



# An Effective Clustering Approach for Mining Text Data Using Side Information

Monica.M<sup>1</sup>, Ganesh.J<sup>2</sup>

PG Scholar, Department of IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India<sup>1</sup>

Assistant Professor, Department of IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India<sup>2</sup>

**Abstract:** Side-information, in text mining, is available with the text documents in many text mining applications which may be of different kinds, such as document origin information, user-access behavior from web logs, the links in the document, or other non-textual attributes which are embedded into the text document. These could contain a huge amount of information for clustering. However, the estimation of the importance of this side-information may be difficult because some of the information may contain noisy data. In such cases, it can be risky to incorporate side-information into the mining process, because it can either improve the quality of the mining process, or can add noise to the process. Therefore, we need a principled way to perform the mining process, so as to maximize the advantages of using this side information. In this paper, we design an algorithm which combines classical partitioning algorithms with probabilistic models in order to create an effective clustering approach. We then show how to extend the approach to the classification problem. We also present experimental results on a number of real data sets in order to illustrate the advantages of using such an approach.

**Keywords:** Text Mining, Clustering, Side-information.

