

NMF Ensembles? Not for Text Summarization!

Alka Khurana

Department of Computer Science
University of Delhi
Delhi, India
akhurana@cs.du.ac.in

Vasudha Bhatnagar

Department of Computer Science
University of Delhi
Delhi, India
vbhatnagar@cs.du.ac.in

Abstract

Non-negative Matrix Factorization (NMF) has been used for text analytics with promising results. Instability of results arising due to stochastic variations during initialization makes a case for use of ensemble technology. However, our extensive empirical investigation indicates otherwise. In this paper, we establish that ensemble summary for single document using NMF is no better than the best base model summary.

1 Introduction

Non-negative Matrix factorization (NMF) has demonstrated promise in text analytic tasks like topic modeling (Suh et al., 2017; Qiang et al., 2018; Belford et al., 2018), document summarization (Lee et al., 2009; Khurana and Bhatnagar, 2019) and document clustering (Shahnaz et al., 2006; Shinnou and Sasaki, 2007). The method finds favour due to the presence of non-negative elements in resultant factor matrices, which enhance intuitive understanding of the underlying latent semantic structure of the text (Lee and Seung, 1999).

Recent applications of ensemble methods for NMF based topic modeling has shown considerable promise (Suh et al., 2017; Qiang et al., 2018; Belford et al., 2018). These observations drive our motivation for exploring NMF ensembles for document summarization task.

1.1 NMF for Text Summarization

Consider a (pre-processed) document D consisting of n sentences (S_1, S_2, \dots, S_n) and m terms (t_1, t_2, \dots, t_m) represented by a Boolean term-sentence matrix $A_{m \times n}$. NMF decomposition of A results into two non-negative factor matrices W and H , where W is $m \times r$ term-topic (feature) matrix and H is $r \times n$ topic-sentence (co-efficient) matrix, with $r \ll \min\{m, n\}$.

Columns in W correspond to document topics represented as $\tau_1, \tau_2 \dots \tau_r$ in the latent semantic space, and columns in H represent sentences in D . Element w_{ij} in W signifies the contribution of term t_i in topic τ_j , and element h_{ij} in H denotes the strength of topic τ_i in sentence S_j . Deft manipulation of the elements of two factor matrices yields distinctive sentence scores (Lee et al., 2009; Khurana and Bhatnagar, 2019). Top scoring sentences are selected to generate summary of desired length.

Instability of NMF for Text Summarization:

Even though NMF based automatic text summarization is unsupervised and carries advantages of language, domain and collection independence, yet it has been used sporadically for summarization. The reason can be linked to stochastic variations in factor matrices due to random initialization.

Repeated NMF factorization of the input term-sentence matrix results into different sentence scores, generating different summaries. This ambivalence renders the resulting NMF summaries dubitable. In authors' opinion, this has retarded development in this line of research.

Lee et al. (2009) suggested a simplistic fix to this problem by using static initialization for W and H . Experiment using DUC2002¹ data-set with varying initial seed values for W and H shows that the best initialization value is document specific (Fig. 1). Hence, fixed initialization of factor matrices is not a prudent idea.

Another fix for the problem is to use NNDSVD (Non-negative Double Singular Value Decomposition (Boutsidis and Gallopoulos, 2008)) based initialization for NMF factor matrices. Our earlier work (Khurana and Bhatnagar, 2019) establishes that this initialization method improves the sum-

¹ Well studied data-set for single document summarization available at <https://duc.nist.gov> consisting of 533 unique documents.

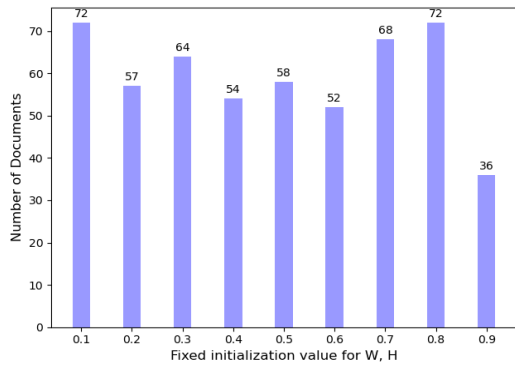


Figure 1: Number of documents with highest ROUGE-L recall score for different initial seed values of NMF factor matrices.

mary quality over fixed initialization for several benchmark data-sets.

1.2 NMF Ensembles

Random initialization has been exploited by clustering and topic modeling researchers to create ensembles (Greene et al., 2008; Belford et al., 2018; Qiang et al., 2018). Ensembling is a machine learning technique, which combines (multiple) varying base models to construct a consensus model, which is expected to perform better than individual base models.

Effectiveness of NMF ensembles in text analytics, specifically in topic modeling (Belford et al., 2018; Qiang et al., 2018), motivated the current research. We initiated the study with the aim to leverage stochastic variations in NMF factors and resulting diverse base summaries, to combine into stable ensemble summary.

Extrapolating earlier studies, we expected that NMF ensemble summary, smoothed over multifarious summaries obtained by randomly initialized NMF factors, will accomplish higher ROUGE scores. However, our investigations establish that NMF ensembles are not effective. Rather, despite heavy overhead of creating multiple base models and combining them, NMF ensemble often perform worse than the best base model for single document extractive summarization.

2 NMF Ensembles for Text Summarization

Ensemble methods are employed in supervised, semi-supervised and unsupervised learning settings. In all scenarios, they comprise two phases.

In the first phase, *diverse* base models are generated. Diversity in models is recognized to be the key factor for improvement (Kuncheva and Hadjitodorov, 2004), which is commonly sourced from variations in base algorithm, algorithmic parameters or data itself.

In second phase, multiple base models are combined using a consensus (aka integration) function. Wide variety of choices for creating diversity and combining base models gives rise to numerous possibilities for creating ensembles (Zhou, 2012).

2.1 Generation of diverse base models

Repeated application of NMF on the term-sentence matrix leads to generation of multiple base models. In the present context, we achieve diversity using two methods.

- (i) Repeated factorization² of A using NMF with random initial seed values for W and H , and
- (ii) Repeated factorization by varying the number of latent topics into which the document is decomposed, as suggested in (Greene et al., 2008). This strategy implicitly embeds variations that arise out of random initialization.

In practice, variation in choice of NMF solvers and initialization methods is also a source of diversity. We however, refrain from following this direction because of weak scientific ground.

2.2 Combining NMF base models

We examine six combining methods, in increasing order of complexities, to generate consensus summaries. First three are simple aggregation methods, where sentence scores are combined directly. Next two methods are based on rank manipulation of the scored sentences. Finally, we use *Stacking*, which is a sophisticated combining method (Zhou, 2012; Belford et al., 2018).

- i. **Average:** We calculate the consensus score of each sentence in the document by averaging sentence scores over all base models, and use it for summary sentence selection.
- ii. **Median:** Since average is sensitive to outliers, we calculate median score of each sentence across all base models and use it for summary sentence selection.
- iii. **Quartile:** We obtain consensus score by considering third quartile of the sentence scores across all base models, and use it for summary sentence

²We calculate r according to the formula proposed in (Aliguliyev, 2009).

selection.

iv. **Voting:** We rank sentences based on their scores. Consensus rank of a sentence is the most frequent rank (majority) of the sentence amongst base models. Top ranking sentences are selected for summary.

v. **Ranking:** We count the number of times a sentence appears among the top k scoring sentences (k is the desired number of sentences). Finally, sentences are ranked based on the frequency of appearing among top- k ranked sentences. Top scoring sentences are included in summary.

vi. **Stacking:** *Stacking* is a well established combining method, which combines base level models to create a meta-training set (Zhou, 2012). Subsequently, ensemble model is trained on this meta-training set. We stack topic-term (W^T) matrices (base level models) as meta-training set (\widehat{W}) for producing the *stacked ensemble* (Belford et al., 2018).

\widehat{W} matrix is factorized using NMF with NNDSVD initialization to obtain ensembled topic-term matrix, which along with A is used for scoring sentences using term-oriented sentence selection method, NMF-TR, proposed in (Khurana and Bhatnagar, 2019). Finally, top- k scoring sentences are included in summary.

3 Performance Evaluation

In this section, we present extensive experimentation³ carried out to investigate the performance of NMF ensembles for extractive summarization. First, we evaluate the performance based on combining methods described in Sec. 2 and the sizes of ensemble. Next, we study the effect of diversity in base models on ensemble performance. Finally, we test the statistical significance of our results. All experiments are performed on DUC2001⁴ data-set consisting of 308 documents and DUC2002¹ data-set. We report macro averaged ROUGE recall scores (R-1: ROUGE-1, R-2: ROUGE-2, R-L: ROUGE-L) (Lin, 2004).

In interest of brevity, all results are reported as performance gain (+) or loss (-) over NMF-TR scores as baseline method. Table 1 shows ROUGE scores of this baseline for DUC2001 and DUC2002 data-sets reported in (Khurana and Bhatnagar, 2019).

³Code is available <https://github.com/alkakhurana/NMF-Ensembles>

⁴Available at <https://duc.nist.gov>

	R-1	R-2	R-L
DUC2001	44.7	15.9	39.3
DUC2002	49.0	21.5	44.1

Table 1: Macro-averaged ROUGE recall performance of NMF-TR method (Khurana and Bhatnagar, 2019).

3.1 Examining Combining Methods

Primary objective of this experiment is to examine the comparative performance of model integration methods. However, size of an ensemble is a crucial factor that determines the quantum of performance gain. Oversized ensembles have obvious computational and memory overheads, while undersized ensembles run the risk of little performance gains and reduced stability. Ergo, we evaluate the performance of each combining method on ensembles of varying sizes. We compute macro-averaged ROUGE recall scores, each score averaged over ten executions to combat random variations. Table 2 shows the performance differential for six combining methods for ten different sizes for DUC2002 data-set. A cursory glance is sufficient to conclude that there are more *negs* than *pos*'s. Degradation in performance is most unexpected. Macro-level analyses in the bottom row and rightmost column consolidate the surprise.

The bottom row 'Total' shows the number of times the ensemble improves summary quality across all combining methods. For ensemble size 100, there is $\approx 50\%$ chance (9/18) of improving the summary quality across all methods. The rightmost column 'Total' shows the number of times a combining method improves summary across all sizes for each integration method. It suggests that simple combining methods improve marginally even for large size ensemble.

To confirm the trend, we repeated the same experiment with DUC2001 data-set (Table 3). Apparently there is better chance of improvement for this data-set using NMF ensembles, but the gain is meagre (less than 0.5 in each case) and does not justify the computational overhead.

Consolidating observations from Table 2 & 3, none of the the combining methods yield noticeably better quality summaries than the baseline method. Further, increasing the size of ensemble also does not hold promise.

Since Table 2 & 3 exhibit similar trends, we choose to perform remaining experiments on DUC2002 data-set as it clearly demonstrates infirmity of NMF ensembles.

