CHAPTER 7

RESULTS AND DISCUSSIONS

7.1. INTRODUCTION

The performance analysis of document clustering results of tests is shown in below figures, in which the partition based approach and frequent pattern mining approach is compared to the different scenarios such as k-means, EM, sGEM, Apriori algorithm, FP-Growth and FP-Bonsai Tree.

7.2. FINAL COMPARISON RESULT: F-MEASURE

![Comparison of f-measure](image_url)

Figure 7.1 Final comparison result: Comparison of f-measure
• Figure 7.1 shows the comparison of f-measure performance of k-means, EM, sGEM, Apriori, FP-Growth and FP-Bonsai in Reuters-21578, 20 NG, and TDT-2 datasets.

• sGEM clustering f-measure is 98.90%, which is more than all other techniques. The relative clustering formed by 20 NG is 98.20% over TDT-2 moreover, the f-measures clearly proves that sGEM approach outperforms all other f-measures.

• The main objective result of the proposed method in 20 NG dataset has higher f-measure than the existing methodology.

7.3 FINAL COMPARISON RESULT: PURITY

![Comparison of Purity Measure](image)

Figure 7.2 Final comparison result: Comparison of Purity Measure

• Figure 7.2 shows the comparison of purity performance of k-means, EM, sGEM, Apriori, FP-Growth and FP-Bonsai in Reuters-21578, 20 NG, and TDT-2 datasets. FP-Bonsai clustering purity is 96.20%, which is more than all other techniques.
The relative clustering formed by 20 NG is 87.5% over TDT-2 moreover, the purity measures clearly prove that the FP-Bonsai approach outperforms all other purity measures.

Document clustering purity performance of the TDT-2 dataset under investigation for the topic category and the improvement in document clustering purity is more noticeable than in the other document categories.

The main objective result of the proposed method in Reuters-21578 dataset has higher purity than the existing methodology.

7.4. FINAL COMPARISON RESULT: EXECUTION TIME

Figure 7.3 Final comparison result: Comparison of Execution Time

Figure 7.3 provides a comparison of Execution Time, to various techniques (k-means, EM, sGEM, Apriori, FP-Growth and FP-Bonsai) with respect to the various dataset (Reuters-21578, 20 NG, and TDT-2 datasets).
• The relative Execution Time improvement of the FP-Bonsai approach achieves an efficiently less execution time compare to the existing and another standard approach.
• More over the FP-Bonsai approach outperforms the other approaches in terms of Execution time.
• The FP-Bonsai approach shows significant document clustering when compared with the k-means, EM, and sGEM for all the datasets.
• The percentage of improvement values shows that the FP-Bonsai on the 20 NG outperforms the other approaches.
• Expectation maximization algorithm, deals with the reassignment meanwhile maximization deals with re-computation. These steps are used in an iterative manner for successful document clustering.
• sGEM algorithm, re-allocates the new cluster membership to make the cluster process more efficient.
• K-means is a restrictive algorithm, can perform quite quickly. K-means leads to clusters that look like the same sphere of equal sized groups. More optimizations is required to reduce the time.
• Execution time is more in k-means when compared to EM and sGEM.
7.5. FINAL COMPARISON RESULT: MEMORY UTILIZATION

Figure 7.4 Final comparison result: Comparison of Memory Utilization

- The k-means, EM, sGEM, Apriori, FP-Growth and FP-Bonsai approaches are tested against the Reuters-21578 datasets, 20 NG datasets, and TDT-2 datasets.
- The experimental results are provided in terms of clustering memory utilization as shown in figure 7.4.
- When considering the Reuters-21578 datasets, it is observed that the memory utilization of term clustering of FP-Bonsai is 9 MB, but the memory utilization of the FP-Bonsai in TDT-2 dataset is 19 MB.
- Similarly, for other datasets, the memory utilization of semantic clustering technique is better than term clustering technique.
- Figure 7.4 makes it clear that the memory utilization of 20 NG dataset clustering approach is very significant when compared with term clustering technique for all the datasets.
7.6. SUMMARY

The improvement in performance of Ontology based partitioning system over k-means, EM, and sGEM techniques and frequent pattern mining approach Apriori algorithm, FP-Growth and FP-Bonsai Tree has been presented. In order to confirm, that the improved performance is significantly efficient in Ontology framework, the above inferences have been inferred.