.

4 Experiments and Results

4.1 Data Acquisition and Descriptive Statistical Analysis

On July 23, 2016, we retrieved 6,230 document records from CNKI' s China Academic Journals Database using the search query: "Publication date between (2006-01-01, 2016-06-30) AND Subject=Traffic AND Subject=Law OR Subject=Regulation (exact match)" . Descriptive statistical analysis was performed on four dimensions: publication year, discipline, institution, and funding source, as shown in Figure 3 [Figure 3: see original paper].

Figure 3 reveals that transportation law research literature shows rapid growth with a slight decline in recent years. Publications concentrate in administrative law and local legislation, highway and waterway transportation, criminal law, and transportation economics. Major research institutions include Jilin University, Southwest University of Political Science and Law, Chang' an University, East China University of Political Science and Law, and China University of Political Science and Law. Primary funding sources include the National Natural Science Foundation, National Social Science Foundation, National Science and Technology Support Program, and National High-tech R&D Program (863 Program).

4.2 Corpus Generation and Information Transformation

We extracted abstracts from 6,230 Chinese transportation law documents. After text cleaning and segmentation, we removed pronouns and modal particles using a stopword dictionary. Simple segmentation yielded top 10 high-frequency terms: "mechanism," "norm," "construction," "problem," "development," "management," "research," "impact," "society," and "road" —terms with ambiguous meanings that offer limited value for topic analysis.

To improve semantic recognition, we adopted a compound word approach. We extracted keyword fields from all 6,230 documents, obtaining 11,565 unique terms after deduplication to serve as a compound word dictionary. All compound words were segmented and stored. For each abstract, we checked whether it contained all segmented components of a compound word; if so, we removed those components and added the compound word. This approach ensures analysis relies on abstract content rather than author-provided keywords, as the results do not correspond one-to-one with "keyword + abstract" combinations.

Figure 4 [Figure 4: see original paper] illustrates the text preprocessing for a sample article: "On the Distinction of Several Relationships in Road Traffic Accident Liability Determination" by Wang Feiyue from Central South University, published in *Politics and Law* (2016, Issue 6). Simple segmentation produced 189 terms, which reduced to 80 terms after removing duplicates and stopwords and adding compound words. Seven new compound terms were added: "liability presumption," "road traffic accident," "tort liability law," "public security management," "criminal liability," "traffic accident," and "traffic law." The author's original keywords were "traffic accident," "traffic-related accident," "non-action traffic violation," and "liability presumption" —not a one-to-one correspondence.

Analysis of the added compound words reveals: some match the article's keywords (e.g., "liability presumption," which never appeared in other documents' keywords); some are highly similar (e.g., "road traffic accident," similar to the article's "traffic accident" and "traffic-related accident," sourced from 367 documents including Sun Yurong's 2014 article in *Law Magazine*); and some are not in the keywords but appear in the abstract (e.g., "tort liability law" from 24 documents including Li Zhihao's 2010 article). These compound words align closely with abstract content, providing clearer semantics.

4.3 Hotspot Term Analysis Using CA-LDA Model

We applied the CA-LDA model to analyze topics in each article's abstract. Variable parameters include hyperparameters α, ϕ, and topic count K. α varies with topic count, typically set to $\alpha = 50/K$, with initial $\phi_0 = 0.01$. K is usually determined by testing different values and selecting the optimal through cross-validation using perplexity [18], calculated as:

$$\text{Perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(d_d)}{\sum_{d=1}^{M} N_d}\right)$$

where N_d is document d length (total terms) and $p(d_d)$ is the probability of generating document d_d from the test model. Lower perplexity indicates better generalization.

Based on perplexity calculations from 10 experiments on corpus D, the model achieved minimum perplexity at $K = 50$ (Figure 5 [Figure 5: see original paper]).

Using CA-LDA with co-word network topology parameters (betweenness centrality) to adjust topic generation weights, we generated 50 topics. Extracting the top 20 terms from each topic yielded 1,000 terms, forming a 1000$\times$1000 co-occurrence matrix (sparsity=98.16%). Using TF-IDF weighting [18] and removing low-frequency terms (sparsity=90%), we obtained 533 domain hotspot terms. The 467 removed terms (e.g., "security company," "deep-water channel") had maximum frequency 7, while remaining terms averaged frequency 64, with "traffic safety" reaching 353. This sparse matrix dimensionality reduction significantly decreases computational load with minimal information loss in large-scale text processing.