DUC2002												
		$\ell = 2$	$\ell = 4$	$\ell = 6$	$\ell = 8$	$\ell = 10$	$\ell = 20$	$\ell = 30$	$\ell = 40$	$\ell = 50$	$\ell = 100$	Total
Average	R-1	-0.572	-0.268	-0.225	-0.098	-0.118	-0.060	+0.040	+0.108	+0.151		4
	R-2	-0.727	-0.429	-0.447	-0.293	-0.278	-0.220	-0.123	-0.084	-0.128	-0.027	0
	R-L	-0.557	-0.279	-0.292	-0.164	-0.195	-0.132	-0.039	+0.022	-0.030	+0.069	2
Median	R-1	-	-0.165	-0.048	-0.050	-0.125	+0.047	+0.170	+0.166	+0.159	+0.219	5
	R-2	-	-0.362	-0.248	-0.287	-0.279	-0.123	-0.003	-0.053	-0.043	+0.024	1
	R-L	-	-0.209	-0.098	-0.138	-0.185	-0.021	+0.093	+0.063	+0.069	+0.124	4
Quartile	R-1	-	-0.367	-0.355	-0.145	-0.212	-0.093	+0.029	-0.044	+0.041	+0.154	3
	R-2	-	-0.550	-0.606	-0.406	-0.401	-0.292	-0.212	-0.257	-0.187	-0.057	0
	R-L	-	-0.392	-0.408	-0.227	-0.294	-0.167	-0.073	-0.135	-0.054	+0.066	1
Voting	R-1	-	-0.561	-0.470	-0.295	-0.333	-0.103	-0.125	-0.029	+0.068	+0.138	2
	R-2	-	-1.002	-0.898	-0.669	-0.625	-0.409	-0.328	-0.238	-0.121	-0.102	0
	R-L	-	-0.579	-0.501	-0.345	-0.403	-0.193	-0.214	-0.110	-0.006	+0.046	1
Ranking	R-1	-	-3.078	-2.815	-2.806	-2.829	-2.564	-2.319	-2.192	-2.112	-1.913	0
	R-2	-	-3.485	-3.271	-3.213	-3.247	-3.027	-2.729	-2.625	-2.567	-2.402	0
	R-L	-	-2.735	-2.486	-2.508	-2.551	-2.356	-2.093	-1.992	-1.890	-1.742	0
Stacking	R-1	-0.152	-0.093	-0.076	-0.063	-0.136	-0.202	-0.181	-0.084	-0.208	-0.184	0
	R-2	-0.416	-0.307	-0.285	-0.226	-0.331	-0.322	-0.352	-0.295	-0.383	-0.385	0
	R-L	-0.208	-0.095	-0.113	-0.078	-0.157	-0.212	-0.205	-0.122	-0.248	-0.196	0
Total		0	0	0	0	0	1	4	4	5	9	

Table 2: Performance differential in macro-averaged ROUGE recall scores w.r.t NMF-TR for different ensemble sizes and combining methods. ℓ is the size of ensemble. ‘-’ indicates that the integration method is not meaningful.

DUC2001												
		$\ell = 2$	$\ell = 4$	$\ell = 6$	$\ell = 8$	$\ell = 10$	$\ell = 20$	$\ell = 30$	$\ell = 40$	$\ell = 50$	$\ell = 100$	Total
Average	R-1	-0.328	-0.158	-0.011	+0.066	+0.054	+0.052	+0.177	+0.105	+0.070	+0.135	7
	R-2	-0.076	+0.152	+0.337	+0.324	+0.294	+0.257	+0.338	+0.313	+0.295	+0.341	9
	R-L	-0.118	+0.003	+0.154	+0.208	+0.163	+0.158	+0.255	+0.187	+0.151	+0.191	9
Median	R-1	-	-0.121	+0.037	+0.104	+0.068	+0.043	+0.047	+0.115	+0.180	+0.210	8
	R-2	-	+0.145	+0.301	+0.302	+0.301	+0.252	+0.224	+0.288	+0.340	+0.315	9
	R-L	-	-0.002	+0.188	+0.188	+0.140	+0.112	+0.108	+0.196	+0.230	+0.232	8
Quartile	R-1	-	-0.391	+0.062	-0.168	-0.087	-0.101	-0.001	+0.005	-0.004	+0.073	3
	R-2	-	-0.014	+0.274	+0.205	+0.199	+0.223	+0.226	+0.266	+0.255	+0.273	8
	R-L	-	-0.205	+0.243	+0.003	+0.033	+0.055	+0.100	+0.124	+0.120	+0.150	8
Voting	R-1	-	-0.464	-0.225	-0.093	+0.008	-0.011	+0.007	+0.046	-0.026	+0.009	4
	R-2	-	-0.290	+0.005	+0.210	+0.268	+0.290	+0.360	+0.379	+0.328	+0.428	8
	R-L	-	-0.192	+0.022	+0.130	+0.189	+0.152	+0.159	+0.202	+0.123	+0.184	8
Ranking	R-1	-	-2.299	-2.295	-2.180	-2.158	-1.768	-1.474	-1.607	-1.391	-1.052	0
	R-2	-	-1.659	-1.680	-1.495	-1.497	-1.164	-0.912	-1.032	-0.895	-0.662	0
	R-L	-	-1.619	-1.599	-1.508	-1.535	-1.152	-0.849	-0.945	-0.812	-0.522	0
Stacking	R-1	-0.009	+0.015	-0.032	-0.021	+0.085	+0.092	+0.032	+0.161	+0.042	-0.034	6
	R-2	+0.216	+0.250	+0.228	+0.201	+0.333	+0.271	+0.300	+0.344	+0.311	+0.242	10
	R-L	+0.178	+0.164	+0.115	+0.163	+0.252	+0.217	+0.211	+0.279	+0.211	+0.127	10
Total		2	6	12	12	14	13	14	15	13	14	

Table 3: Performance differential in macro-averaged ROUGE recall scores w.r.t NMF-TR for different ensemble sizes and combining methods. ℓ is the size of ensemble. ‘-’ indicates that the integration method is not meaningful.

