Accident Analysis and Prevention 118 (2018) 131–145

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Visualization and analysis of mapping knowledge domain of road safety studies

Xin Zou[a,*], Wen Long Yue[b], Hai Le Vu[c]

[a] School of Natural and Built Environments, University of South Australia, Mawson Lakes, SA, 5095, Australia
[b] Australian Road Research Board, Port Melbourne, VIC, 3207, Australia
[c] Institute of Transport Studies, Monash University, Clayton, VIC, 3800, Australia

ARTICLE INFO

Keywords:
Road safety
Bibliometrics
Mapping knowledge domain
Visualization
VOSviewer
Sci2 Tool

ABSTRACT

Mapping knowledge domain (MKD) is an important application of visualization technology in Bibliometrics, which has been extensively applied in psychology, medicine, and information science. In this paper we conduct a systematic analysis of the development trend on road safety studies based on the Science Citation Index Expanded (SCIE) and Social Sciences Citation Index (SSCI) articles published between 2000 and 2018 using the MKD software tools VOSviewer and Sci2 Tool. Based on our analysis, we first present the annual numbers of articles, origin countries, main research organizations and groups as well as the source journals on road safety studies. We then report the collaborations among the main research organizations and groups using co-authorship analysis. Furthermore, we adopt the document co-citation analysis, keywords co-occurrence analysis, and burst detection analysis to visually explore the knowledge bases, topic distribution, research fronts and research trends on road safety studies. The proposed approach based on the visualized analysis of MKD can be used to establish a reference information and research basis for the application and development of methods in the domain of road safety studies. In particular, our results show that the knowledge bases (classical documents) of road safety studies in the last two decades have focused on five major areas of "Crash Frequency Data Analysis", "Driver Behavior Questionnaire", "Safety in Numbers for Walkers and Bicyclists", "Road Traffic Injury and Prevention", and "Driving Speed and Road Crashes". Among the research topics, the five dominant clusters are "Causation and Injury Severity Analysis of Road Accidents", "Epidemiologic Study and Prevention of Road Traffic Injury", "Intelligent Transportation System and Active Safety", "Young drivers' driving behavior and psychology", and "Older drivers' psychological and physiological characteristics". Finally, the burst keywords in research trends include Cycling, Intelligent Transportation Systems, and Distraction.

1. Introduction

Road crashes result in about 1.3 million fatalities and 50 million severely injuries worldwide every year. The number of road deaths is on the rise despite the continuing improvements in road safety (International Transport Forum, 2017). Road accidents put significant financial strain on families and the society. Many families fell into poverty due to the burden of long-term medical expenses, the loss of incomes, or the need to care for the disabled relative (Mohan et al., 2006). To this end, road safety research plays a significant role in the improvement of the socio-economic by reducing the occurrences and severities of road accidents.

Through a periodical review of research in road safety, it can be beneficial to understand the research status and identify gaps. However, the majority of literature reviews in this domain have been based on the summary of literatures with, and require a long-term accumulation, summarization and extraction of research activities in the domain. The existing reviewing approach thus often does not give the entire picture of road safety research which motivates the development of a new method in this paper. Using high-performance computers, more complete databases and information visualization technology, the mapping knowledge domain (MKD) methods provide a new way to conduct such a literature survey. In recent years, the MKD methods have been widely applied in exploring disciplinary development status, research frontiers, research hotspots, and systematic reviews (Zhu and Hua, 2017; van Nunen et al., 2017; Vega-Almeida et al., 2018; de la Hoz-Correa et al., 2018; Gaede and Rowlands, 2018). This paper proposes the use of MKD and employs bibliometric methods to proceed quantitative analyses, explores the research status in road safety research to objectively reveal the research development status, identify the research fronts and

* Corresponding author.
E-mail address: zouxy011@mymail.unisa.edu.au (X. Zou).

https://doi.org/10.1016/j.aap.2018.06.010
Received 19 March 2018; Received in revised form 12 June 2018; Accepted 15 June 2018

hotspots in the domain of road safety studies.

2. Methodology

Bibliometric analysis is a kind of document analysis methods which takes the advantage of bibliometric theory to analyze pertinent literatures through the use of mathematical and statistical approaches. It studies the distribution profiles, quantitative relations and clusters of the literatures. Within bibliometric analysis, MKD is a method to visually present the knowledge on a certain field of subject by means of data mining, information analysis, scientific measurement and graphic plotting. It has the function of knowledge navigation and belongs to the category of Scientometrics (Shiffrin and Börner, 2004). Bibliometric methods and MKD, with the advantages of comprehensive quantitative statistics, visual information display, objective description and evaluation, have become an important tool for global analysis and investigation in various scientific areas.

