ITCS 5156 - [Project] Progress 1 Project Topic : Topic Modeling using Latent Dirichlet Allocation (LDA) model in NLP

By : Pranjali Mehta (801255574)

Contents

- Summary of Paper 1: A Comparative Study of NLP Topic Modeling Methods and Tools
- Summary of Paper 2: Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis
- Summary of Paper 3: Topic Modeling in Natural Language Processing

Summary : A Comparative Study of NLP Topic Modeling Methods and Tools

The chosen primary research paper presents two categories of Topic modeling - Methods and Tools. The first category discusses three methods of topic modeling- LSA, pLSA and LDA. It specified the characteristics and theoretical backgrounds of these methods as well as its limitations and advantages over one another. The paper provides a high level view of each of these methods without going into specific details. LDA is the most widely used method of topic modeling. A number of methods, including both probabilistic and non probabilistic, were introduced before LDA for the sake of topic modeling and information gathering. Probabilistic models like pLSA and LDA improved on the previous non-probabilistic models. The second category provides a brief view of tools that can enable the use of topic modeling in various applications.

Summary : Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis

This paper determined the most important elements of the MOOC learning experience by combining LDA topic models and qualitative analysis to analyze post-course open-ended survey responses from learners. Important technical factors included comprehensive course information, language-related support, platform usability, accessibility of learning materials, support for slow Internet speeds in developing nations, and certificate cost.

The approach employed in this work may be replicated for exploratory analysis on big textual data and is applicable to comparable tasks in educational research, such as the examination of open-ended comments or discussion forums. It's also crucial to keep in mind that while the LDA topic model makes an excellent beginning point for exploratory data analysis, qualitative analysis must be used to analyze the results in order to draw meaningful conclusions.

Summary : Topic Modeling in Natural Language Processing

In this paper, I studied about how they have used the word embedding model for analysis of different topics present in new headlines which is capable of capturing syntactic similarity and semantic relation, the context of a word in a document, relation with other words. This particular paper covers a detailed flow of what all steps are incorporated while performing Topic modeling on the documents. It has a detailed description of different techniques which can be employed to perform topic modeling such as : Word embeddings, word2vector, continuous bag of words, skip-gram-model and K-means clustering. I plan to use these techniques on my dataset before applying LDA model.