

Abstract :

An unprecedented way is accomplished by using concept words derived from statistical context analysis between sentences which is better than traditional methods that uses only keyword representation. Through scaling to a very large dataset we proposed an algorithm which discovers and describes events with effective keyword networks, based on their coexisting peripheral co-occurrence. In our experiments, we used real-world news and supervised them into paraphrases by weighting the all attempted events. We evaluated our system by a set of terms that maximally discriminated the percussion in news which also keep the evidences.

Contribution :

★ Real World Data Collection

- Prothom Alo News Collection by Crawler (2016-2018)

★ Data Cleaning

- Proper Refining for Bengali font and Making Ready to Use

★ Feature Collection

- Feature Collected as per Event's Keyword Networks
- Feature Collected as per Event's Outbreak Area Network
- Every Feature is Weighted for Better Realization & Evaluation

■ Specimen

Year	Month	Day	খুন	অপহরণ	নারানারি	বিক্ষোভ	দুর্ঘটনা	আতঙ্ক	বিশাল	চটপুত্রাম	ঢাকা	খুলনা	রাজশাহী	রূপুর	সিলেট	ময়মনসিংহ
2017	1	1	24	0	0	2	0	0	0	0	2	0	0	0	16	0
2017	1	1	21	0	8	0	0	1	0	0	7	8	0	0	0	2
2017	1	1	4	1	0	0	0	1	0	16	2	0	1	0	0	0
2017	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
2017	1	1	15	0	0	0	0	1	0	2	2	0	3	0	0	0
2017	1	1	2	2	1	9	1	1	0	1	2	0	1	0	0	0
2017	1	1	2	0	1	0	2	0	0	0	0	0	0	0	0	0
2017	1	1	18	8	6	0	2	3	0	1	0	0	0	1	0	0
2017	1	1	4	2	2	1	1	5	0	0	0	0	0	0	0	0
2017	1	1	9	7	11	7	3	3	1	2	2	1	0	0	1	0

★ Classification And Clustering

- Evaluation for Better Classification & Clustering Through Ensemble Learning Algorithms

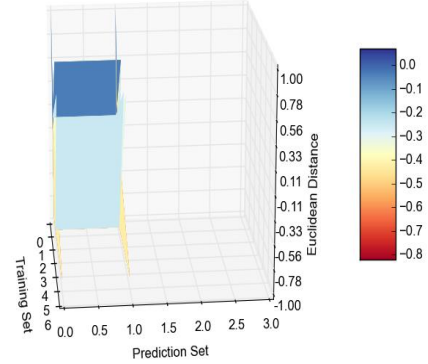
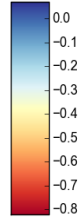
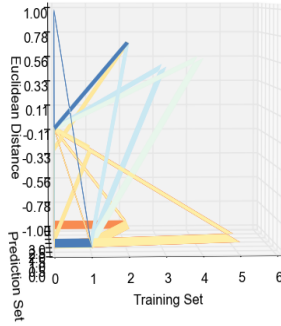
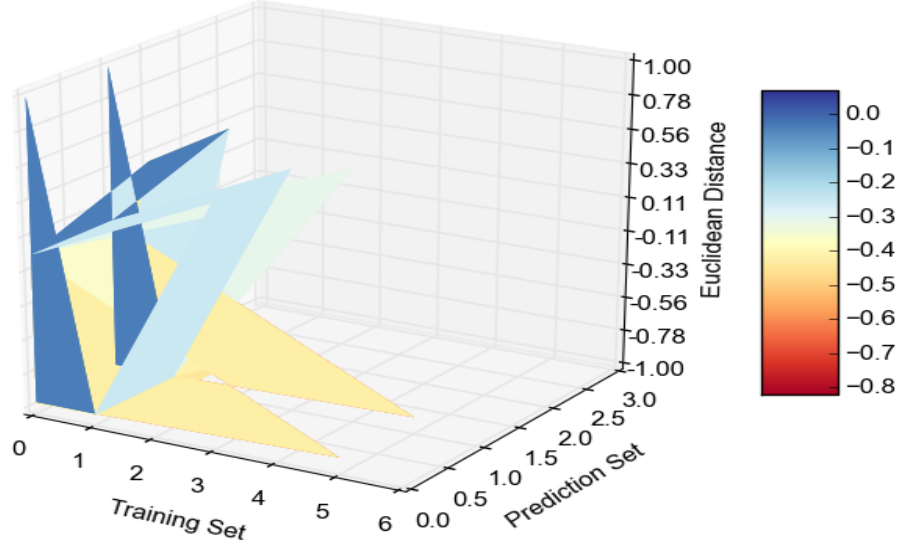
■ Specimen

- Total Number of Misclassified for Decision Tree : 1408
- Accuracy on the training subset: 0.782
- Accuracy on the test subset: 0.817
- Misclassified Index :
[1][2][4][5][7][8][9][10][14][17].....[7697][7698]

★ Machine Learning

- Data Analysis for Automates analytical model & identical patterns

★ Graph & Tree Visualization



□ Feature Extraction Keyword Network Specimen

- WordForEvents=["খুন", "নিহত", "ঘাত", "আঘাত", "হত্যা", "গুলি", "চাকু", "বন্দুক", "পিস্তল", "আগ্নেয়াস্ত্র", "ছুড়ি", "অস্ত্র", "সশস্ত্র", "রক্তপাত", "মার", "মেরে", "লাশ", "মৃত", "ঘাতক", "পিটিয়ে"]
- WordForArea=["সিলেট", "মৌলভীবাজার", "হবিগঞ্জ", "সুনামগঞ্জ", "বালাগঞ্জ", "বিয়ানীবাজার", "বিশ্বনাথ", "কোম্পানীগঞ্জ", "ফেঞ্চগঞ্জ", "গোলাপগঞ্জ", "গোয়াইনঘাট", "জৈন্তাপুর", "কানাইঘাট", "জকিগঞ্জ", "দক্ষিণ সুরমা", "ওসমানী", "বড়লেখা", "কমলগঞ্জ", "কুলাউড়া", "রাজনগর", "শ্রীমঙ্গল", "জুড়ী", "নবীগঞ্জ", "বাহুবল", "আজমিরীগঞ্জ", "বানিয়াচং", "লাখাই", "চুনারাঘাট", "মাধবপুর", "বিশ্বম্ভরপুর", "ছাতক", "জগন্নাথপুর", "দোয়ারাবাজার", "তাহিরপুর", "ধর্মপাশা", "জামালগঞ্জ", "শাল্লা", "দিরাই"]