Discrete-Time Hopfield Neural Network Based Text Clustering Algorithm

Zekeriya Uykan[1], Murat Can Ganiz[2], Çağla Şahinli[2]

[1]Electronics and Communications Engineering Dept
[2] Computer Engineering Dept.
Dogus University, Istanbul, Turkey
{zuykan, mcganiz, csahinli}@dogus.edu.tr

Abstract. In this study we propose a discrete-time Hopfield Neural Network based clustering algorithm for text clustering for cases $L = 2^q$ where L is the number of clusters and q is a positive integer. The optimum general solution for even 2-cluster case is not known. The main contribution of this paper is as follows: We show that *i)* sum of intra-cluster distances which is to be minimized by a text clustering algorithm is equal to the Lyapunov (energy) function of the Hopfield Network whose weight matrix is equal to the Laplacian matrix obtained from the *document-by-document* distance matrix for 2-cluster case; and *ii)* the Hopfield Network can be iteratively applied to text clustering for $L = 2^k$. Results of our experiments on several benchmark text datasets show the effectiveness of the proposed algorithm as compared to the *k*-means.

Keywords: Text clustering, discrete-time Hopfield Neural Networks, Lyapunov function, max-cut graph partitioning.

1 Introduction

Clustering is one of the most important techniques for organizing documents in an unsupervised manner since large amounts of textual data accumulated in organizations. In document clustering documents are automatically grouped into predefined number of clusters based on their similarity or dissimilarity. In general, dissimilarity or distance is determined in vector space either using a distance metric such as Euclidean distance or Manhattan distance or a similarity metric such as Cosine similarity. There are various types of document clustering methods such as partitioning (e.g. [1], [2]), hierarchical (e.g. [3]), density-based (e.g. [4]), grid-based and model-based methods (e.g. [5]). A comparative study for various document clustering methods is given in [5]. For more information, a survey on document clustering algorithms can be found in e.g. [6]. One of the most commonly used clustering algorithms is *k*-means due to its simplicity and effectiveness. It is also frequently used in text clustering studies as baseline clustering algorithm due to its low complexity and its relatively good performance when combined with appropriate distance metric [7].

In this paper, we present a discrete-time Hopfield Neural Network (HNN) based clustering algorithm for text clustering. Hopfield Network has been an important focus of research area since early 1980s whose applications varied from combinatorial optimization to image restoration, from various control engineering optimization problems in robotics to content-addressable memory systems, among many others. For further info about HNN, see any related texbooks. Hopfield Network has been applied, in the context of text/document clustering, to bipartite graph representing term-by-document association matrix in e.g., [8] [9] for Latent Semantic Indexing. In this paper, our approach is completely different than theirs because we apply the Hopfield Network to the graph of the *document-by-document* matrix for direct clustering of the documents. This yields a simple but efficient text clustering algorithm as shown by the experiment results.

The rest of the paper is organized as follows: Section 2 summarizes the approach followed. Section 3 presents and analyzes the proposed text clustering algorithm. Experiment setup is explained in section 4. The experiment results are shown in section 5, followed by the conclusions in section 6.

2 Approach

As a high level definition, document clustering can be defined as partition of N documents into a predetermined number of L subsets so that documents assigned to each subset are more similar to each other than the documents assigned to different subsets (e.g. [10]). Although the Euclidean distance is one of the most commonly used similarity metric in clustering problems, it's reported in various studies that its performance is rather poor in document clustering. Instead, the cosine similarity performs better than the Euclidean distance in this context (see e.g. [11], [12]). Therefore, the cosine similarity is the main distance metric for the evaluation of our results.