3.2 Diversity in Base Models

Since summary scores are sensitive to the number of latent topics into which document is decomposed, varying the number of latent topics while decomposing the term-sentence matrix is a potential source of diversity in NMF base models. We explore two different ways to accomplish this.

Selecting latent topics from range: We create 100 base models with number of latent topics randomly chosen from the range $[r, 2r]$, where r is determined using method proposed by (Aliguliyev, 2009). We expect that random initialization and variation in the number of topics would inject diversity in base models. We do not test this method with *stacking* because it requires stacked matrices to have same number of columns. Results for this experiment (Table 4) belie our expectation. Thus *varying the number of topics does not improve the*

	Avg	Med	Quart	Vote	Rank
R-1	-0.005	+0.170	+0.041	+0.017	-1.745
R-2	-0.095	-0.035	-0.064	-0.126	-2.227
R-L	+0.002	+0.129	0.000	-0.053	-1.622

Table 4: Performance differential for random variation in number of latent topics in the range $[r, 2r]$ for ensemble size 100.

quality of consensus summary.

Varying latent topics over range: Suspecting that repetition in the number of latent topics in previous experiment curb diversity in the ensemble, we attempt to create diversity by generating base models with all values in the range $[r, 2r]$. Hence, the size of ensemble for this experiment is document specific. Here too, it is not possible to create a *stacking* ensemble because of different number of latent topics in each base model. Results for this diversity creation method are presented in Table 5.

Thus systematically varying the number of topics

	Avg	Med	Quart	Vote	Rank
R-1	-0.265	-0.540	+0.029	-0.397	-2.515
R-2	-0.363	-0.597	-0.268	-0.725	-2.935
R-L	-0.271	-0.566	+0.022	-0.503	-2.262

Table 5: Performance gain/loss when the number of latent topics are in generated from the range $[r, 2r]$.

also fails to infuse diversity and shows no promise.

3.3 Comparison with best summary

With no success in injecting diversity in base models, we proceed to perform deeper analysis to diagnose the cause of degradation. We wanted to answer the question ‘How many base models are responsible for pulling down the score of consensus summary?’.

To answer this, we evaluated all base summaries and noted their ROUGE scores. The score of the best base summary was compared against that of ensemble summary and translated to win (if ensemble summary score is higher or equal), and loss otherwise. This exercise was done for all integration methods and ensemble size 30 and 100.

Results shown in Table 6 are almost startling. E.g. 23/510 for ensemble size 30 means that out of 533 total documents, for 23 documents the ensemble summary was atleast as good as the best base summary. For 510 documents, ensemble summary score was worse than that of the base summary. Thus NMF ensemble summaries fail miserably to improve quality over the best base summary.

size = 30						
	Avg	Med	Quart	Vote	Rank	Stk
R-1	23/510	23/510	28/505	22/511	42/491	27/506
R-2	28/505	25/508	26/507	24/509	52/481	28/505
R-L	23/510	21/512	22/511	18/515	41/492	20/513
size = 100						
	Avg	Med	Quart	Vote	Rank	Stk
R-1	11/522	10/523	11/522	6/527	20/513	12/521
R-2	16/517	17/516	18/515	11/522	26/507	16/517
R-L	11/522	11/522	12/521	7/526	18/515	13/520

Table 6: Wins/losses for ensemble summary compared with best base model summary for DUC2002 corpus.

3.4 Statistical Significance of Combining methods

We investigate the statistical significance of our results for all combining methods. We employ bootstrap approach recommended in (Dror et al., 2018), and test the null hypothesis

H0: NMF ensemble method performs no worse than the baseline NMF-TR, against the alternative hypothesis,

H1: NMF ensemble method performs worse than NMF-TR.

