Simulated Annealing Based Algorithm for Tuning LDA Hyper Parameters



Nikhlesh Pathik and Pragya Shukla

Abstract LDA is a very popular unsupervised model used to find thematic information about the documents. The performance of the LDA greatly depends on its hyper parameter values. If its parameters are tuned properly then LDA may produce much better results. This paper mainly focused on finding good LDA configurations for finding improved LDA output as compare to the traditional LDA model. We have proposed and implemented SA-LDA algorithm that uses Simulated Annealing (SA) to find optimal values of LDA parameters. An empirical evaluation using customer review datasets from three different domains namely Mobile, Hotel and Movie is conducted. The experiment results show that SA-LDA gives better performance when it is evaluated by the coherence score.

Keywords Hyper parameters · LDA · Simulated annealing · Reviews

1 Introduction

Topic modeling is an unsupervised natural language processing approach that is used to get various topics from a text document for interpretation. Topic modeling is a method to understand, organize, and summarize huge collections of textual information and documents. Topic model works in unsupervised manner to learn topics on unlabeled documents. Basically, a document holds multiple topics in different proportions. The topics are a set of similar words that are produced by topic modeling techniques. Widely used topic modeling algorithms are Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) or Latent Semantic Indexing (LSI), and Hierarchical Dirichlet Process (HDP). The most frequently adopted topic modeling technique is LDA to extract topics from the textual data. LDA [1, 2] generates a

N. Pathik (✉) · P. Shukla
Institute of Engineering & Technology, Devi Ahilya University, Indore, India
e-mail: pathiknikhlesh@gmail.com

P. Shukla
e-mail: pragyashukla_iet@yahoo.co.in

© Springer Nature Singapore Pte Ltd. 2020

M. Pant et al. (eds.), *Soft Computing: Theories and Applications*,
Advances in Intelligent Systems and Computing 1154,
https://doi.org/10.1007/978-981-15-4032-5_47

probability distribution of a topic for each document. Corpus of documents is pre-processed and transformed into $d \times w$ document-word co-occurrence matrix $M1$, where w is number of unique terms in all the documents and d represents number of documents.

This $M1$ is input for LDA, which in turn will generate two matrices $M2$ and $M3$. $M2$ is $d \times t$ document- topic matrix, where d represents number of documents and t representing the number of topics. $M3$ represents $t \times w$ topic-word matrix where t represents number of topics and w representing the number of words.

In the last decade, LDA is extensively used for finding hidden topics from the text. The Dirichlet priors α, β has a great impact on the performance of LDA. If α, β are set to appropriate value, LDA output will significantly improve. This paper proposed and implemented SA-LDA for tuning hyper parameters for finding the best LDA configurations.

Rest of the paper is organized as follows: In Sect. 2, related research works are mentioned. Section 3 discussed the proposed algorithm. Section 4 describes the experimental setup and results. Section 5 concludes the paper.

2 Related Work

Recently different optimization approaches are used to tuned LDA parameters like Ant Colony Optimization (ACO), Genetic Algorithm (GA), Hill climbing, Simulated Annealing (SA), etc. We have summarized recent optimization research here.

Yarnguy and Kanarkard [3] proposed ACO-LDA method for tuning Latent Dirichlet Allocation (LDA) parameters to improve its performance. They have opted Ant Colony Optimization (ACO) to solve the computational problem by adding parameter tuning. Experiments have conducted on various datasets fetched from the UCI and evaluated on perplexity score. Dit et al. [4] have proposed the concept of LDA-GA and represent an experiment on TraceLab. They extend their approach for software engineering tasks towards a new IR method. A multiobjective optimization algorithm based on simulated annealing approach is suggested by Bandyopadhyay et al. [5]. Gultekin et al. [6] proposed an asymptotic version of simulated annealing for variational inference. Foulds and Smyth [7] introduced new Annealed importance sampling annealing paths a Monte Carlo integration technique. This estimates the relative performance of the models. This strategy can exhibit much lower empirical variance than earlier approaches. The described strategy computes the ratio of the possibility of two models rather than computing the possibility of each model separately. Elhaddad [8] has proposed a hybrid optimization technique by combining GA and SA. This algorithm has been tested on instances of symmetric Traveling Salesman Problem (TSP) from known TSPLIB.

A heuristic approach is proposed by Zhao et al. [9] to estimate the most appropriate number of topics for LDA. The rate of perplexity change (RPC) is calculated and determined by the change point of RPC to the appropriate number of topics. George and Doss [10] combined the Markov Chain Monte Carlo and sampling, for estimating

the hyper parameters. The choice of hyper parameters is completely on the Bayes approach, by putting a prior on the hyper parameter vector, which adds one layer to the hierarchical model. Gultekin et al. [11] have evaluated a stochastic annealing strategy for Bayesian posterior optimization with variational inference. Kuzmenko [12] presented a collapsed Gibbs sampling formula for the LDA model. The results they have shown allow the topics to have asymmetric Dirichlet priors on words and topics in LDA. Wallach et al. [13] explored various classes of structured priors for topic models and shown that an asymmetric Dirichlet prior has significant advantages over a symmetric prior. This prior structure approximate efficient hyper parameter optimization steps. Pathik and Shukla [14] explained various extended LDA models and hyper parameter values statistics along with justification for optimizing hyper parameter values.

From the above study, it can be concluded that Dirichlet priors α, β has a great impact on the performance of LDA. If they are set to appropriate values, LDA output will significantly improve. We are proposing SA-based approach for finding the best LDA configuration for producing high-quality output. SA-LDA optimizes the coherence of the topic produced by LDA for the considered input dataset.

3 Proposed Algorithm

SA is a random search approach which is developed in 1983. One of the reasons for popularity of SA is its simplicity. SA allows the transition to weaker solutions or we can say downward steps which make it more feasible for global optimal. Unlike hill-climbing type search which may stick in local optimal solution. SA can be considered as advanced hill climbing which allows downhill as well.

We have used Gensim implementation of LDA and optimized it using our developed SA algorithm. Hyper parameters α and β are tuned using SA and find out the optimal values so that LDA output will improve. Once the best values for α and β are obtained we optimize number of topic T corresponding to it. The reason for tackling T separately is that range of T is very different from α and β. Also dataset size affects T very much whereas α and β are not affected much.

3.1 SA-LDA Algorithm

1. Initialize and define hyper parameter ranges, number of iterations and control parameter. Hyper parameters are α and β, iterations i and Control Parameter is CP.
2. Generate a random population and its corresponding solution S (coherence value) based on the initial parameters value.
3. Generate a random neighboring population and calculate its corresponding solution.

4. Apply acceptance criteria on step 3
   (a) If $S_{new} > S_{old}$: move to new solution.
   (b) If $S_{new} < S_{old}$: move with acceptance probability if more than random no otherwise go to step 3.
5. Repeat 3–5 for a large number of iterations so that max search space is explored.
6. Once we got the best parameter value for α and β we again apply SA-LDA with α_{best} and β_{best} for optimizing number of topics also. After this step, we will get best LDA configuration in terms of parameters α, β, and T.

Figure 1 demonstrates the flow of the proposed algorithm.
Using Metropolis Criterion acceptance probability is defined as follows:

$$P(X_n) = e^{(dS/(i*S_{max,i}))}$$

where:

$P(X_n)$: Probability of acceptance of nth point X_n
dS: $S_{new} > S_{old}$, i.e. change in coherence value
$S_{max,i}$: Max value of coherence for up to ith iteration
i: number of iterations
CP: Control parameter CP $= i*S_{max}$.

S_{new} and S_{old} are Coherence values for LDA configuration. Performance of LDA will be evaluated based on this value. Coherence is our objective function.

Acceptance Criteria is a function of both the control parameter and objective function. As the value of the control parameter increases, the probability of acceptance decreases. The same is desired initially that the probability of acceptance should be higher so that majority of points should be accepted. It will initially explore solution space to maximum. As control parameter increases after every iteration and reduces the probability which helps to converge on some global optimal values.

4 Experimental Setup and Results

We have used the Gensim implementation of LDA and optimized it using our developed SA algorithm. Gensim provides various text processing libraries implemented in Python and customized it by implementing SA-LDA in Python. The experiment was conducted on Desktop Workstation with Intel Xeon CPU X5675 3.07 GHz with 16.0 GB RAM and 64 bit Operating System. Anaconda platform is used for Python programming.

We have considered three datasets from three different domains namely Hotel reviews, Product review, and Movie review. As our approach is unsupervised so we have only considered review text from datasets, i.e. we haven't used labeled dataset (Table 1).

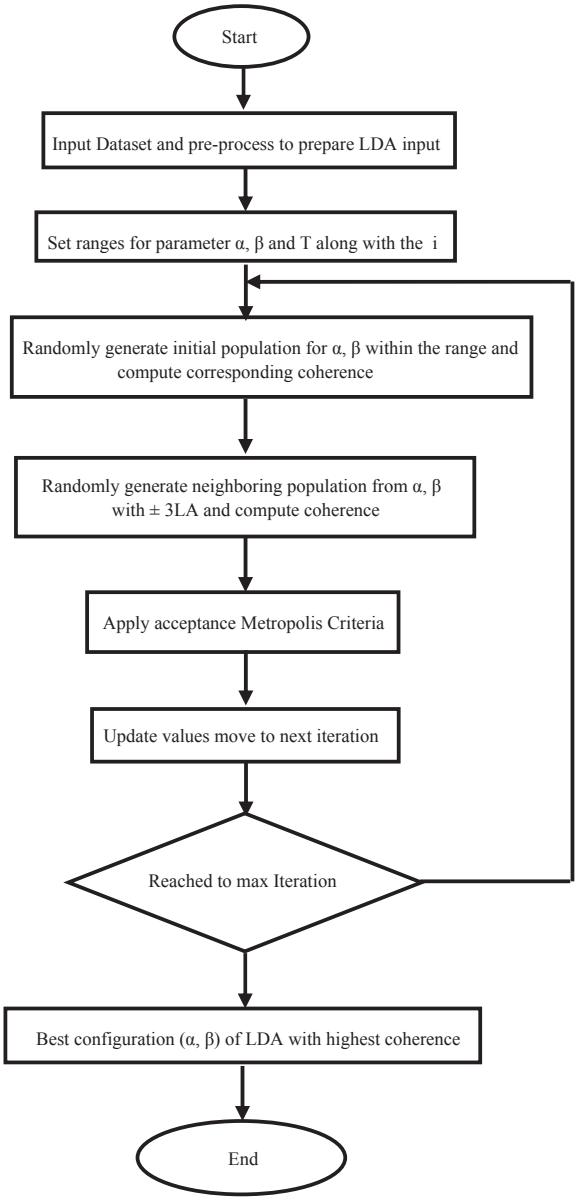


Fig. 1 Flow chart for SA-LDA algorithm

Table 1 Dataset statistics

Domain	Source of data	No of documents	No of sentences	No of words
Mobile reviews	Amazon	310	807	52,565
Hotel reviews	Yelp	3000	29,053	527,271
Movie reviews	IMDB	1000	10,872	272,803

Table 2 Pre-processing statics in term of number of words

Dataset	No of words	After removing punctuations	After removing stop words
Mobile	52,565	45,832	22,411
Hotel	527,271	460,297	200,127
Movie review	272,803	231,882	103,074

Text prepossessing reduces text input by half of its size. Or in text mining if free text is considered as input like text document then after pre-processing it will remain half in size. From Table 2 it is clear that pre-processing has a great impact on the aspect extraction step.

In most of the papers, number of topics are taken between 10 and 20 that is why we have used $T = 15$ for our SA-LDA for tuning α and β. After getting best values of α and β we have tried to find out optimal value of topics also by applying simple iterative process.

For the Hotel review dataset the value of hyper parameter was $\{\alpha, \beta, T\} = \{0.9, 0.69, 20\}$ and optimal value of coherence was 0.704. Same experiment is repeated with Mobile review dataset and we got values of hyper parameter $\{\alpha, \beta, T\} = \{0.79, 0.89, 22\}$ and optimal value of coherence was 0.665. For Movie review dataset $\{\alpha, \beta, T\} = \{0.60, 0.62\}$ and optimal value of coherence was 0.645.

We have compared with Gensim LDA implementation with autoconfiguration of α, β and found significant improvement in the value of coherence. We have also compared it with HDP-LDA and found our SA-LDA better in terms of coherence value. Table 3 shows comparative analysis and Fig. 2 shows a comparative graph for it.

It is clear from Fig. 2 SA-LDA is having better performance in comparison to LDA and HDP-LDA, when it is evaluated by coherence score.

Table 3 Comparison based on coherence value

Datasets	LDA	HDP-LDA	SA-LDA	α	β	T
Mobile review	0.367	0.649	0.665	0.79	0.89	22
Hotel review	0.505	0.641	0.704	0.90	0.69	20
Movie review	0.249	0.644	0.654	0.60	0.62	22



Fig. 2 Comparative Graphs for LDA, HDP-LDA, and SA-LDA

5 Conclusion

This paper proposed SA-LDA algorithm for finding hidden topics from the review datasets. The different datasets (Hotel, Mobile, and Movie) were taken. The performance of LDA is improved by tuning hyper α and β. Simulated Annealing based optimization gives best LDA configuration that produces most coherent topics. For evaluation of our SA-LDA algorithm, we have compared coherence of LDA, HDP-LDA with it for all three datasets. The comparison results show that SA-LDA gives better performance as compared to LDA and HDP-LDA in terms of coherence score.

References

1. Blei, D., Carin, L., Dunson, D.: Probabilistic topic models. IEEE Signal Process. Mag. **27**(6), 55–65 (2010)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**(1), 993–1022 (2003)
3. Yarnguy, T., Kanarkard, W.: Tuning latent Dirichlet allocation parameters using ant colony optimization. J. Telecommun. Electron. Comput. Eng. (JTEC) **10**(1–9), 21–4 (2018)
4. Dit, B., Panichella, A., Moritz, E., Oliveto, R., Di Penta, M., Poshyvanyk, D., De Lucia, A.: Configuring topic models for software engineering tasks in Tracelab. In: 2013 7th International Workshop on Traceability in Emerging Forms of Software Engineering (TEFSE), IEEE, pp. 105–109 (2013)
5. Bandyopadhyay, S., Saha, S., Maulik, U., Deb, K.: A simulated annealing-based multiobjective optimization algorithm: AMOSA. IEEE Trans. Evol. Comput. **12**(3), 269–283 (2008)
6. Gultekin, S., Zhang, A., Paisley, J.: Asymptotic simulated annealing for variational inference. In: IEEE Global Communications Conference (GLOBECOM), IEEE, pp. 1–7 (2018)
7. Foulds, J.R., Smyth, P.: Annealing Paths for the Evaluation of Topic Models. UAI, pp. 220–229 (2014)
8. Elhaddad, Y.R.: Combined simulated annealing and genetic algorithm to solve optimization problems. World Acad. Sci. Eng. Technol. **68**, 1508–1510 (2012)
9. Zhao, W., Chen, J.J., Perkins, R., Liu, Z., Ge, W., Ding, Y., Zou, W.: A Heuristic approach to determine an appropriate number of topics in topic modeling. In: BMC Bioinformatics, vol. 16, no. 13, p. S8. BioMed Central (2015)
10. George, C.P., Doss, H.: Principled selection of hyper-parameters in the latent Dirichlet allocation model. J. Mach. Learn. Res. **18**, 162 (2017)
11. Gultekin, S., Zhang, A., Paisley, J.: Stochastic annealing for variational inference. arXiv:1505.06723 (2015)
12. Kuzmenko, A.: Gibbs sampling for LDA with asymmetric Dirichlet priors. akuz.me/wp-content/uploads/2014/01/akuz_lda_asym.pdf (2011)
13. Wallach, H.M., Mimno, D.M., McCallum, A.: Rethinking LDA: Why priors matter. In: Advances in Neural Information Processing Systems, pp. 1973–1981 (2009)
14. Pathik, N., Shukla, P.: An Ample analysis on extended LDA models for aspect based review analysis. Int. J. Comput. Sci. Appl. **14**(2), 97–106 (2017)