Abstract

Text mining is one of the most significant fields in the digital era due to the rapid growth of textual information. Topic models are gaining popularity in the last few years. A topic comprises of a group of words that are often take place together. Topic models are better performing techniques to extract semantic knowledge presented in the data. The various methods used for topic models are, LSA (Latent Semantic Analysis), PLSA (Probabilistic Latent Semantic Analysis), LDA (Latent Dirichlet Allocation). These methods gained popularity in extracting hidden themes from the document (corpus). Various topic modeling algorithms are developed to inquiry, summarize and extract hidden semantic structures of large corpus. In this paper, we present a detailed survey covering the various topic modeling techniques proposed in last decade. Additionally, we focus on different strategies of extracting the topics in social media text, where the goal is to find and aggregate the topic within short texts. Further, we summarize the various applications and quantitative evaluation of the various methods, with statistical and mathematical knowledge to predict the convergence of results.
References

allocation and the Dirichlet multinomial mixture model on short text. In Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016 (pp. 1-6). IEEE.


A Detailed Survey on Topic Modeling for Document and Short Text Data


51. Zhao, H., Du, L. and Buntine, W., 2017, November. A word embeddings informed
focused topic model. In Asian Conference on Machine Learning (pp. 423-438).


Index Terms

Computer Science

Information Sciences

Keywords
Text mining, Topic Modeling, Short Text, Semantic Analysis.