Building a co-word network from these 533 high-frequency terms reduced the topic count to 28 (Figure 6 [Figure 6: see original paper]). These hotspot terms essentially cover transportation law research hotspots from 2006-2016. Sorting by publication year reveals hotspot evolution.

4.4 Comparison Between CA-LDA and Traditional LDA Models

We compared CA-LDA and traditional LDA on the same dataset (6,230 transportation law abstracts). Results are shown in Table 1 . Both models use LDA' s bag-of-words approach, producing identical vocabularies but differing term importance rankings. CA-LDA' s high-frequency co-occurrence term "China Academic Journals" is unrelated to transportation law, primarily due to non-topic content mixing in web data.

Both models' top 50 terms center on "transportation," "traffic management," and "traffic accidents," showing consistent core content. However, differences include: (1) 18 distinct terms (shaded in Table 1); (2) significant differences in term frequency ranking; (3) traditional LDA generates more semantically singular key terms (e.g., "urban rail transit," "urban traffic," "public transport," "legal liability," "judicial interpretation"), while CA-LDA results include contextual terms like "automobile society," "low-carbon economy," and "blue economy"; legal documents such as "Road Traffic Safety Law" and "Interpretation"; controversial research hotspots like "limit scope," "discretion," "aggravated offense," "traffic accident determination document," and "personal injury compensation" ; and

management methods like "electronic police," "traffic safety education," and "bus priority."

Overall, CA-LDA provides richer research auxiliary information than traditional LDA, capturing genuine hotspot research content. For clarity, we generated term networks from both models' top 50 terms, with node size representing frequency (or weighted frequency). Figure 7 [Figure 7: see original paper] shows substantial differences: traditional LDA produces many isolated high-frequency terms, indicating generation from few documents, whereas true hotspots should appear across multiple documents. Traditional LDA also shows uneven frequency distribution, potentially emphasizing terms with low absolute but high relative frequency. CA-LDA exhibits smaller frequency differences, stronger associations, and clear term clustering, demonstrating superior topic cohesion.

5 Conclusion

This paper proposes a CA-LDA model incorporating co-word network analysis, using network topology parameters as regulating variables for topic classification to control term-to-topic assignment, with stochastic gradient descent techniques improving computational efficiency. Co-word network topology parameters modify term assignment from a vector association perspective, yielding results that reflect not only term frequency probability but also provide information through node betweenness centrality. This enables discovery of hub terms from lexical associations, reflecting critical technologies in longitudinal domain evolution and effective solutions for different problems in horizontal comparison. Applied to transportation law research hotspot analysis, the model achieves good results in processing large-scale document data and can be extended to automated processing across various fields.

Future research directions include: (1) investigating other complex network topology parameters (e.g., clustering coefficient, cliques, community) that reflect social network relationships in co-word networks and their impact on LDA topic generation; (2) improving semantic analysis by establishing domain-specific professional vocabularies, as the current compound word method (using literature keywords) is not universally applicable (e.g., for online shopping reviews); and (3) developing more scientific methods for determining the number of topics K, as perplexity-based cross-validation may produce large values that hinder document classification, requiring further processing to consolidate topics.

References

[1] Fan Yunman, Ma Jianxia. Review on the LDA-based Techniques Detection for the Field Emerging Topic [J]. New Technology of Library and Information Service, 2012(12): 58-65.

[2] Day W H E, Edelsbrunner H. Efficient Algorithms for Agglomerative Hierarchical Clustering Methods [J]. Journal of Classification, 1984, 1(1): 7-24.

[3] Cao Gaohui, Jiao Yuying, Cheng Quan. Research on Tag Cluster Based on Hierarchical Agglomerative Clustering Algorithm [J]. New Technology of Library and Information Service, 2008(4): 23-28.

[4] Katz S. Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer [J]. IEEE Transactions on Acoustics, Speech, & Signal Processing, 1987, 35(3): 400-401.

