International Journal of Electrical and Computer Engineering (IJECE)
Vol. 8, No. 5, October 2018, pp. 3204~3213
ISSN: 2088-8708, DOI: 10.11591/ijece.v8i5.pp3204-3213

Optimisation towards Latent Dirichlet Allocation: Its Topic Number and Collapsed Gibbs Sampling Inference Process

Bambang Subeno[1], Retno Kusumaningrum[2], Farikhin[3]
[1]Information System, Universitas Diponegoro, Indonesia
[2]Department of Informatics, Universitas Diponegoro, Indonesia
[3]Department of Mathematics, Universitas Diponegoro, Indonesia

Article Info

Article history:

Received Nov 20, 2017
Revised Jan 19, 2018
Accepted Aug 23, 2018

Keyword:

Latent Dirichlet allocation
Likelihood
Minimum description length
Number of topics
Optimisation

ABSTRACT

Latent Dirichlet Allocation (LDA) is a probability model for grouping hidden topics in documents by the number of predefined topics. If conducted incorrectly, determining the amount of K topics will result in limited word correlation with topics. Too large or too small number of K topics causes inaccuracies in grouping topics in the formation of training models. This study aims to determine the optimal number of corpus topics in the LDA method using the maximum likelihood and Minimum Description Length (MDL) approach. The experimental process uses Indonesian news articles with the number of documents at 25, 50, 90, and 600; in each document, the numbers of words are 3898, 7760, 13005, and 4365. The results show that the maximum likelihood and MDL approach result in the same number of optimal topics. The optimal number of topics is influenced by alpha and beta parameters. In addition, the number of documents does not affect the computation times but the number of words does. Computational times for each of those datasets are 2.9721, 6.49637, 13.2967, and 3.7152 seconds. The optimisation model has resulted in many LDA topics as a classification model. This experiment shows that the highest average accuracy is 61% with alpha 0.1 and beta 0.001.

Corresponding Author:

Retno Kusumaningrum,
Department of Informatics,
Universitas Diponegoro,
Jl. Prof. Soedarto, SH, Tembalang, Semarang, 50275, Indonesia.
Email: retno@live.undip.ac.id

1. INTRODUCTION

Nowadays, text mining is widely implemented due to a wide variety of text types, such as news articles, scientific articles, books, email messages, etc. Furthermore, it encourages an increased need to extract the information contained in a document. Furthermore, it encourages an increased need to extract the information contained in a document to generate useful knowledge [1], [2], [3], [4]. The difference between news articles or textual articles disseminated through electronic media with other documents is the model of information flow. The news flow is a dynamic and continuously updated stream; the more the news article in electronic media is, the more extensive the data collection as it always increases [5]. With enormous data variations, problems occur when needing to take on the different news while having the same theme. So, to facilitate navigation, news articles must be grouped by the same topic.

One way to get the topic information contained in the corpus of a news article document is to use topic modelling. Latent Dirichlet Allocation (LDA) is a topic modelling technique that can group words into specific topics from various materials [6]. The number of topics contained in the corpus with multiple variations is necessary to optimise the number of topics listed within the corpus. There are several estimation algorithms used in LDA including Expectation-Maximization algorithm [6], Expectation-Propagation algorithm to obtain better accuracy [7], as well as Collapsed Gibbs Sampling [8]. EM variations require high computation and learning models to be biased and inaccurate. Also, all of these algorithms and the number of topics should be set beforehand.

Determining the number of K topics is very important in LDA. Incorrectly identifying the number of K topics can result in limited word correlation with the topic [9]. Too large or too small number of the topic will affect the inference process and cause inaccuracies in grouping topics in the training model [10]. The use of Bayesian nonparametric methods, such as Hierarchial Dirichlet Process (HDP) in determining the number of topics, experienced bottlenecks during high computation [11]. The use of stochastic variational inference and parallel sampling is not consistent with the determination of the number of topics in the LDA model [12].

In this study, we optimise the number of topic LDA using maximum likelihood and Minimum Description Length (MDL) towards the usage Indonesian news articles. Basically, LDA Collapsed Gibbs Sampling (CGS) runs based on the number of documents [13], [14], [15], so that the reports dramatically affects the computation time. In this study, the number of documents does not affect the computation time, while the number of words greatly affects the computing time. To obtain the optimal number of topic K based on likelihood, LDA CGS will run from the smallest amount of K to the most significant number of K. For each K, we will calculate log-likelihood value and perplexity with specific iteration. The iteration will stop itself if perplexity value convergences. The optimal number of the topic will automatically be obtained based on the maximum log-likelihood value of the K range. For MDL as opposed to likelihood, LDA CGS will run from maximum number of K to minimum number of K. The smallest MDL value of the K range represents the optimal number of topics.