2.1. Data source and research process

The SCIE and SSCI citation index database in the Web of Science (WOS) Core Collection were retrieved as the source for this study. The retrieval topic was “Road Safety”, the timespan was “from 2000 to 2018”, and the document type was “Article”. A total of 9835 pertinent publications were collected, and the last update of the data was on May 9, 2018. The retrieved results were saved as a “Plain Text” with “Full Record and Cited References”. Note that only the “road safety” keyword was used in this study and the number might change with additional keywords added to the search. Nevertheless, the methodology proposed herein remains unchanged and the findings in this paper are applicable for multiple keywords search.

2.2. Analytical tool and method

The main method in this study is based on Scientometrics, and the analysis software used include VOSviewer and Sci2 (Science of Science) Tool. VOSviewer, developed by van Eck and Waltman (Leiden University) in Netherlands, is a literature knowledge unit visualization software based on Visualization of Similarities (VOS) technology, which has unique advantages in the display of mapping knowledge domains, especially in the aspect of clustering (van Eck and Waltman, 2010). The layer label display technology of VOSviewer enables the node of dense network to be clearly displayed through interactions, therefore, it is especially suitable for analyzing large-scale data and constructing complex networks. In the latest VOSviewer version 1.6.6, the co-occurrence network, citation network and coupling network can be constructed using file formats such as. net,. mat and. txt. The Sci2 Tool, developed by the research team of Börner (Indiana University) and Boyack (SciTech Strategies) in USA, is a modular toolset specifically designed for the study of science (Sci2 Team, 2009). The Sci2 Tool, embedded with a variety of database functions, can load data sets in the different formats to conduct fundamental analysis on Scientometrics, such as co-occurrence analysis, citation analysis, coupling analysis and burst detection analysis. The maps of knowledge domains can also be created by a variety of visualization algorithms built in the Sci2 Tool.

Mapping the knowledge domain is a kind of image which shows the development process and structural relationship on the scientific knowledge. It has the dual nature and characteristics of “graph” and “genealogy”, which means it is not only a visual knowledge graph but also a serialized knowledge genealogy, showing many implied complex relationships including network, structure, interaction, intersection, evolution, or derivative among knowledge units or knowledge clusters. Understanding these complex knowledge relationships can thus produce a new knowledge. The drawing (or creating) of mapping knowledge domains includes co-citation analysis, co-occurrence analysis and burst detection analysis as explained below.

- ***Document co-citation analysis***: In Scientometrics, citing documents form the research frontier in a certain field, and the cited documents form a knowledge base in that field. The document co-citation analysis is based on the statistics of the number of two documents being cited by one or more documents at the same time, so as to conduct the network analysis and cluster analysis for the cited documents and thus analyze the knowledge base of the certain subject represented by these documents.
- ***Keywords co-occurrence analysis***: The keywords in academic publications are the natural language words that express the thematic concepts of documents. The keywords condense authors’ academic viewpoints, making it an important indicator in bibliometrics. Keywords co-occurrence analysis is based on the statistics of the number of a pair of keywords being cited in the same document, so as to conduct the network analysis and cluster analysis for these words and thus reveal the knowledge structure and research frontier of a certain subject.
- ***Burst detection analysis***: Burst detection analysis takes into account the change of keyword frequencies and identifies the keywords with burst growth characteristics in a certain time period on a certain research sector, which can be used to study the development trend of a certain topic. It is different from the burst detection analysis based on threshold values, although each keyword may be used relatively less frequently, the burst keywords can be found according to the change of keyword frequencies over time, and thus the latest research trend can be predicted through such keywords.

2.3. Construction of co-word similar matrix

The construction of the keywords co-occurrence matrix is the basis of conducting cluster analysis. Counting the number of times of any two keywords appearing in the same documents, and n keywords can build a $n \times n$ co-occurrence matrix, which defines the similarity matrix S:

$$S = (s_{ij}) \tag{1}$$

where $s_{ij} \geq 0$ is a similarity measure defined below and $s_{ij} = s_{ji}$, $i, j \in \{1, 2, \cdots, n\}$

The idea of using the association strength in VOS to express the relationship between two objects is to cluster the objects with high similarity as much as possible, while the objects with low similarity should be separated as far as possible (van Eck and Waltman, 2007). VOSviewer uses association strength method to calculate the similarity s_{ij} between the objects i and j in a map as:

$$s_{ij} = \frac{c_{ij}}{W_i W_j} \tag{2}$$

where s_{ij} denotes the similarity between the objects i and j,

c_{ij} denotes the co-occurrence times of the objects i and j,

W_i and W_j denote the occurrence times of the objects i and j respectively.

Note that van Eck and Waltman (2009) specially discussed the advantages of this measure comparing to other similarity measures (such as the cosine, the inclusion index, and the Jaccard index).

2.4. Construction of knowledge domain maps

In order to make the clustering effect more obvious, the main idea of VOS is to minimize the sum of the weighted Euclidean distances of all the objects in each cluster, where the distance of each cluster can be determined with following equation:

$$E(X; S) = \sum_{i<j} s_{ij} \| x_i - x_j \|^2 \tag{3}$$

where $| \cdot |$ denotes the Euclidean norm. In order to avoid all similar objects appearing in the same position, the following constraint condition is used: