exemplar itself, the availability of point k as an exemplar can be increased if some other points have positive responsibilities for point k being their exemplar. The total sum is thresholded to limit the influence of strong incoming positive responsibilities so that it cannot go above zero. The ―self-availability‖ a(k,k) is updated differently:

This message reflects that point k is an exemplar sent to candidate exemplar k from other points. The above update rules require only local and simple computations that are easily implemented in equation and messages need be exchanged between pairs of points with known similarities Availabilities and Responsibilities can be combined to identify exemplars at any point during affinity propagation. For point i, the value of k that maximizes a(i,k) + r(i,k) either identifies point i as an exemplar if k = i, or identifies the data point that is the exemplar for point i. After changes in the messages fall below a threshold , the message-passing procedure may be terminated after a fixed number of iterations. To avoid numerical oscillations that arise in some circumstances, it is important that they be damped when updating the messages. Each message is set to l times its value from the previous iteration plus 1 – λ times its prescribed updated value, where the damping factor l is between 0 and 1. Each iteration of affinity propagation consists of :

I. Updating all responsibilities given the availabilities.
II. Updating all availabilities given the responsibilities and
III. Combining availabilities and responsibilities to monitor the exemplar decisions and terminate the algorithm.

Disadvantages of Affinity Propagation

1) It is hard to know the value of the parameter preferences which can yield an optimal clustering solution.
2) When oscillations occur, AP cannot automatically eliminate them.

4. Incremental Affinity Propagation

A semi-supervised scheme called incremental affinity propagation clustering. Incremental Affinity Propagation integrates AP algorithm and semi-supervised learning. The labeled data information is coded into similarity matrix in some way. And the incremental study is performed for amplifying the prior knowledge. In the scheme, the pre-known information is represented by adjusting similarity matrix. An incremental study is applied to amplify the prior knowledge. To examine the effectiveness of this method, concentrate it to text clustering problem. In this method, the known class information is coded into the similarity matrix initially. And then after running AP for a certain number of iterations , the most convinced data are put into the "labeled data set" and reset the similarity matrix. This process is repeated until all the data are labeled. Compared with the method in , the dealing with constrained condition in this scheme is soft and objective. Furthermore, the introduction of incremental study amplifies pre-knowledge about the target data set and therefore leads to a more accurate result. Also, the parameter of "preference" in the method is self-adaptive mainly according to the target number of clusters. Focused on text clustering problem, Cosine coefficient method is used to compute the similarity between two different points (texts) in the specific I-APC algorithm.

In summary, the I-APC scheme can be described as follows:

1. Initialize: including the initializations of labeled data set *L*, the responsibility matrix and availability matrix, similarities between different points and self-similarities say, set as a common value as well.
2. If the size of *L* reaches a pre-given number *P*, goto step 5, else run AP.
3. Select the most convinced data point and then update *L*. Reset similarities between different points , and self-similarities
4. Go to step 2.
5. Output results, end.

The flow chart of the I-APC scheme is shown in Fig 2.

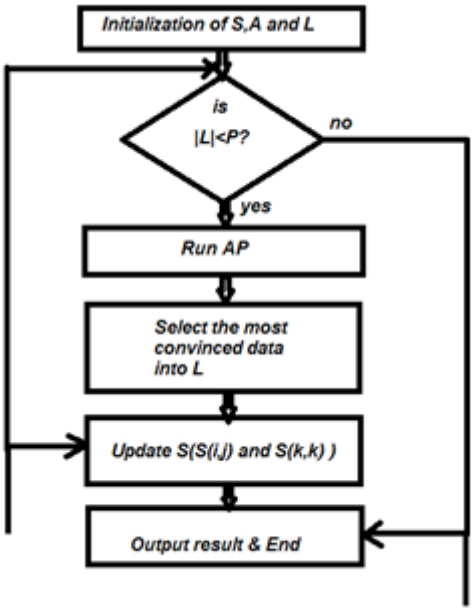

Figure 2: Flowchart of I-APC scheme

5. Conclusion

In this survey, various clustering approaches and algorithms in document clustering are described. . A new clustering algorithm which combines Affinity Propagation with semi-supervised learning ,i.e Seeds Affinity Propagation algorithm is present. In comparison with the classical clustering algorithm k-means, SAP not only improves the accuracy and reduces the computing complexity of text clustering and but also effectively avoids being trapped in local minimum and random initialization. Whereas Incremental Affinity Propagation integrates AP algorithm and semi-supervised learning. The labeled data information is coded into similarity matrix. The Adaptive Affinity Propagation algorithm first computes the range of preferences, and then searches the space to find the value of preference which can generate the optimal clustering results compare to AP approach which cannot yield optimal clustering results because it sets preferences as the median of the similarities.

The area of document clustering has many issues, which need to be solved. We hope , the paper gives interested readers a broad overview of the incremental affinity propagation clustering technique. As a future work, improvement over the existing systems with better results which offer new information representation capabilities with different techniques can be attempted.

References

[1] Leilei Sun, Chonghui Guo, "Incremental Affinity Propagation Clustering Based on Message Passing," IEEE Transactions On Knowledge And Data Engineering Vol:Pp No:99 Year 2014
[2] X. Zhang, C. Furtlehner, and M. Sebag, "Frugal and Online Affinity Propagation," Proc. Conf. francophone sur l'Apprentissage (CAP '08), 2008.
[3] J. Beringer and E. Hullermeier, "Online Clustering of Parallel Data Streams," Data and Knowledge Engineering, vol. 58, no. 2, pp. 180-204, Aug. 2006.
[4] J. Beringer and E. Hullermeier, "Online Clustering of Parallel Data Streams," Data and Knowledge Engineering, vol. 58, no. 2, pp. 180-204, Aug. 2006.
[5] B.J. Frey and D. Dueck, "Response to Comment on 'Clustering by Passing Messages Between Data Points'," Science, vol. 319, no. 5864, pp. 726a-726d, Feb. 2008.
[6] F.R. Kschischang, B.J. Frey, and H.A. Loeliger, "Factor Graphs and the Sum-product Algorithm," IEEE Trans. Information Theory, vol. 47, no. 2, pp. 498-519, Feb. 2001.
[7] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms," IEEE Trans. Information Theory, vol. 51, no. 7, pp. 2282-2312, July 2005.
[8] L. Ott and F. Ramos, "Unsupervised Incremental Learning for Long-term Autonomy," Proc. 2012 IEEE Int. Conf. Robotics and Automation (ICRA '12), pp. 4022-4029, May 2012,
[9] Adil M. Bagirov, Julien Ugon, Dean Webb, " Fast modifiedglobal k-means algorithm for incremental cluster construction," Pattern Recognition 44 (2011) 866–876
[10] A.K. Jain, "Data Clustering: 50 Years Beyond K-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, June 2009.