2. RESEARCH METHOD

This section discusses the implementation of likelihood and MDL to find the optimal number of topic LDA. The process of optimising the number of topic LDA is a one-time execution. The optimisation process stages are documented with their input, pre-processing, Bag of Word (BoW), determining the maximum number of topic K, and optimising number of topic. The process of optimising the number of topic LDA can be seen in Figure 1.

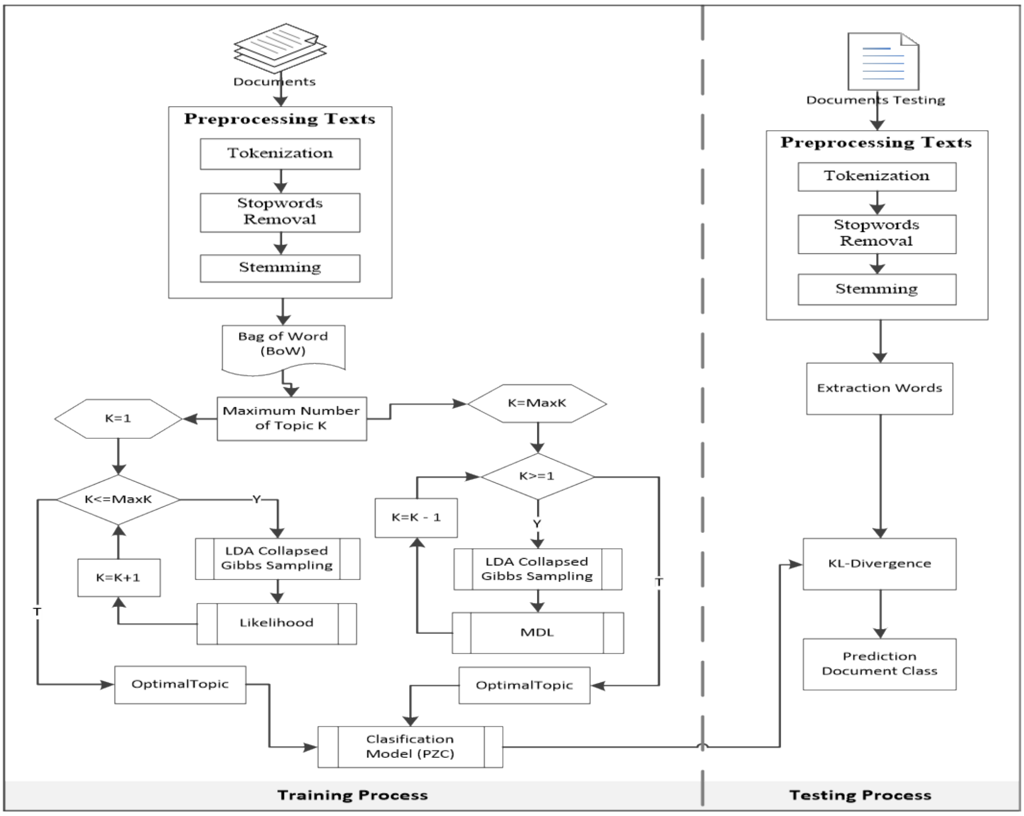

Figure 1. Process of optimisation number of topic LDA

2.1. Maximum Number of Topic

Bag of Word (BoW) pre-processing results still come in random data, which can be made into group data. Lists containing grouped data by a specific interval class or by a particular category are called frequency distribution [16], [17]. The formula for calculating the number of groups is as follows [16], [17]:

$$K = 1 + 3.322\, log_{10}(\, N\,) \approx 1 + log_2\, (N) \tag{1}$$

Where N is the number of data. For example, the resulted words are *"makan"*, *"jeruk"*, *"mangga"*, *"beli"*, *"jeruk"*, *"apel"*, *"tarif"*, *"sopir"*, *"angkut"*, *"mahal"*, *"bbm"*, *"naik"*, *"bbm"*, *"solar"*, *and "mahal"*. Based on equation 1, the data can be grouped into 4 or 5 groups.

2.2. LDA Collapsed Gibbs Sampling

Latent Dirichlet Allocation is a topic modelling technique that describes the probability procedure of document [6]. Applying topic modelling to a document will be able to produce a set of low-dimensional polynomial distributions called topic. Each topic will be used to combine some information from documents that have the same word relationship. The resulted topic can be extracted into a semantic structure with comprehensive results, even in large data [18], [19].

LDA model is a probability model that can explain the correlation between words with hidden topics in the document, find topics, and summarize text documents [20]. The main idea of topic modelling assumes that each document can be represented as a distribution of several topics whereby each topic is the probability distribution of the words [21]. The development of LDA method used today is LDA as a generative model and LDA as inference model, which can be seen in Figure 2 [22]. Pseudo code of CGS Standard, Pseudo code of Efficient CGS-Shortcut, Pseudo code of Collapsed Gibbs Sampling (CGS) optimisation [13] as shown in Figure 3,4,5.

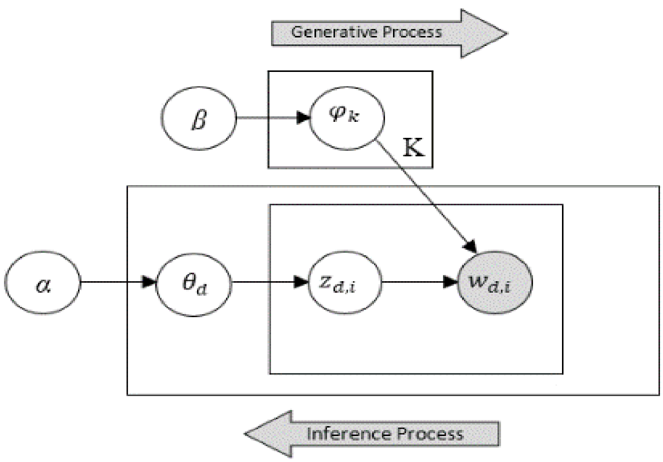

Figure 2. LDA representation model

LDA as a generative model is used to generate a document based on the probability value of word topic (φk) and proportion topic of document (θd). LDA as an inference model using Collapsed Gibbs Sampling (CGS) is the reverse of generative process as it aims to determine or find hidden value variables, i.e., probability word topic (φk) and proportion topic of documents (θd) from the predefined observation data [22]. In CGS processes, every word in the document will be determined at random at the beginning of the topic. Then, each word will be processed to determine a new topic based on the probability value of each topic. To calculate the probability value, the following formula is used [14]:

$$p(z_i = k|\vec{z}_{-i}, w) = \frac{n_{k,-i}^{(w)} + \beta}{n_{k,-i}^{(\cdot)} + (V * \beta)} * \left(n_{d,-i}^{(k)} + \alpha\right) \tag{2}$$

Where V is number of vocabulary; $n_{k,-i}^{(w)}$ is the number of words w on topic k, except token i; $n_{d,-i}^{(k)}$ is the number of words in document d specified as topic k, except token i; and $n_{k,-i}^{(.)}$ is the total word on topic k, except the token i. To determine the probability words topic and proportion topic of the document after going through the Gibbs Sampling process, the following formula is used [22]:

$$\varphi_{k,t} = PWZ = \frac{n_k^{(t)} + \beta_t}{\sum_{t=1}^{V}(n_k^{(t)} + \beta_t)} \tag{3}$$

$$\theta_{d,k} = PZD = \frac{n_d^{(k)} + \alpha_k}{\sum_{k=1}^{K}(n_d^{(k)} + \alpha_k)} \tag{4}$$

```
for ( d=1 to D ) do
  for ( i=1 to N_d) do
    v ← w_di , I_di ← N_di
    for ( j = 1 to I_di ) do
      k = z_dij
      N_wk ← N_wk − 1, NN_dk ← N_dk − 1
      for ( k = 1 to K ) do
        pk = (N_wk + β) x (N_dk + α) / (∑_v N_vk + Vβ)
      x ~ uniform (0, pk)
      k ← binarySearch(k: p_{k−1} < x < p_k)
      N_wk ← N_wk + 1, NN_dk ← N_dk + 1
      z_dij = k
```

Figure 3. Pseudo code of CGS Standard [13]

```
for ( d=1 to D ) do
  for ( i=1 to N_d) do
    v ← w_di , k = z_di
    N_wk ← N_wk − 1, N_dk ← N_dk − 1
    for ( k = 1 to K ) do
      pk = (N_wk + β) x (N_dk + α) / (∑_v N_vk + Vβ)
    x ~ uniform (0, pk)
    k ← binarySearch(k: p_{k−1} < x < p_k)
    N_wk ← N_wk + 1, NN_dk ← N_dk + 1
    z_dij = k
```

Figure 4. Pseudo code of Efficient CGS-Shortcut [13]

```
for ( i=1 to N_d) do
  v ← w_di , old_k ← z_di
  for ( k = 1 to K ) do
    if ( k = z_di) then
      N_wk ← N_wk − 1, N_dk ← N_dk − 1
    pk = (N_wk + β) x (N_dk + α) / (∑_v N_vk + Vβ)
    new_k ← index_k (max(p_k ))
    z_di = k
    if (old_k = new_k) then
      N_wk ← N_wk + 1, N_dk ← N_dk + 1
```

