OCTIS 2.0:
Optimizing and Comparing Topic Models in Italian Is Even Simpler!

Silvia Terragni and Elisabetta Fersini
University of Milano-Bicocca, Milan, Italy
`s.terragni4@campus.unimib.it, elisabetta.fersini@unimib.it`

Abstract

English. OCTIS is an open-source framework for training, evaluating and comparing Topic Models. This tool uses single-objective Bayesian Optimization (BO) to optimize the hyper-parameters of the models and thus guarantee a fairer comparison. Yet, a single-objective approach disregards that a user may want to simultaneously optimize multiple objectives. We therefore propose OCTIS 2.0: the extension of OCTIS that addresses the problem of estimating the optimal hyper-parameter configurations for a topic model using multi-objective BO. Moreover, we also release and integrate two pre-processed Italian datasets, which can be easily used as benchmarks for the Italian language.

Italiano. *OCTIS è un framework open-source per il training, la valutazione e la comparazione di Topic Models. Questo strumento utilizza l'ottimizzazione Bayesiana (BO) a singolo obiettivo per ottimizzare gli iperparametri dei modelli e quindi garantire una comparazione più equa. Tuttavia, questo approccio ignora che un utente potrebbe voler ottimizzare pi'u di un obiettivo. Proponiamo perciò OCTIS 2.0: l'estensione di OCTIS che affronta il problema della stima delle configurazioni ottimali degli iperparametri di un topic model usando la BO multi-obiettivo. In aggiunta, rilasciamo e integriamo anche due nuovi dataset in italiano pre-processati, che possono essere facilmente utilizzati come benchmark per la lingua italiana.*

1 Introduction

Topic models are statistical methods that aim to extract the hidden topics underlying a collection of documents (Blei et al., 2003; Blei, 2012; Boyd-Graber et al., 2017). Topics are often represented by sets of words that make sense together, e.g. the words "cat, animal, dog, mouse" may represent a topic about animals. Topic models' evaluations are usually limited to the comparison of models whose hyper-parameters are held fixed (Doan and Hoang, 2021; Terragni et al., 2020a; Terragni et al., 2020b). However, hyper-parameters can have an impressive impact on the models' performance and therefore fixing the hyper-parameters prevents the researchers from discovering the best topic model on the selected dataset.

Recently, OCTIS (Terragni et al., 2021a, Optimizing and Comparing Topic Models is Simple) has been released: a comprehensive and open-source framework for training, analyzing, and comparing topic models, over several datasets and evaluation metrics. OCTIS determines the optimal hyper-parameter configuration according to a Bayesian Optimization (BO) strategy (Archetti and Candelieri, 2019; Snoek et al., 2012; Galuzzi et al., 2020). The framework already provides several features and resources, among which at least 8 topic models, 4 categories of evaluation metrics, and 4 pre-processed datasets. However, the framework uses a single-objective Bayesian optimization approach, disregarding that a user may want to simultaneously optimize more than one objective (Terragni and Fersini, 2021). For example, a user may be interested in obtaining topics that are coherent but also diverse and separated from each other.

Contributions. In this paper, we propose OCTIS 2.0, an extension of the existing framework that integrates both a single-objective and multi-objective hyper-parameter optimization

strategy, using Bayesian optimization. Moreover, we also pre-process and include two novel datasets in Italian. We will then briefly show the potentiality of the extended framework by comparing different topic models on the new released Italian datasets. We believe these resources can be useful for the topic modeling and NLP communities, since they can be used as benchmarks for the Italian language.

2 OCTIS: Optimizing and Comparing Topic Models Is Simple!

2.1 OCTIS 1.0

OCTIS (Terragni et al., 2021a, Optimizing and Comparing is Simple!) is an open-source evaluation framework for the comparison of topic models, that allows a user to optimize the models' hyper-parameters for a fair experimental comparison. The evaluation framework is composed of different modules that interact with each other: (1) dataset and pre-processing tools, (2) topic modeling, (3) hyper-parameter optimization, (4) evaluation metrics. OCTIS can be used both as a python library and through a web dashboard. It also provides a set of pre-processed datasets, state-of-the-art topic models and several evaluation metrics.

We will now briefly describe the two components that we will extend in this work: the pre-processed datasets and the hyper-parameter optimization module.

Pre-processing and Datasets. OCTIS currently provides functionalities for pre-processing the texts, which include the lemmatization of the text, the removal of punctuation, numbers and stop-words, and the removal of words based on their frequency. Moreover, the framework already provides 4 pre-processed datasets, that are ready to use for topic modeling. These datasets are 20 NewsGroups,[1] M10 (Lim and Buntine, 2014), DBLP,[2] and BBC News (Greene and Cunningham, 2006). All the datasets are split into three partitions: training, testing and validation.

All the currently provided datasets are in English. OCTIS already provides language-specific pre-processing tools (e.g. lemmatizers for multiple languages), but it does not present datasets in other languages. Creating benchmark datasets for

other languages is useful for investigating the peculiarities of different topic modeling methods.

Single-Objective Hyper-parameter Optimization. OCTIS uses single-objective Bayesian Optimization (Snoek et al., 2012; Shahriari et al., 2015) to tune the topic models' hyper-parameters with respect to a selected evaluation metric. In particular, the user specifies the search space for the hyper-parameters and an objective metric. Then, BO sequentially explores the search space to determine the optimal hyper-parameter configuration. Since the models are usually probabilistic and can give different results with the same hyper-parameter configuration, the objective function is computed as the median of a given number of model runs (i.e., topic models run with the same hyper-parameter configuration) computed for the selected evaluation metric. OCTIS uses the Scikit-Optimize library (Head et al., 2018) for the implementation of the single-objective hyper-parameter Bayesian optimization.

The use of a single-objective approach is however limited. In fact, this strategy disregards other objectives. For example, a user may require to optimize the coherence of the topics and their diversity at the same time.

2.2 OCTIS 2.0

New dataset resources for the Italian language. Since OCTIS provides only English datasets, we extend the set of datasets by including two new datasets in Italian. We build the two datasets from the Italian version of the Europarl dataset[3] and from the Italian abstracts of DBPedia.[4] In particular, we randomly sample 5000 documents from Europarl and we randomly sample 1000 Italian abstracts for 5 DBpedia types (event, organization, place, person, work), for a total of 5000 abstracts.

We preprocess the datasets using the following strategy: we lemmatize the text, we remove the punctuation, numbers and Italian stop-words, we filter out the words with a document frequency higher than the 50% and less than the 0.1% for Europarl and 0.2% for DBPedia and we also remove the documents with less than 5 words. These values have been chosen by manually inspecting the resulting pre-processed datasets.

We report the most relevant statistics of the

---

[1] http://people.csail.mit.edu/jrennie/20Newsgroups/

[2] https://github.com/shiruipan/TriDNR/tree/master/data

[3] https://www.statmt.org/europarl/

[4] https://www.dbpedia.org/resources/ontology/

novel Italian datasets in Table 1. Following the original paper, we split the datasets in three partitions: training (75%), validation (15%), and testing (15%).

Dataset	Num. of documents	Avg. doc length (Std. dev.)	Num. of unique words
DBPedia	4251	5.5 (11.8)	2047
Europarl	3616	20.6 (19.3)	2000

Table 1: Statistics of the pre-processed datasets.

From Single-objective to Multi-objective Hyper-parameter Bayesian Optimization. Given the limitations of the single-objective hyperparameter optimization approach, we extend OCTIS by including a multi-objective approach (Kandasamy et al., 2020; Paria et al., 2019). Single-objective BO can be in fact generalized to multiple objective functions, where the final aim is to recover the Pareto frontier of the objective functions, i.e. the set of Pareto optimal points. A point is Pareto optimal if it cannot be improved in any of the objectives without degrading some other objective. Using a multi-objective hyper-parameter optimization approach thus allows us not only to identify the best performing model, but also to empirically discover competing objectives.

Since the original Scikit-Optimize library does not provide multi-objective optimization tools, we use the *dragonfly* library[5] (Paria et al., 2019). Like the single-objective optimization, the user must specify the hyper-parameter search space. But in addition, they also need to specify which functions they want to optimize. We report a simple coding example below:

```
# loading of a pre-processed dataset
dataset = Dataset()
dataset.fetch_dataset("DBPedia_IT")

#model instantiation
lda = LDA(num_topics=25)

#definition of the metrics to optimize
td = TopicDiversity()
coh = Coherence()
metrics = [td, coh]

#definition of the search space
config_file = "path/to/search/space/file"
```

---

[5]https://github.com/dragonfly/dragonfly

```
#define and launch optimization
mmm = MOOptimizer(
        dataset=dataset, model=model,
        config_file=config_file,
        metrics=metrics, maximize=True)
mmm.optimize()
```

The snippet will run a multi-objective optimization experiment that will return the Pareto front of the diversity and coherence metrics on the Italian dataset DBPedia by optimizing the hyperparameters (defined in a configuration file) of LDA with 25 topics.

In keeping with the spirit of the first version of OCTIS, the framework extension is open-source and easily accessible, in order to guarantee researchers and practitioners a fairer, accessible and reproducible comparison between the models (Bianchi and Hovy, 2021). OCTIS 2.0 is available as extension of the original library, at the following link: https://github.com/mind-Lab/octis.

3 Experimental Setting

In the following, we will show the capabilities of the extended framework on the new datasets by carrying out a simple experimental campaign.

We assume an experimental setting in which a topic modeling practitioner is interested in discovering the main thematic information of the two novel datasets in Italian. However, the user does not have prior knowledge on the datasets, therefore does not know which topic model is the most appropriate. Moreover, the user aims to get topics which are coherent and make sense together but which are also diverse and separated from the others. Let us notice that a user could consider a different set of metrics to optimize, by selecting one of the already defined metrics available in OCTIS or by defining novel metrics.

3.1 Evaluation Metrics

We briefly describe the two evaluation metrics (one of topic coherence and one of topic diversity) that we will target as the two objectives of the multi-objective Bayesian optimization. Both metrics need to be maximized.

IRBO (Bianchi et al., 2021a; Terragni et al., 2021b) is a measure of topic diversity (0 for identical topics and 1 for completely different topics). It is based on the Ranked-Biased Overlap measure (Webber et al., 2010). Topics with common

words at different rankings are penalized less than topics sharing the same words at the highest ranks.

NPMI (Lau et al., 2014) measures Normalized Pointwise Mutual Information of each pair of words (w_i, w_j) in the 10-top words of each topic. It is a topic coherence measure, that evaluates how much the words in a topic are related to each other.

3.2 Topic Models and Hyper-Parameter Setting

We focus our experiments on four well-known topic models that OCTIS already provides, two of them are considered classical topic models and the others are neural models. In particular, we trained Latent Dirichlet Allocation (Blei et al., 2003, LDA), Non-negative Matrix Factorization (Lee and Seung, 2000, NMF), Embedded Topic Model (Dieng et al., 2020, ETM), Contextualized Topic Models (Bianchi et al., 2021a; Bianchi et al., 2021b, CTM).

Model	Hyper-parameter	Values/Range
All	Number of topics	[5, 100]
LDA	α prior	$[10^{-3}, 10]$
	β prior	$[10^{-3}, 10]$
NMF	Regularization factor	[0, 0.5]
	L1-L2 ratio	[0,1]
	Initialization method	nndsvd, nndsvda, nndsvdar, random
	Regularization	V matrix, H matrix, both
ETM	Activation function	elu, sigmoid, softplus, selu
	Dropout	[0, 0.9]
	Learning rate	$[10^{-3}, 10^{-1}]$
	Number of neurons	$\{100, 200, \ldots, 900, 1000\}$
	Optimizer	adam, sgd, rmsprop
CTM	Activation function	elu, sigmoid, softplus, selu
	Dropout	[0, 0.9]
	Learning rate	$[10^{-3}, 10^{-1}]$
	Momentum	[0, 0.9]
	Number of layers	1, 2, 3, 4, 5
	Number of neurons	$\{100, 200, \ldots, 900, 1000\}$
	Optimizer	adam, sgd, rmsprop

Table 2: Hyper-parameters and ranges.

We summarize the models' hyper-parameters and their corresponding ranges in Table 2. For each model, we optimize the number of topics, ranging from 5 to 100 topics. We select the ranges of the hyper-parameters similarly to previous work (Terragni and Fersini, 2021).

Regarding LDA, we also optimize the hyper-parameters α and β priors that the sparsity of the topics in the documents and sparsity of the words in the topic distributions respectively. These hyper-parameters are set to range between 10^{-3} and 10^{-1} on a logarithmic scale.

The hyper-parameters of NMF are mainly related to the regularization applied to the factorized matrices. The *regularization* hyper-parameter controls if the regularization is applied only to the matrix V, or to the matrix H, or both. The *regularization factor* denotes the constant that multiplies the regularization terms. It ranges between 0 and 0.5 (0 means no regularization). *L1-L2 ratio* controls the ratio between L1 and L2-regularization. It ranges between 0 and 1, where 0 corresponds to L2 regularization only, 1 corresponds to L1 regularization only, otherwise it is a combination of the two types. We also optimize the *initialization method* for the two matrices W and H.

Since ETM and CTM are neural models, their hyper-parameters are mainly related to the network architecture. We optimize the *number of neurons* (ranging from 100 to 1000, with a step of 100). For simplicity, each layer has the same number of neurons. We also consider different variants of *activation functions* and *optimizers*. We set the *dropout* to range between 0 and 0.9 and the *learning rate*, that to range between 10^{-3} and 10^{-1}, on a logarithm scale. We fix the batch size to 200 and we adopted an early stopping criterion for determining the convergence of each model.

Moreover, only for CTM we also optimized the *momentum*, ranging between 0 and 0.9, and the number of layers (ranging from 1 to 5). Following (Bianchi et al., 2021b), we use the contextualized document representations derived from SentenceBERT (Reimers and Gurevych, 2019). In particular, we use the pre-trained multilingual Universal Sentence Encoder.[6]

For all the models, we set the remaining parameters to their default values. Finally, we train each model 30 times and consider the median of the 30 evaluations as the evaluation of the function to

---

[6]Let us notice that there is not a Sentence BERT-like model for Italian. Therefore we used a multilingual one: `distiluse-base-multilingual-cased-v1`.

be optimized. We sample the n initial configurations using the Latin Hypercube Sampling, with n equal to the number of hyperparameters to optimize plus 2 to provide enough configurations for the initial surrogate model to fit. The total number of BO iterations for each model is 125. We use Gaussian Process as the probabilistic surrogate model and the Upper Confidence Bound (UCB) as the acquisition function.

4 Results

In the following, we report the results of the comparative analysis between the considered models on the Italian datasets.
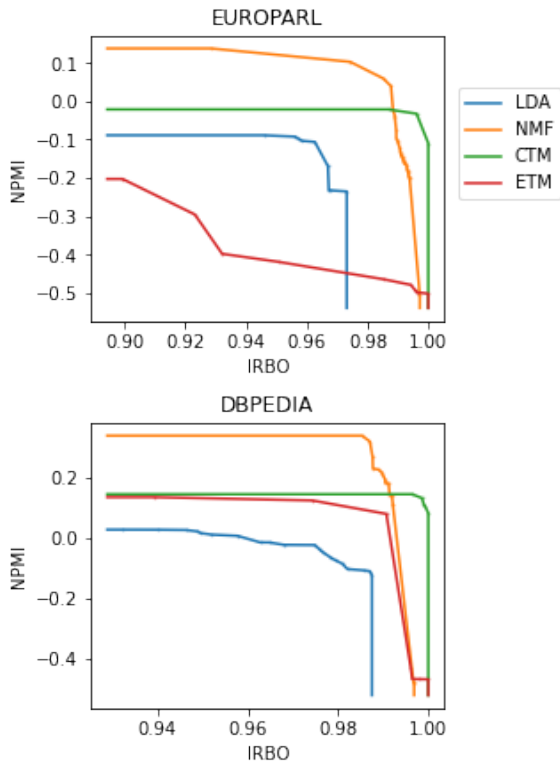
4.1 Quantitative Results

Figure 1: Pareto front of the performance of the considered models for the analyzed Italian datasets.

We jointly consider the results of both objectives by plotting the Pareto frontier of the results of topic diversity and topic coherence. Figure 1 shows the frontier of each model for the pair of metrics (NPMI, IRBO). We can notice that the topic models have similar frontiers in each dataset. The most competitive models are NMF and CTM. In particular, NMF outperforms the others for the

topic coherence but gets a lower coherence as the diversity increases. Therefore, CTM is the model to prefer if a user wants to get totally separated topics but good coherence. Instead, LDA and ETM have lower performance than the others. We also noticed from our experiments that the performance of ETM is affected when the documents are shorter (on the Europarl dataset), often originating the phenomenon of mode collapsing, i.e. obtaining all the topics equal to the others.

4.2 Qualitative Results

In Table 3 we report an example of topics discovered by the models. We selected the best hyperparameter configuration discovered by the models with 5 topics and randomly sampled a model run among the 30 runs. Let us notice that, for the sake of simplicity, we have to fix the number of topics here and select a run among the total of 30 runs. Therefore, the qualitative results reported in Table 3 may not reflect the overall results.

We can notice that NMF obtains more coherent and stable topics. CTM and LDA obtain topics that have a higher variance: in particular, CTM discovers a topic (the fourth one, NPMI=-0.51) that lowers the average coherence, while LDA discovers a topic (the second one, NPMI=0.48) that effectively increases the average coherence. On the other hand, the topics discovered by ETM are more stable but have a lower coherence on average. As already observed in previous work (Al-Sumait et al., 2009; Doogan and Buntine, 2021), obtaining junk or mixed topics is common in topic models and this problem can be addressed by filtering out the topics that are less relevant.

5 Conclusion

In this paper, we presented OCTIS 2.0, the extension of the evaluation framework OCTIS for topic modeling. This tool can now address the problem of estimating the optimal hyper-parameter configurations of different topic models using a multi-objective Bayesian optimization approach. Moreover, we also released two novel datasets in Italian which can be used as benchmark datasets for the Italian topic modeling and NLP communities.

We conducted a simple experimental campaign to show to potentiality of the extended framework. We have seen that using a multi-objective hyperparameter optimization approach allows us not only to identify the best performing model over the oth-

Model	Top words	NPMI
LDA	de album pubblicare italiano the uniti situare fondare università noto	-0.05
	torneo giocare tennis edizione tour atp ambito open categoria cemento	0.48
	film pubblicare the album serie musicale venire statunitense rock band	0.11
	guerra battaglia venire situare statunitense spagnolo partito esercito distretto mondiale	-0.14
	comune campionato squadra abitante calcio regione situare società francese vincere	-0.03
NMF	comune abitante dipartimento regione situare francese alta distretto est grand	0.29
	torneo giocare tennis tour atp open edizione ambito categoria cemento	0.48
	album pubblicare studio the musicale statunitense records singolo cantante rock	0.29
	calciatore ruolo allenatore calcio centrocampista difensore attaccante portiere settembre aprile	0.24
	contea america uniti situare comune censimento designated census place capoluogo	0.39
CTM	album the pubblicare band statunitense singolo brano of musicale rock	0.26
	superare argentino calciatore el buenos maria en svezia situare chiesa	-0.29
	partito battaglia guerra venire politico de linea isola stazione regno	-0.08
	st stella vendetta dollaro robert company ritorno west superiore soggetto	-0.51
	edizione tennis giocare torneo vincere tour campionato maschile disputare squadra	0.18
ETM	sede de italiano fondare nome azienda noto francese compagnia parigi	0.06
	guerra partito battaglia venire nord politico tedesco esercito regno militare	0.03
	torneo situare comune giocare abitante edizione tennis tour regione uniti	-0.10
	film serie the dirigere gioco pubblicare statunitense televisivo venire romanzo	0.07
	album pubblicare campionato squadra musicale the calcio statunitense singolo vincere	-0.12

Table 3: Example of top words of 5 topics for each considered model and the corresponding topic coherence (NPMI).

ers, thus guaranteeing a fairer comparison among different models, but also to empirically discover the relationships between different objectives.

As future work, we aim to extend the framework by considering additional datasets in different and possibly low-resource languages, which require different pre-processing strategies and would allow researchers to investigate the peculiarities of different topic modeling methods.

References

Loulwah AlSumait, Daniel Barbará, James Gentle, and Carlotta Domeniconi. 2009. Topic Significance Ranking of LDA Generative Models. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009*, volume 5781 of *Lecture Notes in Computer Science*, pages 67–82. Springer.

Francesco Archetti and Antonio Candelieri. 2019. *Bayesian Optimization and Data Science*. Springer International Publishing.

Federico Bianchi and Dirk Hovy. 2021. On the gap between adoption and understanding in nlp. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3895–3901.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021a. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021*, pages 759–766. Association for Computational Linguistics.

Federico Bianchi, Silvia Terragni, Dirk Hovy, Debora Nozza, and Elisabetta Fersini. 2021b. Cross-lingual contextualized topic models with zero-shot learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021*, pages 1676–1683. Association for Computational Linguistics.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

David M Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Jordan L. Boyd-Graber, Yuening Hu, and David M. Mimno. 2017. Applications of topic models. *Found. Trends Inf. Retr.*, 11(2-3):143–296.

Adji Bousso Dieng, Francisco J. R. Ruiz, and David M. Blei. 2020. Topic modeling in embedding spaces. *Trans. Assoc. Comput. Linguistics*, 8:439–453.

Thanh-Nam Doan and Tuan-Anh Hoang. 2021. Benchmarking neural topic models: An empirical study. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages

4363–4368, Online, August. Association for Computational Linguistics.

Caitlin Doogan and Wray L. Buntine. 2021. Topic model or topic twaddle? re-evaluating semantic interpretability measures. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 3824–3848. Association for Computational Linguistics.

Bruno Giovanni Galuzzi, Ilaria Giordani, Antonio Candelieri, Riccardo Perego, and Francesco Archetti. 2020. Hyperparameter optimization for recommender systems through bayesian optimization. *Computational Management Science*, pages 1–21.

Derek Greene and Pádraig Cunningham. 2006. Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. In *Proceedings of the 23rd International Conference on Machine learning (ICML'06)*, pages 377–384. ACM Press.

Tim Head, Gilles Louppe MechCoder, Iaroslav Shcherbatyi, et al. 2018. scikit-optimize/scikit-optimize: v0. 5.2.

Kirthevasan Kandasamy, Karun Raju Vysyaraju, Willie Neiswanger, Biswajit Paria, Christopher R. Collins, Jeff Schneider, Barnabás Póczos, and Eric P. Xing. 2020. Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly. *Journal of Machine Learning Research*, 21:81:1–81:27.

Jey Han Lau, David Newman, and Timothy Baldwin. 2014. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014*, pages 530–539.

Daniel D. Lee and H. Sebastian Seung. 2000. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000*, pages 556–562. MIT Press.

Kar Wai Lim and Wray L. Buntine. 2014. Bibliographic analysis with the citation network topic model. In *Proceedings of the Sixth Asian Conference on Machine Learning, ACML 2014*.

Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. 2019. A Flexible Framework for Multi-Objective Bayesian Optimization using Random Scalarizations. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 115 of *Proceedings of Machine Learning Research*, pages 766–776, Tel Aviv, Israel. AUAI Press.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, (EMNLP-IJCNLP)*, pages 3980–3990. Association for Computational Linguistics.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. 2015. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. 2012. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems*, pages 2960–2968.

Silvia Terragni and Elisabetta Fersini. 2021. An empirical analysis of topic models: Uncovering the relationships between hyperparameters, document length and performance measures. In *Recent Advances in Natural Language Processing (RANLP)*.

Silvia Terragni, Elisabetta Fersini, and Enza Messina. 2020a. Constrained relational topic models. *Information Sciences*, 512:581 – 594.

Silvia Terragni, Debora Nozza, Elisabetta Fersini, and Messina Enza. 2020b. Which matters most? comparing the impact of concept and document relationships in topic models. In *Proceedings of the First Workshop on Insights from Negative Results in NLP*, pages 32–40.

Silvia Terragni, Elisabetta Fersini, Bruno Giovanni Galuzzi, Pietro Tropeano, and Antonio Candelieri. 2021a. OCTIS: Comparing and Optimizing Topic models is Simple! In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, EACL 2021*, pages 263–270. Association for Computational Linguistics.

Silvia Terragni, Elisabetta Fersini, and Enza Messina. 2021b. Word embedding-based topic similarity measures. In *Natural Language Processing and Information Systems - 26th International Conference on Applications of Natural Language to Information Systems, NLDB 2021*, volume 12801 of *Lecture Notes in Computer Science*, pages 33–45. Springer.

William Webber, Alistair Moffat, and Justin Zobel. 2010. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.*, 28(4):20:1–20:38.