[5] Chen Langzhou, Huang Taiyi. A Novel Word Clustering Algorithm and Vari-Gram Language Model [J]. Chinese Journal of Computers, 1999, 22(9): 942-948.

[6] Salton G, Wong A, Yang C S. A Vector Space Model for Automatic Indexing [J]. Communications of the ACM, 1975, 18(11): 613-620.

[7] Pang Jianfeng, Bu Dongbo, Bai Shuo. Research and Implementation of Text Categorization System Based on VSM [J]. Application Research of Computers, 2001, 27(9): 23-26.

[8] Porteous I, Newman D, Ihler A, et al. Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation [C]. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2008: 569-577.

[9] Newman D, Asuncion A, Smyth P, et al. Distributed Inference for Latent Dirichlet Allocation [C]. In: Proceedings of the 21st Annual Conference on Neural Information Processing Systems. 2007: 1081-1088.

[10] Asuncion A U, Smyth P, Welling M. Asynchronous Distributed Learning of Topic Models [C]. In: Proceedings of the 22nd Annual Conference on Neural Information Processing Systems. 2008: 81-88.

[11] Blei D M, Lafferty J D. A Correlated Topic Model of Science [J]. The Annals of Applied Statistics, 2007, 1(1): 17-35.

[12] Sato I, Nakagawa H. Topic Models with Power-law Using Pitman-Yor Process [C]. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2010: 673-682.

[13] Teh Y W. Dirichlet Process [A]. //Sammut C, Webb G I. Encyclopedia of Machine Learning [M]. Springer US, 2011: 280-287.

[14] Callon M, Courtial J P, Turner W, et al. From Translations to Problematic Networks: An Introduction to Co-word Analysis [J]. Social Science Information, 1983, 22(2): 191-235.

[15] Callon M, Courtial J P, Laville F. Co-word Analysis as a Tool for Describing the Network of Interactions Between Basic and Technological Research: The Case of Polymer Chemsitry [J]. Scientometrics, 1991, 22(1): 155-205.

[16] Coulter N, Monarch I, Konda S. Software Engineering as Seen Through Its Research Literature: A Study in Co-word Analysis [J]. Journal of the American Society for Information Science, 1998, 49(13): 1206-1223.

[17] Zhang Xiaodong, Zhou Hongli, Hu Yang, et al. Research Hotspots of Computer Integrated Manufacturing of China Based on Co-word Analysis and Social Network Analysis [J]. Science and Technology Management Research, 2016(11): 145-149.

[18] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[19] Newman D, Bonilla E V, Buntine W. Improving Topic Coherence with Regularized Topic Models [C]. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. 2011: 496-504.

[20] Jordan M I, Ghahramani Z, Jaakkola T S, et al. An Introduction to Variational Methods for Graphical Models [J]. Machine Learning, 1999, 37(2): 183-233.

[21] Hoffman M, Blei D, Wang C, et al. Stochastic Variational Inference [J]. Journal of Machine Learning Research, 2013, 14(1): 1303-1347.

[22] Brandes U. A Faster Algorithm for Betweenness Centrality [J]. Journal of Mathematical Sociology, 2001, 25(2): 163-177.

[23] Newman M E J. The Structure and Function of Complex Networks [J]. SIAM Review, 2003, 45(2): 167-256.

Author Contributions: Ma Hong: conceptualized research, designed study, analyzed conclusions; Cai Yongming: data acquisition, cleaning and preprocessing, algorithm design, program development.

Conflict of Interest Statement: All authors declare no conflict of interest.

Supporting Data: Available at journal website http://www.infotech.ac.cn. [1] Ma Hong, Cai Yongming. CNKI Retrieval Raw Data.rar. CNKI retrieval raw data. [2] Ma Hong, Cai Yongming. Compound Word Library and Stopword Library.rar. Compound word library and stopword library. [3] Ma Hong, Cai Yongming. Descriptive Statistical Analysis Data.xlsx. Descriptive statistical analysis data. [4] Ma Hong, Cai Yongming. Preprocessed Data.rar. Preprocessed data.

Received: August 1, 2016 **Revised:** November 2, 2016

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