Figure 5. Pseudo code of Collapsed Gibbs Sampling (CGS) optimisation

2.3. Likelihood

Maximum Likelihood is the estimated standard used to determine the point estimation of an unknown parameter of a probability distribution with maximum probability. Pseudo code of likelihood standard, and pseudo code of likelihood optimisation as shown in Figure 6 and Figure 7. The estimation obtained by the likelihood maximum method is called likelihood maximum estimate [23]. There are several likelihood sample models developed for estimation on topic modelling such as Importance Sampling, Harmonic Mean, Mean Field Approximation, Left-to-Right Samplers, Left-to-Right Participant Samplers, Left-to-Right Sequential Samplers [24]. The log-likelihood function on topic LDA modelling is as follows [14]:

$$p(w_d|M) = \sum_{t=1}^{V} n_d^{(t)} \log(\sum_{k=1}^{K} \varphi_{k,t} \, . \, \theta_{d,k}) \tag{5}$$

```
for ( v=1 to V ) do
  for ( d=1 to D ) do
    for ( k = 1 to K ) do
      // calculate_matrix
      C_{v,d} = C_{v,d} + ( φ_{v,k} x θ_{d,k} )

    Loglik = N_{d,t} x log(C_{v,d})
    sumLoglik = sumLoglik + Loglik
```

Figure 6. Pseudo code of Likelihood standard

```
B ← newBoW
foreach B ∈ { i=1 to N_d }
  d ← B_{index_doc}
  v ← w_{di}
  for ( k = 1 to K ) do
    // calculate_matrix
    C_{v,d} = C_{v,d} + (φ_{v,k} x θ_{d,k})

  Loglik = N_{d,t} x log(C_{v,d})
  sumLoglik = sumLoglik + Loglik
```

Figure 7. Pseudo code of Likelihood optimisation

2.4. Minimum Description Length

Minimum Description Length (MDL) is a method used to optimize parameter estimation of a statistical distribution and model selection in a modelling process. In this MDL principle, the Bayesian theory is used to determine estimation by consideration of the likelihood data and existing knowledge of the prior probability [25]. Implementation of the MDL principle comes from the normalization of maximum likelihood to measure the model complexity of the data sets [26]. The formula for calculating the MDL is as follows [27]:

$$MDL = -\log\big(p(x|\theta)\big) + \tfrac{1}{2} L \log(NT) \,, \tag{6}$$
$$L = \frac{1}{100}\left(1 + T + \frac{T(T+1)}{2} - 1\right)$$

Where $\log\big(p(x|\theta)\big)$ is *log-likelihood* value, T is the number of topics used, and N is the number of words in the document.

2.5. Perplexity

Perplexity is another way to calculate the likelihood used to measure the performance of the LDA model. The smallest perplexity value is the best LDA model [14]. The formula for calculating the perplexity is as follows:

$$Perplexity = \exp\left\{-\frac{\sum_{d=1}^{D} \log p(w_d|M)}{\sum_{d=1}^{D} N_d}\right\} \tag{7}$$

Where D is the number of documents, $\log p(\, w_d|M)$ is log-likelihood according to the equation (5), and N is the number of words in the document.

3. RESULTS AND ANALYSIS

Section IV consists of three subsections, i.e., experiments set up, the scenario of experiments, experiments result, and analysis.

3.1. Experiments Set Up

In this study, we use Indonesian news articles from online portal of detik.com and Radar Semarang. The numbers of documents we use are 25, 50, 90, and 600 with the numbers of pre-processing words of each document are 3898, 7760, 13005, and 4365. Implementation of experiments use PHP programming language, MySQL database, and hardware specifications as follows:

a. Intel® Core™ i3 1.8GHz
b. 4 GB of memory
c. 500 GB of hard disk drive

The algorithms in Figure 4 and Figure 6 of the document looping process are omitted because document index information appears in BoW results. Optimisation process based on maximum likelihood and MDL once executed will automatically earn the optimal number of topic K, along with the value of perplexity, probability word topic, proportion topic for document, and probability topic of each class

3.2. Scenario of Experiments

Based on experiments set up, we perform four experimental scenarios using combinations of alpha 0.1, 0.001 and beta 0.1, 0.001. Scenario 1 aims to compare the execution time between standard algorithm and CGS optimisation, where we used several datasets for alpha 0.1 and beta 0.1. The datasets consist of a various number of documents, i.e., 25, 90, and 600. Scenario 2 aims to know the parameters that affect the time of optimisation of the number of topics. Scenario 3 aims to know the parameters that affect the optimal number of topics by using Likelihood and MDL. Scenario 4 aims to know the application of the resulted optimal number of the topic with LDA CGS as the classifying model.

LDA CGS implementation results in the optimal number of topics as a classification model. We use 100 articles divided into 90%, or 90 document articles as training data and 10%, or 10 article documents as testing data. The article document is divided into five classes: each class for training data consisting of 18 news articles. In the testing process, we use Kullback-Leibler Divergence (KLD) to measure the distribution similarity between the proportion of document testing topics and the proportion of topics for each class produced in the training process. The prediction of the document testing class is taken from the smallest value of KLD. Detailed information of KLD can be found in [22].

3.3. Experiments Result and Analysis

The results of the experimental scenario 1 can be seen in Figure 8, and Figure 9. While the results of the experimental scenario 2 can be seen in Table 1, Figure 10, and Figure 11. The results of the experimental scenario 3 can be seen in Table 2 and Figure 12. Furthermore, the result of experimental scenario 4 can be seen in Table 3 and Figure 13.

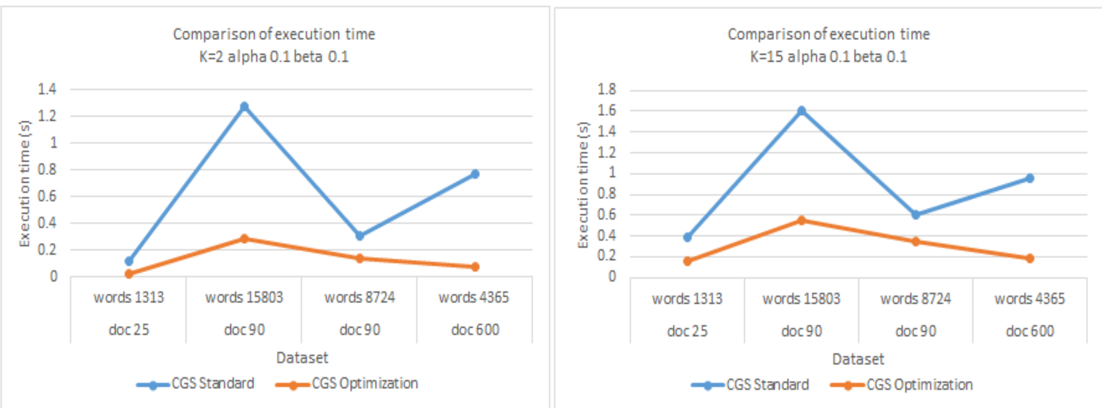

Figure 8. Comparison CGS Standard and Optimisation

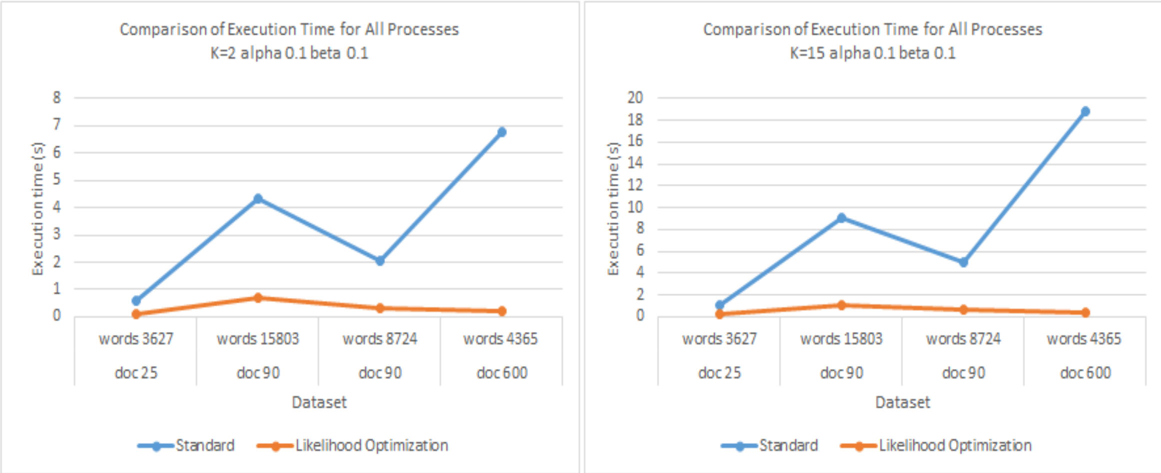

Figure 9. Comparison of Execution Time Standard and Optimisation for All Processes

Results in Table 1, Figure 10, and Figure 11 shows that the number of words used will affect the computational time: the greater the number of words is, the longer the computational time will increase. The number of documents and combinations of alpha, beta does not affect the computational time. The use of algorithms shown in Figure 5 and Figure 7 greatly concerns the optimisation of the execution time. Looping document is removed because the Bag of Word (BoW) pre-processing results show a document index. This is shown by the experimental results of the first scenario, which is illustrated in Figure 8 and Figure 9.

Table 1. Time Optimisation Process Result

No	Doc	Words	Alpha	Beta	Computing Time (second)	
					Likelihood	MDL
1	25	3898	0.1	0.1	2.97216	2.97216
2	25	3898	0.1	0.001	2.96717	2.96717
3	25	3898	0.001	0.1	2.95516	2.95516
4	25	3898	0.001	0.001	2.97816	2.97816
5	50	7760	0.1	0.1	6.496371	6.496371
6	50	7760	0.1	0.001	6.467370	6.467370
7	50	7760	0.001	0.1	6.476370	6.477377
8	50	7760	0.001	0.001	6.457369	6.458369
9	90	13005	0.1	0.1	13.29676	13.29676
10	90	13005	0.1	0.001	13.31676	13.31676
11	90	13005	0.001	0.1	13.30975	13.30975
12	90	13005	0.001	0.001	13.30476	13.30476
13	600	4365	0.1	0.1	3.715208	3.725208
14	600	4365	0.1	0.001	3.715212	3.715212
15	600	4365	0.001	0.1	3.716212	3.716212
16	600	4365	0.001	0.001	3.715212	3.715212

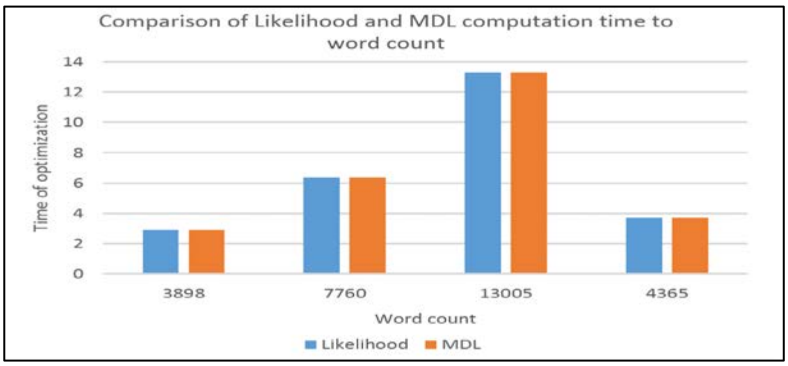

Figure 10. Comparison of Likelihood and MDL computation time to word count

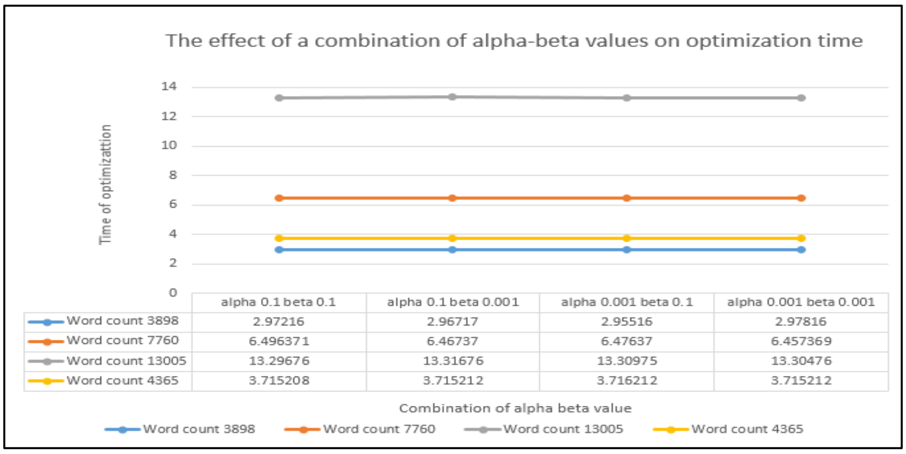

Figure 11. The effect of a combination of alpha-beta values on optimisation time

Based on the experimental results in Table 2 and Figure 10, hyper-parameter alpha, beta can affect the optimal number of topics on likelihood and MDL. Although the use of alpha, beta values may affect the number of topics, the Likelihood and MDL processes will result in the same optimal number of topics. Table 3 shows the result of LDA CGS implementation as a classification model using 10 fold. The highest accuracy of document classification is 0.80 or 80% with alpha 0.1 and beta 0.001.

Table 2. Optimal Number of Topics Based on Likelihood and MDL

No	Doc	Words	Alpha	Beta	Optimal Number of Topic	
					Likelihood	MDL
1	25	3898	0.1	0.1	11	11
2	25	3898	0.1	0.001	12	12
3	25	3898	0.001	0.1	13	13
4	25	3898	0.001	0.001	13	13
5	50	7760	0.1	0.1	13	13
6	50	7760	0.1	0.001	14	14
7	50	7760	0.001	0.1	14	14
8	50	7760	0.001	0.001	14	14
9	90	13005	0.1	0.1	15	15
10	90	13005	0.1	0.001	15	15
11	90	13005	0.001	0.1	15	15
12	90	13005	0.001	0.001	15	15
13	600	4365	0.1	0.1	12	12
14	600	4365	0.1	0.001	12	12
15	600	4365	0.001	0.1	13	13
16	600	4365	0.001	0.001	13	13

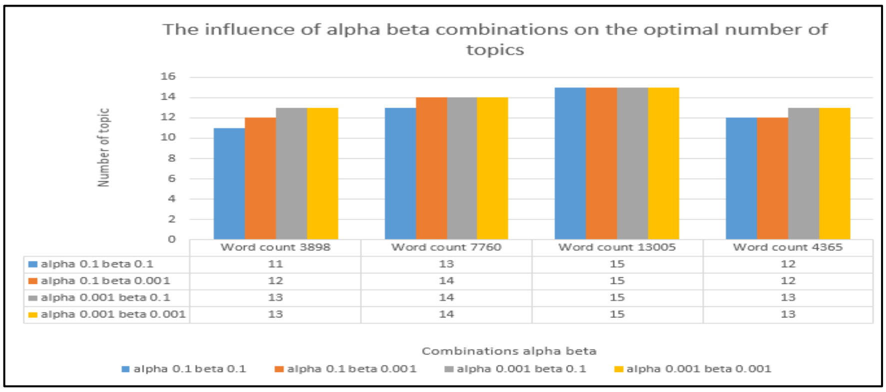

Figure 12. The influence of alpha, beta combinations on the optimal number of topics

Table 3. Average Accuracy Classification of Every Fold

Fold	Accuracy of Document Classification			
	Alpha 0.1 Beta 0.1	Alpha 0.1 Beta 0.001	Alpha 0.001 Beta 0.1	Alpha 0.001 Beta 0.001
1	60%	70%	40%	50%
2	60%	50%	50%	40%
3	50%	50%	60%	50%
4	50%	80%	50%	50%
5	40%	60%	40%	50%
6	50%	70%	40%	50%
7	50%	70%	50%	70%
8	50%	50%	30%	60%
9	50%	60%	40%	50%
10	50%	50%	50%	50%
Average	51%	61%	45%	52%

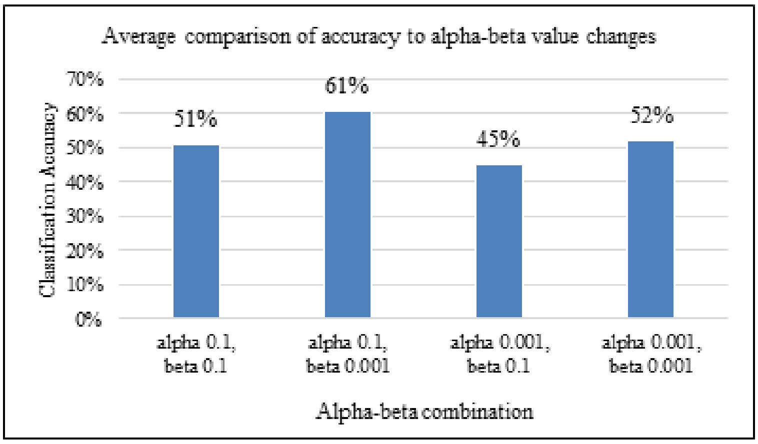

Figure 13. Average comparison of accuracy to alpha-beta value changes

Based on the experimental result in Table 3 and Figure 11, it is shown that the average highest classification accuracy of each fold is 61% with hyper-parameter alpha 0.1 and beta 0.001. The use of alpha and beta greatly affects the accuracy of document classification. The method of appropriate hyper-parameter alpha, beta will produce a high degree of accuracy as in fold 4 with 0.80 or 80% efficiency.

4. CONCLUSION

The optimisation number of topic with LDA, using Likelihood and MDL, yields the same optimal number of topic. The number of documents does not have a significant effect on the optimisation process, but the number of words does. The more number the words used, the longer the computational time was. Combination of alpha, beta values will conduct an effect on the optimal number of topic but does not give a significant effect on computational time.

Moreover, optimising the number of topics with LDA, we have gathered that CGS can be applied as a classification model, but to get good accuracy, one should do several iterations and use appropriate alpha, beta values. The incorrect use of alpha, beta values will affect the optimal number of topics, and the classification accuracy is not good. In this study, the highest mean value earned for 10-fold is 0.61 or 61% with alpha 0.1 and beta 0.001. The best classification accuracy is shown in fold 4 with 0.80 or 80% accuracy value.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the research funding supported by Universitas Diponegoro under the grant of research for international scientific publication-Year 2017 (number 276-36/UN7.5.1/PG/2017). This research funding is granted to the second author.

REFERENCES

[1] S. Moro, P. Cortez and P. Rita, "Business intelligence in banking: A literature analysis from 2002 to 2013 using Text Mining and Latent Dirichlet Allocation," *Expert Systems with Applications*, vol. 42, pp. 1314 - 1324, 2015.

[2] N. Naw and E. E. Hlaing, "Relevant Words Extraction Method for Recommendation System," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 3, pp. 169-176, 2013.

[3] N. Naw, "Relevant Words Extraction Method in Text Mining," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 3, pp. 177-181, 2013.

[4] R. S. A and S. Ramasamy, "Context Based Classification of Reviews Using Association Rule Mining, Fuzzy Logics and Ontology," *Bulletin of Electrical Engineering and Informatics*, vol. 6, no. 3, pp. 250-255, 2017.

[5] D. Bracewell, Y. Jiajun and R. Fuji, "Category Classification and Topic Discovery of Japanese and English News Articles," 2009.

[6] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research 3, pp. 993-1022, 2003.

[7] T. Minka and J. Lafferty, "Expectation-propagation for the generative aspect model," In UAI, p. 352–359, 2002.

[8] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proceeding of the National Academy of Sciences, vol. 101, pp. 5228 - 5235, 2004.

[9] A. Kulesza, N. R. Rao and S. Singh, "Low-Rank Spectral Learning," International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 33, 2014.

[10] J. Tang, Z. Meng, X. Nguyen and Q. Mei, "Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis," International Conference on Machine Learning, vol. 32, 2014.

[11] D. Cheng, X. He and Y. Liu, "Model Selection for Topic Models via Spectral Decomposition," International Conference on Artificial Intelligence and Statistics, vol. 38, 2015.

[12] S. Williamson, A. Dubey and E. P. Xing, "Parallel Markov Chain Monte Carlo for Parallel Markov Chain Monte Carlo," Journal of Machine Learning Research, vol. 28, pp. 98-106, 2013.

[13] T. S. Xiao Han, "Efficient Collapsed Gibbs Sampling For Latent Dirichlet Allocation," Asian Conference on Machine Learning (ACML2010), 2010.

[14] G. Heinrich, Parameter estimation for text analysis, 2.9 ed., Darmstadt, Germany: Fraunhofer IGD, 2009.

[15] R. K. V and K. Raghuveer, "Legal Documents Clustering and Summarization using Hierarchical Latent Dirichlet Allocation," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 2, no. 1, pp. 27-35, 2013.

[16] H. Sturges, "The choice of a class interval," Journal of the American Statistical Association, pp. 65-66, 1926.

[17] D. W. Scott, "Sturges Rule," Wires Computational Statistics, 2009.

[18] S. Arora, R. Ge and A. Moitra, "Learning Topic Models - Going beyond SVD," IEEE 53rd Annual Symposium on Foundations of Computer Science, vol. 2, pp. 1-10, 2012.

[19] Z. Liu, High Performance Latent Dirichlet Allocation for Text Mining, London: Brunel University, 2013.

[20] D. M. Blei, "Probabilistic Topic Models," Communication of The ACM, vol. 55, no. 4, pp. 77-84, 2012.

[21] Z. Qina, Y. Cong and T. Wan, "Topic modeling of Chinese language beyond a bag-of-words," Computer Speech and Language, vol. 40, pp. 60-78, 2016.

[22] R. Kusumaningrum, W. Hong, R. Manurung and M. Aniati, "Integrated Visual Vocabulary in LDA base scene clasification for IKONOS images," Journal of Appled Remote Sensing, vol. 8, 2014.

[23] I. J. Myung, "Tutorial on maximum likelihood estimation," Journal of Mathematical Psychology, vol. 47, pp. 90-100, 2002.

[24] W. Buntine, "Estimating Likelihoods for Topic Models," The 1st Asian Conference on Machine Learning, 2009.

[25] J. I. Myung, D. J. Navarro and M. A. Pitt, "Model selection by normalized maximum likelihood," Journal of Mathematical Psychology, vol. 50, pp. 167 - 179, 2005.

[26] D. W. Heck, M. Moshagen and E. Erdfelder, "Model selection by minimum description length: Lower-bound sample sizes for the Fisher information approximation," Journal of Mathematical Psychology, vol. 60, pp. 29-34, 2014.

[27] W. Xiaoru, D. Junping, W. Shuzhe and L. Fu, "Adaptive Region Clustering in LDA Framework for Image Segmentation," Proceedings of 2013 Chinese Intelligent Automation Conference, pp. 591-602, 2013.