Agriculture, Ecosystems and Environment 267 (2018) 1–14

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

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

Topic modelling of ecology, environment and poverty nexus: An integrated framework

Xin Cheng[a,b,d], Chuanmin Shuai[a,c,d,*], Jiali Liu[e], Jing Wang[a,d], Yue Liu[f], Wenjing Li[a,d,g], Jing Shuai[c,d]

[a] Department of Management Science and Engineering, School of Economics and Management, China University of Geosciences, Wuhan, 430074, China
[b] Department of Environmental Sciences, Faculty of Science and Engineering, Macquarie University, North Ryde, Sydney, NSW, 2109, Australia
[c] Department of Applied Economics, School of Economics and Management, China University of Geosciences, Wuhan, 430074, China
[d] Research Centre of Resource and Environmental Economics, China University of Geosciences, Wuhan, 430074, China
[e] Senior Network Security Researcher, Wuhan Greenet Information Service Co., Ltd., Wuhan, 430074, China
[f] Business School, Jianghan University, Wuhan, 430056, China
[g] School of Natural and Environmental Sciences, Newcastle University, NE1 7RU, Newcastle Upon Tyne, United Kingdom

ARTICLE INFO

Keywords:
Integrated framework
Topic modelling
Ecology
Environment
Poverty
Research concerns

ABSTRACT

There is little research on the ecology, environment and poverty nexus. To understand the major research concerns and clusters of ecology, environment and poverty (EEP), we have built an integrated framework of their nexus to pursue sustainable development. We have collected 4335 English language publications using the search terms "Ecology and Environment and poverty" by type "Topic" in Web of Science, covering publication years from 1981 to 2017. A 9-topic model, 35 references with the strongest citation bursts and the top-30 citation clusters are used to describe the integrated framework. We introduce topic modelling into the EEP field using scientific articles, combined with bibliometric analysis and visualisation approaches, and finally propose an integrated ecology-environment-poverty model. The greatest concerns are biodiversity and environmental services. Six main clusters related to poverty are ecosystem, health risk, economic and environmental development, natural resources and food production, children, women and inequality, and urban poverty. Therefore, ecological and environmental degradation and poverty are linked, and must be tackled together. The win-win strategies on local and national levels that both restore the environment, enhance incomes and ensure a sustainable livelihood are needed. The joint approaches of topic modelling and bibliometrics are expected to be further applied in ecological and environmental economics and management fields. Our findings provide a theoretical basis for further research and decision making and contribute to sustainable development.

1. Introduction

Since the 1980s, the amount of literature on environmental degradation, poverty and their interrelationships has grown rapidly. After the "United Nations (UN) World Summit on Sustainable Development" in 1992 and the "UN Conference on Environment and Development" in 2002, poverty alleviation and environmental protection have become hot topics throughout the world (Gray and Moseley, 2005; Sachs and Reid, 2006). According to the "Millennium Development Goals (MDG) Report 2015", we are now forging a bold vision for sustainable development (United Nations, 2015a). "No Poverty" remains the first of the United Nations sustainable development goals (United Nations, 2015b) and the human and natural worlds also need integration (Costanza et al., 2016).

Sustainable development, defined by environmentally sound economic growth, is a practical necessity (Cao et al., 2010a, 2017b). However, some conservation and development projects lead to negative impacts on impoverished people (Cao et al., 2010b). As important components of sustainable development (World Bank, 1992), poverty alleviation and environmental protection assume priority especially for ecologically vulnerable zones and poverty-stricken regions in developing countries (Dasgupta et al., 2005). Therefore, conservation strategies must be able to achieve both ecological and social progress without detracting from their primary economic objectives (Cao, 2011).

Analysing current research would help find major concerns in the field and the limitations of current research, and be important for

* Corresponding author at: Department of Management Science and Engineering, School of Economics and Management, China University of Geosciences, Wuhan, 430074, China.
E-mail addresses: xin.cheng2@students.mq.edu.au, cxjournal@163.com (X. Cheng), shuaicm@cug.edu.cn, shuaicharm@qq.com (C. Shuai).

https://doi.org/10.1016/j.agee.2018.07.022
Received 20 March 2018; Received in revised form 25 July 2018; Accepted 27 July 2018

guiding further research. Revealing the complex relationships between ecology, environment (EE) and poverty would also be important for policy-makers making systematic strategies for impoverished regions, which would contribute to the MDGs. Early conceptual studies on relationships between EE and poverty focussed largely on (1) theoretical analysis and evaluation models (Dasgupta et al., 2005), (2) quantitative analysis of their interactions (Ahmed et al., 2011), and (3) analysis of their coordinated development of EE and poverty (Pagiola et al., 2005; Sims, 2010; Shen et al., 2015; Suich et al., 2015).

Global environmental problems are often evaluated independently by ecologists and social scientists (Cao, 2010; Liu, 2010; Cao et al., 2014). Beyond large-scale correlative studies mapping global patterns of poverty (wealth) and environmental change, little attention has been paid to understanding systematic relationships among ecology, environment and poverty based on research publications. Arguments remain about complex relationships between poverty and environmental degradation. In the ecology, environment and poverty (EEP) field, few studies analyse research concerns based on text mining. A large number of studies have used thematic reviews and manual content analyses to perform their work. Compare with these conventional approaches, the text mining techniques would be more time-saving and efficient, and less subjective when processing vast amounts of data (Jiang et al., 2016).

This study considers a more efficient and streamlined way to synthesise the interrelationships among ecology, environment and poverty. The scientific literature is a valuable and rich source of knowledge (Jiang et al., 2016). Scientific knowledge changes frequently, and while most of these changes are incremental, some are revolutionary and fundamental (Chen, 2013). To improve the efficiency of processing scientific documents, some scholars (Nichols, 2014) have introduced text mining approaches, such as topic modelling, into Scientometric and bibliometric studies to help outline the integrated framework of a discipline (Yang et al., 2016). Our study uses a mixed topic modelling and bibliometric methodology, with quantitative and qualitative data analyses. We review the scientific literature regarding the links and mechanisms among ecology, environment and poverty. Specific questions we address are: What are the major academic topics in the EEP field during the last 30 years? What are the interrelationships among ecology, environment and poverty? Therefore, the purpose of this investigation is to find the EEP research concerns and overview interrelationships among ecology, environment and poverty. This study makes a major contribution to studies on EEP by demonstrating and improving knowledge about the nature of their linkages, and focusses on the latent integrated framework and future research agenda.

This paper has been organised in the following way (Fig. 1): the introduction is followed by the second part, which is concerned with the Topic Modelling and Bibliometric methods employed for this study. Part three describes data retrieval, and looks at the detailed datasets that would be used in each method. The fourth part presents the results of the research. Part five is the in-depth analysis, discussing the ecology-environment-poverty nexus. The last part is the conclusions.

2. Methods and data

Studies about EEP began in the 1980s, and empirical studies with case studies, qualitative analyses, including literature reviews, have been used to investigate relationships between ecology and poverty or environment and poverty. Since research on EEP has a four-decade history, it is possible to examine the literature to identify the integrated framework that has been formed but has not yet been revealed, using topic modelling and bibliometric approaches.

2.1. Topic modelling

New approaches have been developed to understand latent intelligence with the growth of unstructured data sources, such as textual data sources (George et al., 2014). Among these approaches, topic modelling is a powerful text mining method that is able to uncover the latent integrated framework in textual data (Griffiths and Steyvers, 2004), and subsequently provide significant support for researchers and practitioners in the broad field of decision-making (Jiang et al., 2016).

Furthermore, with recent advances in computing power, scientific indices and bibliographic techniques, progress is being made and researchers are gradually exploring hidden connections and knowledge domains in the literature (Lee et al., 2016). Intellectual relationships and collaboration networks are fundamental to a knowledge domain (Hu and Racherla, 2008). The visual representation of such "knowledge networks" contributes to the overall understanding of intellectual collaborations in a particular knowledge domain (Lee et al., 2016).

Latent Semantic Analysis (LSA) is a method for automatic indexing and retrieval (Deerwester et al., 1990), which was developed to resolve the so-called vocabulary mismatch problem (Landauer et al., 1998). Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus, where documents are represented as random mixtures over latent topics, and where each topic is characterised by a distribution over words (Blei et al., 2003). The probabilistic LSA (pLSA) (Hofmann, 1999), and LDA (Blei et al., 2003) could be represented as probabilistic graphic models (Fig. A1).

LDA forms topics from words that co-occur frequently, whereas LSA models, such as LSA with Singular Value Decomposition (SVD) and LSA with Non-negative Matrix Factorisation (NMF), have no such preferences and often create low-quality topics from completely unrelated words (Stevens et al., 2012). LDA has also been shown be useful for extracting topics from a collection of academic abstracts or articles in previous studies (Blei, 2012; Steyvers and Griffiths, 2014; Jiang et al., 2016). For applications in which a human end-user will interact with learned topics, the flexibility of LDA and the coherence advantages of LDA warrant strong consideration (Stevens et al., 2012). Therefore, we employ LDA to discover the main themes of eco-environmental research articles.

In the LDA representation, the parameters α and β are corpus-level parameters, assumed to be sampled once while generating a corpus. θ is a document-level variable, sampled once per document. The variables z and w are word-level variables and are sampled once for each word in each document (Table 1).

According to the generative process, the probability that a word $w_{m,n}$ instantiates a particular term t, is obtained by marginalising the latent variable$z_{m,n}$ and omitting the hyper parameters as follows:

$$p(w_{m,n} = t \mid \theta_m, \Phi) = \sum_{k=1}^{K} p(w_{m,n} = t|\varphi_k) p(z_{m,n} = k|\theta_m) \tag{1}$$

Based on the Bayesian networks of LDA, the complete-data likelihood of a document can be specified using a joint distribution of all known and hidden variables, given the hyper parameters (α and β), as follows:

$$p(w_m, z_m, \theta_m, \Phi|\alpha, \beta) = \overbrace{\underbrace{\prod_{n=1}^{N_m} p(w_{m,n}|\varphi_{z_{m,n}}) p(z_{m,n}|\theta_m)}_{\text{Word plate}} \cdot p(\theta_m|\alpha)}^{\text{Document plate}} \cdot \underbrace{p(\Phi|\beta)}_{\text{Topic plate}} \tag{2}$$

$$p(w_m|\alpha, \beta) = \iint p(\theta_m|\alpha) \cdot p(\Phi|\beta) \cdot \prod_{n=1}^{N_m} p(w_{m,n}|\theta_m, \Phi) d\Phi d\theta_m \tag{3}$$

$$p(W|\alpha, \beta) = \prod_{m=1}^{M} p(w_m|\alpha, \beta) \tag{4}$$

LDAvis (Sievert and Shirley, 2014) is a web-based interactive visualisation of topics estimated using LDA. It could provide a global view