CS388 NLP Project Report

Text Mining for Hidden Relations and Trending

C. Vic Hu

vic@cvhu.org

Ali Unwala

aliunwala@gmail.com

Abstract—

I. Introduction

II. Problem Definition and Algorithm

A. Task Definition

B. Algorithm Definition

1) Assign K topics to N particles uniformly
2) Add Gaussian noise to particles
3) Cluster particles into K groups (TF-IDF weights with cosine/Euclidean distances)
4) Compare the clusters with topics from the next year, apply discounts to current weights, and adjust to new weights
5) repeat

III. Experimental Evaluation

A. Methodology

B. Results

C. Discussion

IV. Related Work

V. Future Work

VI. Conclusion

References

[1]
[2] Blei, D. Introduction to Probabilistic Topic Models. Princeton University. 2011.