The cosine similarity takes values between -1 and 1. Representing the documents as vectors, the cosine distance between document i ($\mathbf{y}_i$) and j ($\mathbf{y}_j$) is

$$w_{ij} = 1 - \frac{\sum_k y_{ik} \times y_{jk}}{\sqrt{\sum_k y_{ik}^2} \times \sqrt{\sum_k y_{jk}^2}} \tag{1}$$

We define the document distance matrix as follows:

$$\mathbf{W} = \lfloor w_{ij} \rfloor_{N \times N} \tag{2}$$

The aim of text clustering is to minimize the sum of intra-cluster distances in (3):

$$\min \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \left(x_i - x_j \right)^2 \right\}, \tag{3}$$

$$\text{where } \left(x_i - x_j \right) = \begin{cases} 1, & \text{if doc } i \text{ and doc } j \text{ in the same cluster,} \\ 0, & \text{otherwise.} \end{cases}$$

3 Hopfield Network for Text Clustering

3.1 Text Clustering by Hopfield Network for $L = 2$ case

In what follows, we present one of the main contributions of this paper:

Proposition 1: Determining the weight matrix of the Hopfield Network as the Laplacian matrix obtained from the *document-by-document* distance matrix minimizes the sum of *intra-cluster distances* in (1).

Proof: The volume (i.e., entrywise 1-norm) of the *document-by-document* distance matrix $\mathbf{W}$ is equal to

$$vol(\mathbf{W}) = \|\mathbf{W}\|_1 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \tag{4}$$

Then, considering the grouping of the N documents into L clusters C_1 to C_2, we write

$$vol(\mathbf{W}) = \sum_{l=1}^{2}\left\{\sum_{i\in C_l}^{N_l}\sum_{j\in C_l}^{N_l} w_{ij} + \sum_{i\in C_l}^{N_l}\sum_{j\in \overline{C}_l}^{N-N_l} w_{ij}\right\} \tag{5}$$

where C_l and $\overline{C}_l$ represent the cluster l and the other cluster, respectively, where $l \in \{1, 2\}$. Eq.(5) can be written as

$$vol(\mathbf{W}) = \text{constant} = \sum_{l=1}^{2} vol(C_l) + \sum_{l=1}^{2} cut(C_l, \overline{C}_l) \tag{6}$$

where $vol(C_l) = \sum_{i\in C_l}^{N_l}\sum_{j\in C_l}^{N_l} w_{ij}$ is the sum of *intra-cluster weights* for cluster l, and $cut(C_l, \overline{C}_l) = \sum_{i\in C_l}^{N_l}\sum_{j\in \overline{C}_l}^{N-N_l} w_{ij}$ represents the sum of the *inter-cluster weights* between cluster l and the other cluster. From (6)

$$\min_{\{C_l\}_{l=1}^{L}}\left\{\sum_{l=1}^{2} vol(C_l)\right\} \equiv \max_{\{C_l\}_{l=1}^{L}}\left\{\sum_{l=1}^{2} cut(C_l, \overline{C}_l)\right\} \tag{7}$$

From eq.(6) and (7), minimizing the total intra-cluster weights is equal to well-known weighted *maxCut* problem in graph theory (e.g. [2]). The unnormalized Laplacian matrix (e.g. [2], [12]) is given as

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{8}$$

Table 1. Proposed text clustering algorithm for $L = 2^q$

1. Establish the dissimilarity (distance) matrix.
2. Repeat for *n=1:L*
   - Determine the documents of $\left(N / 2^L\right)$ to be spectrally clustered for *L=2*.
   - Determine the $\left(N / 2^L \times N / 2^L\right)$ dimensional Laplacian matrix.
   - Set the weight matrix of the Hopfield Network as the Laplacian matrix.
   - Determine the partitioning of the graph by the Hopfield Network.
3. Finally, determine the *L* groups of documents according to the signs of the Hopfield Networks in the loop at step 2.

▪

where diagonal matrix $\mathbf{D} = [d_{mn}] = \begin{cases} \sum_{j=1,\ (j \neq i)}^{N} w_{ij}, & \text{if } m = n \\ 0, & \text{otherwise} \end{cases}$.

It's known that (see e.g. [2]), the *maxCut* problem can be formulized for *L=2* as follows

$$\max\left\{\mathbf{x}^T \mathbf{L}\mathbf{x}\right\} = \max\left\{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(x_i - x_j\right)^2\right\}, \quad \text{where } x_i, x_j \in \{-1, +1\} \tag{9}$$

It's well-known that the discrete-time Hopfield Network minimizes the following Lyapunov (energy) function

$$\min\left\{-\mathbf{x}^T \mathbf{L}\mathbf{x}\right\} \qquad \text{where } x_i, x_j \in \{-1, +1\} \tag{10}$$

From (6), (7), (9) and (10), determining the weight matrix of the Hopfield Network as the Laplacian matrix obtained from the *document-by-document* distance matrix minimizes the sum of intra-cluster distances and maximizes the sum of inter-cluster distances for *L=2* case. This completes the proof.

3.2 Hopfield Network based Text Clustering Algorithm for $L = 2^q$ case

The proposed Hopfield Network based text clustering algorithm is based on proposition 1 above: We propose to iteratively run the Hopfield Network for $L = 2^q$ case as explained in Table 1.

4 Experiment Setup

In order to examine the performance of Hopfield Network clustering algorithm, we run it on various benchmark textual datasets and compare the evaluation results with the most well known and most commonly used algorithm in document clustering: The *k*-means. Specifically, we use 19 text classification benchmark data sets [13]. These data sets are mainly from Reuters, TREC and OHSUMED. They are labeled datasets and frequently used in especially text classification research. Use of labeled datasets

allows us to calculate several different evaluation metrics including misclassification rate and analyze the performance of clustering algorithms from different perspectives. We use the number of classes as number of clusters in our experiments, especially to see how the clustering solution overlaps with class labels. In other word the degree of agreement with clustering algorithm and human experts who label these datasets. Since Hopfield network clustering algorithm requires the number of clusters to be the power of two, we created several different subsets of news3 dataset with 32, 16, 8, 4, and 2 classes. This also provides us opportunity to analyze performance of the algorithms in varying cluster/class sizes. Similarly, 16 class subsets are created for fbis, wap, and re1 datasets. Characteristics of our datasets including number of documents, terms, classes, and average number of documents per class are given in Table 2. News3 is the largest dataset. It is important to note that majority of our datasets has skewed class distribution.

Table 2. Datasets

Name	♯ of documents	Dictionary	# of classes	# of docs / class
news3-32c	7510	26833	32	~235
news3-16c	4065	26833	16	~255
news3-8c	2554	26833	8	~319
news3-4c	1035	26833	4	~259
news3-2c	256	26833	2	128
fbis-16c	2425	2001	16	~152
wap-16c	1343	8461	16	~84
re1-16c	1244	3759	16	~78

We implement and run our experiments using WEKA data mining library [14]. The datasets we use are supervised meaning that each document is labeled by a category (a.k.a class) label. Using supervised datasets in evaluation of clustering algorithms is common e.g.[7]-[10], and allows us to do more objective evaluation by using several different evaluation metrics. We use four different external criteria of clustering quality: percentage of incorrectly clustered instances (misclassification %), Normalized Mutual Information (NMI), F-measure and entropy. NMI and entropy are commonly used metrics in document clustering (e.g.[7], [8]). While calculating these metrics, we cluster the dataset into the same number of clusters with the number of categories and label the clusters with category labels in order to apply a wide range of evaluation metrics. If a particular category instances are majority in the cluster, it is labeled with this particular category. After this, we compare how well the clustering results match the category labels of documents. For the formulations of the metrics, namely misclassification rate, Normalized Mutual Information (NMI), F-measure and entropy, see e.g.[7], [8] page 358. We here skip the details due to the space restrictions of the paper. Combination of these four evaluation metrics provides a broad perspective for evaluating the text clustering solutions. Each dataset has been run 10 times with different random seeds. Results show the average of these 10 results and standard deviations for algorithms affected from randomization. We use two different initialization approaches for Hopfield Network. For the first one (Hopfield), the initial state of the Hopfield

Network is fixed to 1, thus its standard deviation is 0. For the second one (Hopfield-Rand), initial state values is assigned to -1 or 1 randomly.

5 Results

Table 3 represents the misclassification rate of the clustering algorithms. We also include the results of a dummy clustering algorithm which randomly assigns documents to *k* clusters (denoted as Random) in our experiments as a baseline in order to see how intelligently the other algorithms create clusters. Dataset names also indicate the number of classes. In *k*-means and Random, the values are assigned as the average of 10 results ± standard deviation. For each dataset we show the results of best performing algorithm with bold font. As can be seen from the Table 3, Hopfield network clustering algorithm (Hopfield) outperforms *k*-means algorithm in all datasets except re1 in terms of misclassification rate. Similar pattern can be seen in NMI results in Table 4, F-measure results in Table 5, and Entropy results in Table 6. For majority of the datasets, Hopfield outperforms *k*-means algorithm. The performance improvement is most obvious in 2-cluster case where Hopfield outperforms *k*-means by a wide margin in terms of all evaluation metrics.

Table 3. Misclassification rate results

Dataset	K-means	Hopfield	Hopfield-Rand	Random
news3-32c	0,529 ± 0,024	**0,461**	0.476 ± 0.007	0,903 ± 0,001
news3-16c	0,462 ± 0,016	**0,411**	0,422 ± 0,015	0,826 ± 0,001
news3-8c	0,300 ± 0,023	**0,268**	**0,268 ± 0,001**	0,727 ± 0,001
news3-4c	0,165 ± 0,054	**0,158**	**0,158 ± 0,001**	0,451 ± 0,001
news3-2c	0,366 ± 0,163	**0,023**	**0,023 ± 0,001**	0,471 ± 0,022
wap-16c	0,377 ± 0,027	**0,313**	0,324 ± 0,008	0,744 ± 0,003
fbis-16c	0,410 ± 0,021	0,372	**0,348 ± 0,010**	0.786 ± 0,002
re1-16c	**0,289 ± 0,009**	0,301	0,313 ± 0,019	0,680 ± 0,007

Table 4. NMI results

Dataset	K-means	Hopfield	Hopfield-Rand	Random
news3-32c	**0,564 ± 0,006**	0,556	0.543 ± 0.001	0,020 ± 0,001
news3-16c	0,546 ± 0,018	**0,555**	0,551 ± 0,060	0,010 ± 0,001
news3-8c	0,569 ± 0,026	0,599	**0,600 ± 0,002**	0,004 ± 0,001
news3-4c	0,641 ± 0,069	**0,642**	**0,642 ± 0,001**	0,002 ± 0,001
news3-2c	0,197 ± 0,268	**0,841**	**0,841 ± 0,001**	0,004 ± 0,005
wap-16c	0,524 ± 0,020	**0,536**	0,529 ± 0,005	0,034 ± 0,003

fbis-16c	0,525 ± 0,020	0,534	**0,545 ± 0,008**	0,019 ± 0,001
re1-16c	**0,489 ± 0,011**	0,465	0,455 ± 0,017	0,043 ± 0,004

Table 5. F-measure results

Dataset	K-means	Hopfield	Hopfield-Rand	Random
news3-32c	0,308 ± 0,020	**0,440**	0.433 ± 0.010	0,099 ± 0,001
news3-16c	0,346 ±0,029	**0,492**	0,475 ± 0,012	0,176 ± 0,001
news3-8c	0,550 ± 0,034	**0,610**	0,600 ± 0,013	0,272 ± 0,003
news3-4c	0,707 ± 0,120	0,682	**0,770 ± 0,066**	0,497 ± 0,003
news3-2c	0,641 ± 0,220	**0,977**	**0,977 ± 0,001**	0,529 ± 0,021
wap-16c	0,532 ± 0,039	**0,581**	0,566 ± 0,011	0,263 ± 0,003
fbis-16c	0,449 ± 0,020	**0,536**	**0,536 ± 0,009**	0,222 ± 0,003
re1-16c	**0,597 ± 0,019**	0,573	0,571 ± 0,015	0,332 ± 0,006

Table 6. Entropy results

Dataset	K-means	Hopfield	Hopfield-Rand	Random
news3-32c	0,461 ± 0,012	**0,414**	0.427 ± 0.001	0,938 ± 0,001
news3-16c	0,457 ± 0,020	0,393	**0,369 ± 0,006**	0,917 ± 0,001
news3-8c	0,388 ± 0,029	0,329	**0,328 ± 0,002**	0,893 ± 0,001
news3-4c	0,288 ± 0,067	**0,253**	**0,253 ± 0,001**	0,842 ± 0,001
news3-2c	0,820 ± 0,271	**0,159**	**0,159 ± 0,001**	0,996 ± 0,005
wap-16c	0,406 ± 0,023	**0,354**	0,361 ± 0,004	0,819 ± 0,002
fbis-16c	0,418 ± 0,017	0,378	**0,367 ± 0,008**	0,861 ± 0,001
re1-16c	**0,349 ± 0,010**	0,363	0,372 ± 0,015	0,737 ± 0.004

We show by proposition 1 that the discrete-time Hopfield Network minimizes the sum of intra cluster distances in (3). Figure 1 shows an example how the cost function in (3) decreases by the number of iterations for the dataset news3-2c as an example.

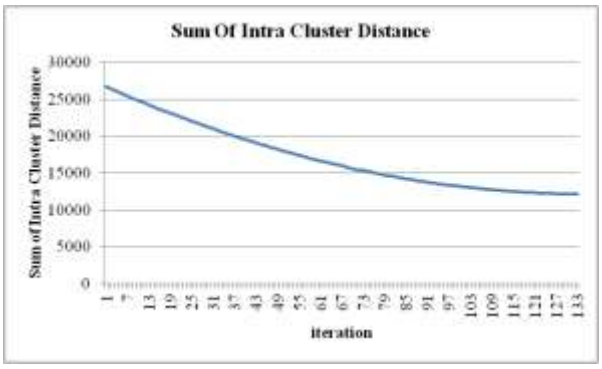

Fig. 1. Sum of intra cluster distance

6 Conclusions

In this study we propose a discrete-time Hopfield Network based text clustering algorithm. The main contribution of this paper is: We show that i) sum of intra-cluster distances which is to be minimized by a text clustering algorithm is equal to the Lyapunov (energy) function of the Hopfield Network whose weight matrix is equal to the Laplacian matrix obtained from the document-by-document distance matrix for 2-cluster case; and ii) the Hopfield Network can be iteratively applied to text clustering for $L = 2^k$. Results of our experiments on several benchmark text datasets show the effectiveness of the proposed algorithm as compared to the *k*-means.

References

1. Jain, A.K., Murty M.N. and Flynn, P.J.: Data clustering: a review. ACM Comput Surv., 31(3):264–323 (1999).
2. Luxburg, U.V.: A Tutorial on Spectral Clustering. Technical Report TR-149. Max-Planck Institute for Biological Cybernetics (August 2006).
3. Kim, H. and Lee, S.: An intelligent information system for organizing online text documents. Knowledge and Information Systems, Springer, 6(2):125–149 (2004).
4. Hinneburg, A. and Keim, D.: A general approach to clustering in large databases with noise. Knowledge and Information Systems, Springer, 5(4):387–415 (2003).
5. Zhong, S. and Ghosh, J.: Generative model-based document clustering: a comparative study. Knowledge and Information Systems, Springer. Vol.8: pp374–384 (2005).
6. Zanasi, A.: Text Mining and its Applications to Intelligence. Crm and Knowledge Management (Advances in Management Information). WIT Press. (2005).
7. Huang, A.: Similarity Measures for Text Document Clustering. NZCSRSC 2008. New Zealand (April 2008).
8. Ding, C.H.Q.: Data clustering: Principal components, Hopfield and self-aggregation networks. NERSC Division, Lawrence Berkeley National Lab., Univ. of California, Berkeley.
9. Ding, C.H.Q.: Document retrieval and clustering: from principal component analysis to self-aggregation networks. Lawrence Berkeley National Laboratory, Berkeley , CA 94720.
10. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. 2nd ed., Morgan Kaufmann Publishers (March 2006).

11. Uykan, Z.: Spectral Based Solutions for (Near) Optimum Channel/Frequency Allocation. In Proc. of IWSSIP 2011. Sarajevo, BiH (June 2011).
12. Luxburg, U.V., Belkin, M. and Bousquet, O.: Consistency of spectral clustering. Annals of Statistics. Vol.36, no. 2, pp.555-586 (2008).
13. Forman, G. & Cohen, I., Learning from Little: Comparison of Classifiers Given Little Training, In Proc. ECML'04 (2004).
14. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations. 11, (1), 10-18 (2009).