Using LDA Model to Quantify and Visualize Textual Financial Stability Report

Guowen Li a,b, Xiaoqian Zhu a, Jun Wang a,b, Dengsheng Wu a,b, Jianping Li a,b,*

* Corresponding author. Tel.: 010-59358805; fax: 010-59358608.
E-mail address: ljp@casipm.ac.cn

Abstract

The financial system plays a crucial role in development of countries, which makes its stability be heated around the world. However, traditional indices which help measure financial stability such as quantile, leverage ratio and liquidity have instinctive shortcomings. For example, these digital indices are usually unavailable and one-sided. Therefore, finding a new approach to quantifying and visualizing financial stability is necessary and desirable. Different from digital data, textual data is more available and usually implies more abundant information and intuitional senses. Since textual data is not visualized, it is vital to find out what texts talk about. Latent Dirichlet Allocation (LDA) is one of the most effective approaches to achieve above goal. In order to apply LDA to measure financial stability, China Financial Stability Report is selected to make an empirical analysis. The results are as follows. Firstly, it is reasonable that LDA model can be applied to analyze China Financial Stability Report. Secondly, dividing core terms of every topic into basic terms and particular terms, we can draw pictures of every embranchment in finance. And we can analyze topics rank of 5 years or in every single year, so that a designing matrix comes into being and we can study financial stability tendencies. At last, the macro-environment in finance can be depicted easily using word cloud.

1. Introduction

Latent Dirichlet Allocation is a classic type of topic model proposed by Blei et al. in 2003 [1]. LDA introduces topics as latent variables to help make documents low the dimension and reveal underlying meanings. Just like Probability Latent Semantic Analysis (PLSA), an original version of LDA, the model is also an unsupervised
Learning method. As a matter of fact, it is usually regarded as a Bayesian version of PLSA. With the help of priori Dirichlet distribution, LDA can deal with extensive document collections, after which two posteriori Dirichlet distributions (one presents the topics distribution under documents and the other presents terms distribution under topics) come into being. The model evolved kinds of mentionable variants. [2,3,4,5] break the hypothesis that both documents and terms are exchangeable; Sammut, C. and G.I. Webb [6] developed a different kind of Dirichlet Process so that LDA could utilize a nonparametric Bayesian method to carry out; Yang Bao et al. [7] proposed a sent-LDA model and used it to discover and quantify risk types simultaneously. As for its application, LDA performs well in sentiment analysis field [8, 9, 10], topics clustering [11, 12] and sequential text mining [13] et al.

As financial system is going to play a big role, the financial stability issue becomes heated around the world [14]. But only using digital indices such as quantile is not all-sided to depict it [15]. To entirely reveal the situation of the financial system, People’s Bank of China issues China Financial Stability Report every year since 2005, in which there are plenty of figures and graph to discuss relevant events from opinions of Chinese perspective. These reports usually cover eight chapters and some special topics, with content from the international financial environment to every single industry in the finance field. China Financial Stability Report counts for stakeholders who concern about financial issues, but it also strikes their pain spot because of its mass of textual content. Under most circumstances, people just want an outline of reports where LDA model could play a role.

This paper uses LDA model to make a topics analysis for these reports. And the result could give readers an overall but specific view. Furthermore, LDA can cluster reports into different parts, which could present a comprehensible stem of reports and be used to define unknown documents. After that, sentiment analysis or other processes would be available.

The rest of this paper is organized as follows. In section 2, a model of LDA is briefly introduced. In section 3, using China Financial Stability Report data from 2013 to 2017, empirical analysis is conducted. At last, conclusions are given in section 4.

2. Methodology

Maybe we could regard the generation of a document as a process of dicing by authors. In other words, the author first chose k topics with probability, and then selected n terms among every topic out with probability to produce a document. From the opposite perspective, if the document already exists, how to infer topics and terms contained by every single topic? LDA proposed by Blei et al. in 2003 gives us answers. The model is the most common topic model applied in text mining currently. When it is applied to a specific problem, what topics these documents cover and what these topics contain would demonstrate using numbers of document collections.

To be more specific, let N, M, K, and V be the number of terms in a document, the number of documents in a corpus, the number of topics, and the vocabulary size, respectively. The notations Dirichlet(·) and Multinomial(·) represent Dirichlet and multinomial distribution with parameter (·), respectively. The notation βk is the V-dimensional word distribution for topic k, and θd is the K-dimensional topic proportion for document d. The notations η and α represent the hyper-parameters of the corresponding Dirichlet distributions. The graphical representation of LDA is shown in Figure 1, and the corresponding generative process is as follows:

1. For each topic k ∈ [2],
   (a) draw a distribution over vocabulary words βk ~ Dirichlet(η).
2. For each document d,
   (a) draw a vector of topic proportions θd ~ Dirichlet(α).
   (b) For each word ωd,n in document d,
      (i) draw a topic assignment zd,n ~ Multinomial(θd);
      (ii) draw a word ωd,n~ Multinomial(βzd,n).
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات