Abstract

Text Mining is the discovery of valuable, yet hidden, information from the text document. Text classification (Also called Text Categorization) is one of the important research issues in the field of text mining. With the dramatic increase in the amount of content available in digital forms gives rise to a problem to manage this online textual data. As a result, it has become a
necessary to classify/categorize large texts (documents) into specific classes. Text Classification assigns a text document to one of a set of predefined classes. This paper covers different text classification techniques and also includes Classifier Architecture and Text Classification Applications.

References

- Jiawei Han, Michelin Kamber, 2001, “Data Mining Concepts and Techniques”, Morgan Kaufmann publishers, USA, 70-181.
- Wenmin Li, Jiawei Han and Jian Pei, 2001, “CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules”, IEEE International Conference on Data Mining - ICDM, pp. 369-376, DOI: 10.1109/ICDM.2001.989541.
Recent Trends in Text Classification Techniques

- Xiaoxin Yin, Jiawei Han. CPAR, 2003, “Classification based on Predictive Association Rules”, in Proceedings of SDM, doi=10.1.1.12.7268.
- Fernando Berzal, Juan-Carlos Cubero, Nicolás Marín, Daniel Sánchez, Jose-María Serrano, Amparo Vila, “Association rule evaluation for classification purposes”.
- Dat Huynh, Dat Tran, Wanli Ma, Dharmendra Sharma, 2011, “A New Term Ranking Method Based on Relation Extraction and Graph Model for Text Classification”, Faculty of Information Sciences and Engineering, University of Canberra ACT 2601, Australia.

Index Terms

Computer Science
Data Mining

Keywords
<table>
<thead>
<tr>
<th>KNN</th>
<th>Naïve Bayes</th>
<th>Support Vector Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Term Graph Model</td>
</tr>
</tbody>
</table>

- Association Based Classification
- Decision Tree Induction
- Centroid based classification
- Classification using neural network