For each combining method, we generate one million bootstrap samples from the ROUGE scores of 533 ensemble summaries. We compute the difference in performance w.r.t baseline and estimate p -value as the ratio of number of times ensemble method beats NMF-TR by twice the margin on the bootstrap samples, to the total number of samples. For p -value > 0.05 , we reject the null hypothesis. Table 7 shows p -values obtained for each ROUGE metric and combining method. According to the computed p -values (Table 7), we fail to accept null hypothesis for each combining method and each ROUGE metric. Therefore, NMF ensemble meth-

	Avg	Med	Quart	Vote	Rank	Stk
R-1	0.37	0.13	0.31	0.23	0.99	0.82
R-2	0.59	0.45	0.66	0.44	1	0.90
R-L	0.45	0.19	0.39	0.34	0.99	0.84

Table 7: Probability values for bootstrap sampling based test of ensemble performance w.r.t baseline.

ods are not statistically significantly better than the baseline method.

4 Discussion and Conclusion

Extensive empirical investigation shows that leveraging stochastic variations due to random initialization of NMF factor matrices for extractive document summarization is not straight-forward. We experimented with different NMF solvers available in (Pedregosa et al., 2011) and found no change in results. In absence of any concrete explanation for degraded performance, we forward two plausible reasons.

First, apparently simple combining methods fail to tease apart the differences in term-topic and topic-sentence strengths in the latent space. Possible future investigation in this direction include projection of these matrices in higher dimension, and drawing from the cluster ensemble research to design more sophisticated combining methods.

Second reason is related to the maxim of *sentence ranking and selection* for extractive document summarization. Combining scores from base models alters sentence ranking, and probably less important sentence get pulled up in the summary. This is most likely to happen with the lowest ranked sentence in the summary. A single *bad* sentence in the summary can lower down the score substantially. Achieving stable ranks in ensemble technology could be another direction of research.

References

- Ramiz M Aliguliyev. 2009. A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*, 36(4):7764–7772.
- Mark Belford, Brian Mac Namee, and Derek Greene. 2018. Stability of topic modeling via matrix factorization. *Expert Systems with Applications*, 91:159–169.
- Christos Boutsidis and Efstratios Gallopoulos. 2008. Svd based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362.
- Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. 2018. The hitchhiker’s guide to testing statistical significance in natural language processing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1383–1392.
- Derek Greene, Gerard Cagney, Nevan Krogan, and Pádraig Cunningham. 2008. Ensemble non-negative matrix factorization methods for clustering protein–protein interactions. *Bioinformatics*, 24(15):1722–1728.
- Alka Khurana and Vasudha Bhatnagar. 2019. Extractive document summarization using non-negative matrix factorization. In *International Conference on Database and Expert Systems Applications*, pages 76–90. Springer.
- L. I. Kuncheva and S. T. Hadjitodorov. 2004. Using diversity in cluster ensembles. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 2, pages 1214–1219 vol.2.
- Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788.
- Ju-Hong Lee, Sun Park, Chan-Min Ahn, and Daeho Kim. 2009. Automatic generic document summarization based on non-negative matrix factorization. *Information Processing & Management*, 45(1):20–34.
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Jipeng Qiang, Yun Li, Yunhao Yuan, and Wei Liu. 2018. Snapshot ensembles of non-negative matrix factorization for stability of topic modeling. *Applied Intelligence*, pages 1–13.
- Fariar Shahnaz, Michael W Berry, V Paul Pauca, and Robert J Plemmons. 2006. Document clustering using nonnegative matrix factorization. *Information Processing & Management*, 42(2):373–386.
- Hiroyuki Shinnou and Minoru Sasaki. 2007. Ensemble document clustering using weighted hypergraph generated by nmf. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 77–80.
- Sangho Suh, Jaegul Choo, Joonseok Lee, and Chandan K Reddy. 2017. Local topic discovery via boosted ensemble of nonnegative matrix factorization. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 4944–4948. AAAI Press.
- Zhi-Hua Zhou. 2012. *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC.