

Topic Grouping by Spectral Clustering

Young-Seob JEONG, Won-Jo LEE, Ho-Jin CHOI

*Department of Computer Science,
KAIST(Korea Advanced Institute of Science and Technology),
291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
{pinode, mochagold, hojinc}@kaist.ac.kr*

Abstract—With the growing number of web documents, it becomes difficult to analyze and obtain information from such an array of documents. Furthermore, unsupervised methods are preferable, as most web documents are unlabeled. Probabilistic topic modeling is one such method. It discovers latent structures among unstructured documents. While many traditional topic models usually assume that the topics are independent of each other, some models have been proposed to obtain correlations between the topics or a hierarchy of the topics. These models are designed to obtain both the topics and the correlations without using any other method. Therefore, very few studies apply other methods to determine a correlation between topics. In this paper, we apply spectral clustering to group the topics obtained from a traditional topic model, in this case the Latent Dirichlet Allocation model. To the best of our knowledge, this is the first approach that uses spectral clustering for the grouping of topics. We demonstrate the experimental results with various settings.

Keyword—Topic model, Spectral clustering, Topic grouping



Young-Seob Jeong is currently a PhD student in the Dept. of Computer Science at KAIST. His current research interests include topic modeling, deep learning, and action prediction based on various sensor data.



Won-Ji Lee received a bachelor's degree of computer science in 2012 from Hanyang University. He is currently a MS candidate student in the department of computer science at Korea Advanced Institute of Science and Technology.



Ho-Jin Choi is currently an associate professor in the Dept. of Computer Science at KAIST. In 1982, he received a BS in Computer Engineering from Seoul National University, Korea, in 1985, an MSc in Computing Software and Systems Design from Newcastle University, UK, and in 1995, a PhD in Artificial Intelligence from Imperial College, London, UK. From 1982 to 1989, he worked for DACOM, Korea, and between 1995 and 1996, worked as a post-doctoral researcher at Imperial College. From 1997 to 2002, he served as a faculty member at Korea Aerospace University, Korea, then from 2002 to 2009 at Information and Communications University (ICU), Korea, and since 2009 he has been with the Dept. of Computer Science at KAIST. Between 2002 and 2003, he visited Carnegie Mellon University (CMU), Pittsburgh, USA, and has been serving as an adjunct professor of CMU for the program of Master of Software Engineering (MSE). Between 2006 and 2008, he served as the Director of Institute for IT Gifted Youth at ICU. Since 2010, he has been participating in the Systems Biomedical Informatics National Core Research Center at the Medical School of Seoul National University. Currently, he serves as a member of the boards of directors for the Software Engineering Society of Korea, for the Computational

Intelligence Society of Korea, and for Korean Society of Medical Informatics. His current research interests include artificial intelligence, data mining, software engineering, and biomedical informatics .