LADy 💃: A Benchmark Toolkit for Latent Aspect Detection Enriched with Backtranslation Augmentation

Farinam Hemmatizadeh
School of Computer Science
University of Windsor, Ontario, Canada
hemmatif@uwindsor.ca

Christine Wong
School of Computer Science
University of Windsor, Ontario, Canada
wong93@uwindsor.ca

Alice Yu
Vincent Massey Secondary School
Windsor, Ontario, Canada
alice.yu@uwindsor.ca

Hossein Fani
School of Computer Science
University of Windsor, Ontario, Canada
hfani@uwindsor.ca

ABSTRACT

We present LADy 💃, a Python-based benchmark toolkit to facilitate extracting aspects of products or services in reviews toward which customers target their opinions and sentiments. While there has been a significant increase in aspect-based sentiment analysis, yet the proposed methods' practical implications in real-world settings remain moot for their closed and irreproducible codebases, inability to accommodate datasets from various domains, and poor evaluation methodologies. LADy is an open-source benchmark toolkit with a standard pipeline and experimental details to fill the gaps. It incorporates a host of canonical models along with benchmark datasets from varying domains, including unsolicited online reviews. Leveraging an object-oriented design, LADy readily extends to new models and training datasets. The first of its kind, LADy also features review augmentation via natural language backtranslation that can be integrated into the training phase of the models to boost efficiency and improve efficacy during inference. LADy's codebase, along with the installation instructions and case studies on five datasets for seven methods with backtranslation augmentation over ten languages, can be obtained under `cc-by-nc-sa-4.0` license at `https://github.com/fani-lab/LADy`.

CCS CONCEPTS

• **Computing methodologies** → **Machine translation**; • **Information systems** → **Information extraction**.

KEYWORDS

Review Analysis; Aspect Detection; Backtranslation Augmentation;

ACM Reference Format:
Farinam Hemmatizadeh, Christine Wong, Alice Yu, and Hossein Fani. 2024. LADy 💃: A Benchmark Toolkit for Latent Aspect Detection Enriched with Backtranslation Augmentation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*

Table 1: Comparison of existing toolkits for aspect detection.

features	absa-pytorch [21]	pyabsa [31]	LADy 💃
maintenance	✗	✓	✓
open-source	✓	✓	✓
reproducible	✓	✓	✓
baseline models	static	static	any
datasets	static	format-specific	any
data augmentation	✗	static	any
evaluation metrics	static	static	any
latent aspect detection	✗	✗	✓

1 INTRODUCTION

Customers express opinions and sentiments about aspects of products or services, also referred to as features, properties, or components, in reviews. For example, in the review *"it's really cheap and you won't regret buying it."*, a customer expresses her opinion *'really cheap'* with *'positive'* sentiments toward the aspect *'price'* of a laptop. Aspects can be explicitly mentioned with a surface form, such as the aspect *'pizza'* in *"pizza was great."*, or they can be *latent* yet implicitly understood through common social knowledge like *'food'* in *"mine was a little burnt!"*. Aspect detection finds immediate application in aiding businesses in improving upon the quality of their offerings based on customers' reviews, thereby enhancing overall customer satisfaction.

While there has been a surge in aspect detection research [8, 10, 14, 16, 20, 24, 25], existing methods lack standardized implementations and experimental details, making it challenging to reproduce results consistently across various datasets and domains. For instance, most studies primarily rely on different versions of SemEval datasets [12, 25–29, 32], falling short of incorporating a new dataset, especially unsolicited reviews from online social media platforms, such as Twitter. This limitation becomes particularly critical when introducing data augmentation methods, which require experimental setup across existing and augmented datasets. Moreover, the reported baselines' results are either sourced from other papers or directly copied from the original papers [3, 23, 28, 33] with little to no reproducibility studies, which poses significant challenges for researchers who seek to adopt or improve such baselines. In terms
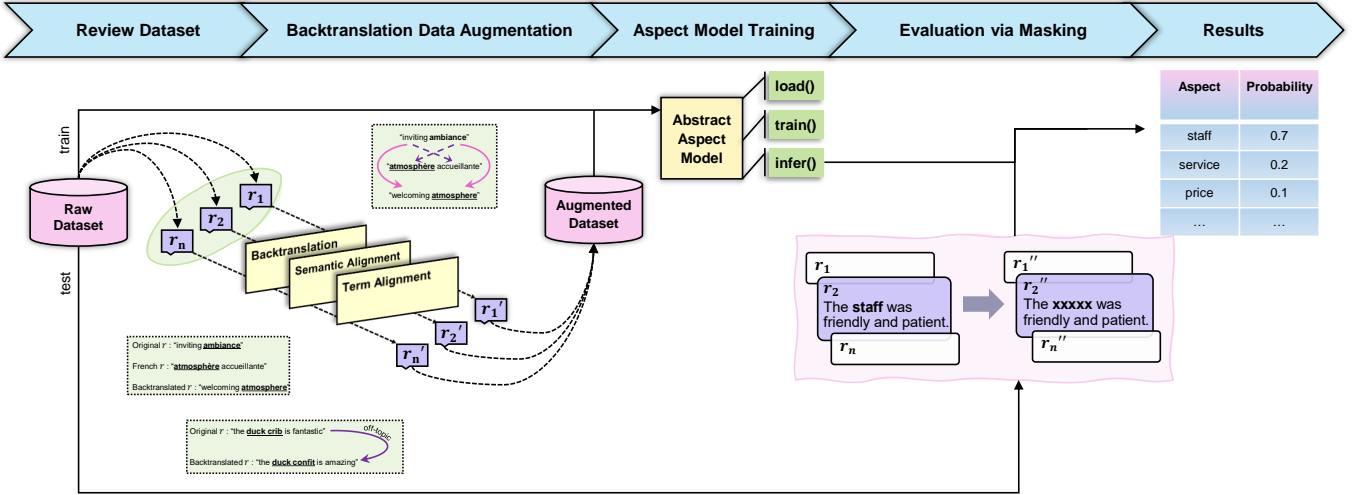
Figure 1: LADy's workflow: *i*) transforming raw reviews into review objects, *ii*) data augmentation via language backtranslation, *iii*) model training for implicit and explicit aspect detection, and *iv*) evaluation.

of evaluation, existing works predominantly focus on standard metrics such as precision, recall, and f-measure, which is insufficient when the aspect is latent and lacks a surface form in reviews. In such cases, an aspect extraction method should generate a ranked list of most likely aspects; therefore, integration of new evaluation metrics that consider ranking positions, e.g., precision@k or recall@k, becomes crucial, ensuring no need for rewriting the codes for the rest of the pipeline [9].

In response to these challenges, we propose LADy, a reproducible pipeline that accommodates new datasets and evaluation techniques to help researchers effectively compare and analyze results using different leading methods via a unified and accessible toolkit. LADy is designed to standardize and streamline the analysis of reviews, with a particular focus on the valuable but often overlooked unsolicited online reviews shared on platforms such as Google Reviews and Twitter. Unsolicited reviews are characterized by their informality and briefness, often containing textual ellipses, which require a specialized approach such as latent aspect detection. To this end, LADy offers a standardized implementation environment, benchmark datasets, adaptability to new metrics, an object-oriented structure, and open-source accessibility, making it a valuable toolkit for the natural language processing research community. It promotes reproducibility and fair model comparisons by providing an array of supervised and unsupervised aspect detection methods and empowers researchers to explore different evaluation approaches across variety of metrics. Last, LADy improves its aspect detection capabilities through backtranslation augmentation to expand the vocabulary and the diversity of detected aspects, and further, improve the extraction of latent aspects [9].

In the literature, limited open-source toolkits and libraries are developed for aspect-based sentiment analysis. Table 1 compares them with LADy in terms of various features, including reproducibility, latent aspect detection capability, evaluation strategies, and more. Also, the primary flow of LADy for detecting latent aspects via backtranslation augmentation is illustrated in Figure 1.

2 ASPECT DETECTION

Given a collection of reviews $\mathcal{R}$ and a set of aspects $\mathcal{A}$, we assume a review $r \in \mathcal{R}$ is about an aspect $a \in \mathcal{A}$ of a product or service, which can be explicitly mentioned like the aspect $a_1 =$ *'dessert.'* in the review $r_1 =$ *"the dessert was the perfect ending!"*, or can be *latent* like $a_2 =$ *'food'* in the review $r_2 =$ *"mine was a little burnt!"*, which lacks an explicit surface of the *latent* aspect a_2 like *'food'* in the review.

Existing aspect detection methods are primarily designed to find the explicit aspects falling short of detecting latent ones. In LADy, we address latent aspect detection by employing the backtranslation augmentation technique to enrich the training data with new words and help find the latent occurrences of aspects. During the data augmentation phase, we translate an input review r with its aspect a to another language l to generate the translated review r' followed by a backtranslation to the original language to generate a backtranslated review r'' with its new aspect a''. The generated collection of backtranslated reviews $\mathcal{R}''$ is then utilized along with the original collection of reviews $\mathcal{R}$ to empower the existing models to detect the latent aspects as well as the explicit ones.

3 SYSTEM DESIGN AND ARCHITECTURE

We present an overview of the LADy toolkit in Figure 1. As seen, LADy transforms raw reviews from diverse domains into review objects with attributes such as aos (aspect, opinion, and sentiment) and authorship details (Section 3.1). It leverages polymorphism to integrate and override functions such as training and inference within the AbstractAspectModel class (Section 3.3). Following this, LADy masks explicit aspect labels in reviews to simulate latent aspects, creating datasets to effectively evaluate its latent aspect detection capabilities (Section 3.4). Additionally, LADy applies data augmentation techniques via backtranslation across different languages to strengthen its capacity to recognize both explicit and latent aspects in reviews (Section 3.2). The subsequent sections provide details of each component and layer in LADy.
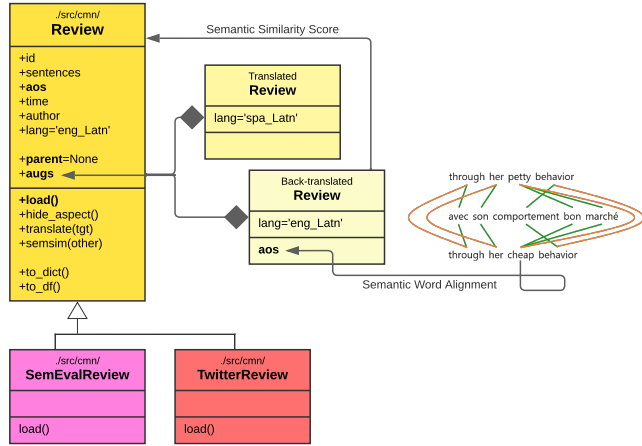
Figure 2: **LADy's** data layer class diagram and inheritance hierarchy.

```
# ./src/cmn/review.py
class Review(object):
    def __init__(self, id, sentences, time, author, aos, parent,
        lang):
        self.id = id
        self.sentences = sentences
        self.time = time
        self.author = author
        self.aos = aos
        self.lang = lang
        self.parent = parent
        self.augs = {}
    def hide_aspects(self, mask):
    def translate(self, tgt, settings): ...
    def semsim(self, other):
    def semalign(self, other): ...
    def preprocess(self): ...
    def load(path): ...
    def get_aos(self): ...
    def get_txt(self): ...
    def get_stats(datapath, output): ...
    def plot_dist(stats, output): ...
```

Figure 3: `Review` class attributes and methods.

3.1 Review Class

Our toolkit is built on a foundational layer known as the data layer, acting as the cornerstone of our architecture. This layer sets up the essential infrastructure for managing reviews by the Review class, as shown in Figure 2. Each review object includes an array of aos triplets, each representing an aspect, opinion, and sentiment for every review sentence, and can contain customized features like timestamps and geographical information, adding significant value for further analytical purposes. Figure 3 presents a code snippet and summary of the review class, detailing its properties and functions.

The Review class provides a unified level of abstraction across diverse review datasets from sources like SemEval and Twitter with varied file formats including json, xml, and csv as well as different annotations and domains such as restaurants, laptops, and social media. Each specific dataset, with its unique file structure and domain, inherits from the Review class and overrides the loader function to output standardized review objects. This uniformity allows various aspect detection and sentiment analysis models to operate with the review objects, eliminating the need for models to adapt to each dataset's specific structure and format. After transforming the reviews within a dataset into the standardized review objects, they are ready for the aspect model layer (aml) for model training purposes. However, before moving forward to the aspect detection phase, and in line with LADy's focus on empowering the models for latent occurrences of aspects, the review objects are first passed to the data augmentation layer.

3.2 Backtranslation Augmentation

Our research has shown inherent limitations in existing aspect term or aspect category detection methods. Aspect term detection models fall short when encountering novel terms or phrases that are not in training data. This limitation becomes apparent in scenarios where new aspects emerge for a product or service. Aspect category detection models, however, are constrained to a fixed set of categories. They might broadly classify a review under a category,

e.g., *'food'*, but lack the granularity to distinguish between sub-categories, e.g., desserts or specific dishes. Moreover, these models typically rely on explicit aspects directly mentioned in reviews.

Despite these challenges, enhancing existing models is feasible through dataset augmentation in the training phase without designing new aspect detection methods [9]. In LADy, we adopt backtranslation augmentation to enrich review datasets with a wider range of vocabulary. Backtranslation involves translating a sentence to a different language and then translating it back to the original language while preserving its meaning. This method adds contextually relevant synonymous terms to the original language, thereby enriching the training datasets. For instance, it might add *'food'* for *'dish'*. Utilizing machine translation, we experiment with translating reviews into a different language (e.g., French) and then back to English, exploring the potential of backtranslation for semantic augmentation. This process helps uncover latent aspects, address out-of-vocabulary terms, and clarify ambiguous expressions. For translations and backtranslations, we employ Meta's *'no language left behind'* (nllb) [4], a neural machine translator supporting 200 languages, chosen for its emphasis on universal translation and its capability to serve low-resource language communities.

From Figure 2, backtranslated versions of a review are then appended to the review object in augs property and added to the training dataset for subsequent training phases, as illustrated in Figure 1. Backtranslation presents its own challenges. There is a risk of topic drift in content during translation, which may lead to the loss of explicit aspect labels. To mitigate this, we have implemented a control mechanism that involves reviewing the backtranslated content and its corresponding labels. This process is shown in Figure 1, forming a part of our primary workflow for augmentation, which consists of two main steps:

(1) *Semantic Alignment*: We employ DeCLUTR [7] to ensure semantic consistency. This tool compares the original and back-translated reviews by assigning a semantic score. Reviews scoring below 0.5 suggest a topic drift and are excluded from the augmented dataset.

(2) *Term Alignment*: Recognizing that backtranslation may alter the wording, we utilize SimAlign [11] to transfer the aspect labels of the original review to its backtranslation versions.
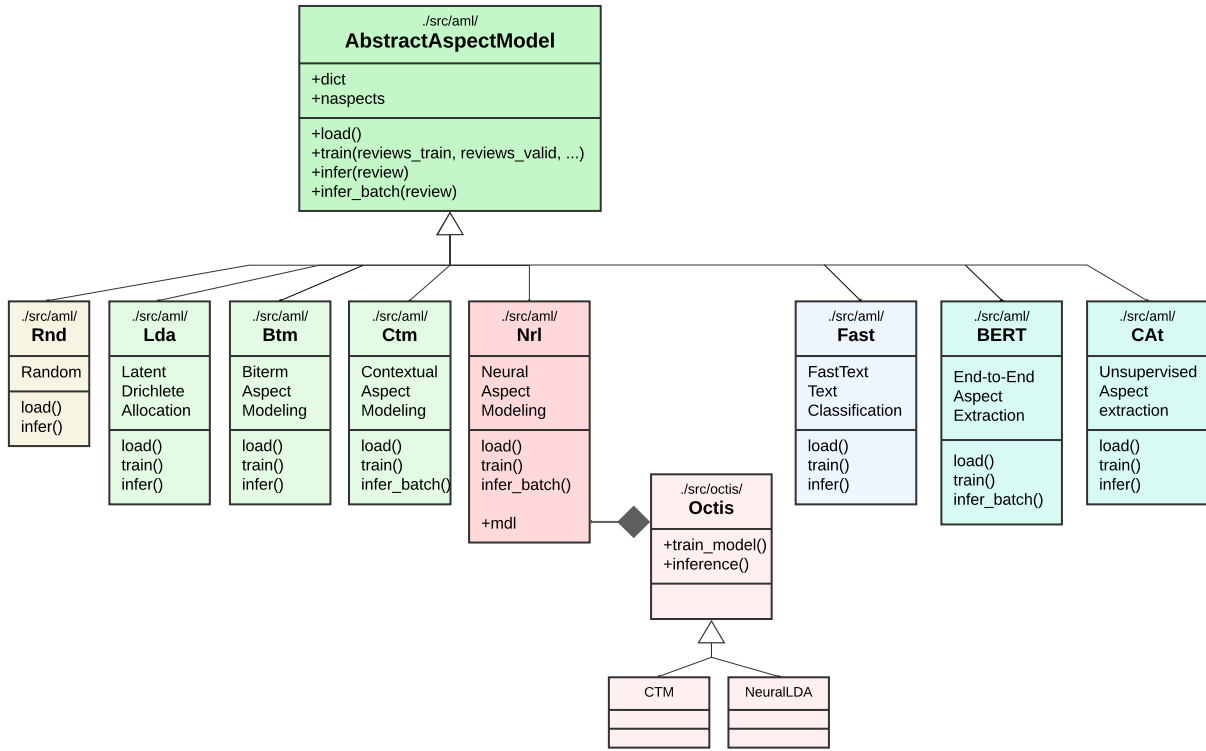
Figure 4: LADy's class diagram and inheritance hierarchy in the aspect model layer (`aml`).

```
# ./src/aml/mdl.py
class AbstractAspectModel:
    def __init__(self, naspects, nwords):
        self.naspects = naspects
        self.nwords = nwords
        self.dict = None
    def load(path): ...
    def train(reviews_train, reviews_valid, settings, output): ...
    def quality(metric): ...
    def infer(reviews, doctype): ...
    def preprocess(reviews, settings=None): ...
```

Figure 5: `AbstractAspectModel` class in the aspect model layer (`aml`).

SimAlign provides index mappings for word alignments between translated sentences, aiding in identifying new aos for the augmented reviews, which are then stored in their respective review objects.

A key point to underscore is that LADy is designed to use any augmentation technique. This flexibility is achieved by adding each augmented review as an attribute to the original review object, regardless of the augmentation method utilized, whether it is back-translation, as demonstrated in our example, or any other method. It is also important to note that the augmentation phase is optional, and we can proceed directly to the aspect detection phase and train the models using the original datasets. After navigating beyond the augmentation layer, review objects are now prepared for the training phase.

3.3 `AbstractAspectModel` Class

The aspect model layer (`aml`) features `AbstractAspectModel`, designed to provide a standardized interface to integrate various aspect modeling techniques. Figure 4 displays its key methods and attributes. A new aspect detection model has to inherit from `AbstractAspectModel` and override its training and inference functions to ensure that all models conform to uniform operations within LADy's pipeline. Moreover, `AbstractAspectModel` includes functions that are universally applicable across different models for fair processing and evaluation. `AbstractAspectModel` is a template that defines a standard set of behaviors and properties, which can then be inherited and customized by various emerging aspect detection models.

Initially, the model undergoes training using input training and validation data, and the resulting model is stored in the output. LADy incorporates a combination of traditional topic models and neural topic models as well as aspect detection models, both in unsupervised and supervised settings, as baselines. Figure 4 shows the class hierarchy of LADy's aspect detection methods. For training and comparative analysis against other baselines, the models are divided into two categories. The first category focuses on baselines that rely on topic modeling methods, explicitly used for detecting topics within documents; each review is considered a document with an associated distribution of topics, and the most probable words of the review's topics are indicative of the latent or explicit aspects of the review. The models in this category include:

- **`loclda`** [2]: A traditional topic modeling method that identifies aspects by associating highly probable words withing topics.

```
# ./src/params.py
settings = {
    'cmd': ['prep', 'train', 'test', 'eval', 'agg']
    'prep': {'langaug': 'pes_Arab', 'spa_Latn', ...]},
    'train': {'ratio': 0.85, 'nfolds': 5, 'nwords': 20,
              'langaug_semsim': 0.5,
              'qualities': ['coherence', 'perplexity']},
    'test': {'h_ratio': 0.5},
    'eval': {'metrics': ['precision', 'recall', ...],
             'topkstr': [1, 5, 10]}
```

Figure 6: LADy's settings and hyperparameters for each layer.

- **btm** [13]: Another traditional topic model that detects aspects based on patterns of word co-occurrence.
- **ctm** [1]: An unsupervised aspect detection model that combines neural networks with contextual topic modeling.
- **neuiallda** [22]: An unsupervised aspect detection model that integrates neural networks and topic modeling.

The second category comprises methods that are specifically designed for the task of aspect detection, including:

- **random**: This technique leverages a random selection process, extracting aspects from the list of words in the training set.
- **cat** [24]: An unsupervised aspect detection that classifies the review with predefined aspects.
- **bert-tfm** [15]: A supervised sequential labeling model utilizing bert [5].

LADy's aspect model layer is designed to facilitate the easy customization of training and inference procedures for a newly integrated model. These models function effectively as black boxes, requiring only hyperparameter settings to produce outputs, which include aspects and their corresponding evaluations.

Finally, we bring a trained model to make inferences (predictions) regarding the aspect of a given review, with special attention to occurrences of latent aspects.

3.4 Evaluation

The evaluation phase of LADy showcases its innovative approach to handling latent occurrences of aspects in reviews for aspect detection. While we aimed to establish a toolkit for detecting latent aspects, we found no existing public datasets featuring reviews with latent aspect labels. Given the absence of labeled datasets with latent aspects, LADy uses a masking approach; that is, the explicit aspects within a review in the *test* set are masked to mimic the latency. This masking is not mandatory, and users can bypass it by setting the masking proportion (h_ratio) to 0.0. This attribute, ranging from 0.0 (no latent aspects, meaning that all aspects are explicit) to 1.0 (all aspects are latent), is adjustable in the hyperparameters' setting, as shown in Figure 6, offering flexibility in test set creation.

Each model generates a ranked list of terms as its prediction for a review's aspect, which is evaluated against the labeled reference using a variety of metrics, including, but not limited to, precision, recall, normalized discounted cumulative gain (ndcg), mean average precision (map), and success at top-k, indicating the model's performance across various ranks of probable aspects. This feature is particularly relevant for latent aspect detection, where a single aspect may not be sufficient, and a ranked list of aspects offers greater insights. LADy's flexibility extends to evaluating explicit

Table 2: Statistics of the datasets.

Dataset	#reviews	average #aspects
semeval-14-laptop	1,488	1.5846
semeval-14-restaurant	2,023	1.8284
semeval-15-restaurant	833	1.5354
semeval-16-restaurant	1,234	1.5235
Twitter	6,248	1.0000

aspects when $k = 1$, and also, easy integration of new ranking metrics.

4 SYSTEM USAGE

LADy is designed to accommodate diverse users, from experienced Python developers to those preferring a more straightforward, user-friendly interface. For advanced users, LADy functions as a library, offering customization options and control over its various aspect models and parameters. This allows experienced programmers and data scientists to fully exploit its capabilities in a programmable environment. For a general user, LADy features an intuitive web dashboard, simplifying the user experience by providing a straightforward, navigable interface for experimenting with models, setting parameters, and viewing results.

4.1 Quick Start

LADy can be obtained at https://github.com/fani-lab/LADy. It incorporates various aspect models with specific hyperparameters, which can be set in the ./src/params.py and presented in Figure 6. LADy offers a host of languages for the backtranslation process on the training dataset. The entry point to the toolkit is ./src/main.py, which executes the entire pipeline, leading to the prediction of aspects for each review, followed by evaluation metrics at top-k predicted aspects. For a quick start experience, a toy sample of reviews is provided as well.

4.2 Case Study

We benchmarked LADy for the state-of-the-art aspect detection methods across various datasets in different domains from SemEval datasets [17–19], encompassing restaurant and laptop domains, as well as Twitter dataset [6] in social media domain. Table 2 shows the statistics about the currently available datasets.

Table 3 shows the partial benchmark results. As seen, the aspect detection methods show performance improvement in detecting explicit aspects when the training dataset is augmented with backtranslated reviews from *'all'* languages in the restaurant domain across a range of information retrieval evaluation metrics. Comprehensive benchmark results from other datasets, baselines, and additional metrics are available in [9] as well as the LADy's codebase.

4.3 Web Application

LADy features a web application to serve as a dynamic platform accommodating various aspect detection models, allowing researchers and general users to experiment with methods on an input review. The user-friendly interface facilitates interaction with different aspect detection methods, particularly highlighting latent aspect detection and the impactful integration of backtranslation during the models' training phase. Researchers can leverage the model

Table 3: The partial benchmark results of some aspect detection baselines with backtranslation augmentation evaluated on one of the datasets. The highest-performing results are highlighted in bold, and the second-best are underlined. Comprehensive results on other datasets, baselines, and additional metrics are available in the LADy's codebase, also report in [9].

| | semeval-15-restaurant | | | | | | | | | | | | | | | | | |
| | bert-tfm [15] | | | cat [24] | | | loclda [2] | | | btm [13, 30] | | | ctm [1] | | | random | | |
	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5
none	**0.7000**	0.6897	0.6757	0.3327	0.5248	0.4343	0.2320	0.3549	0.2925	0.1872	**0.3133**	0.2500	0.0560	0.0493	0.0485	0.0000	0.0005	0.0005
+Chinese	0.6661	0.6928	0.6699	0.3723	0.5287	0.4596	0.1968	0.3408	0.2647	0.1760	0.2783	0.2261	0.0624	0.0717	0.0637	**0.0016**	0.0028	0.0022
+Farsi	0.6742	0.6707	0.6608	0.3703	0.5386	0.4592	0.1840	0.3494	0.2689	0.1776	0.2834	0.2303	0.0560	0.0823	0.0722	0.0000	0.0002	0.0002
+Arabic	0.6661	0.6898	0.6671	0.4139	0.5683	0.4939	0.2000	0.3654	0.2887	0.1568	0.2956	0.2269	0.0592	0.0649	0.0577	0.0000	0.0000	0.0000
+French	0.6565	0.7030	0.6734	0.4040	0.5584	0.4883	0.2512	0.3577	0.3032	0.1968	0.3048	0.2481	0.0720	0.0837	0.0733	0.0000	0.0008	0.0006
+German	0.6710	0.6927	0.6721	0.3980	0.5505	0.4787	0.2416	0.3648	0.2976	0.1808	0.2691	0.2242	0.0560	0.0717	0.0603	0.0000	**0.0061**	**0.0036**
+Spanish	0.6645	0.7099	0.6769	0.3921	0.5663	0.4842	0.2224	0.3737	0.3035	0.2000	0.2975	0.2466	0.0464	0.0531	0.0458	0.0000	0.0000	0.0000
+all	0.6613	**0.7182**	**0.6823**	**0.5980**	**0.7861**	**0.7096**	**0.2592**	**0.3744**	**0.3104**	**0.2128**	0.2986	**0.2515**	**0.2192**	**0.2470**	**0.2263**	**0.0016**	0.0008	0.0010
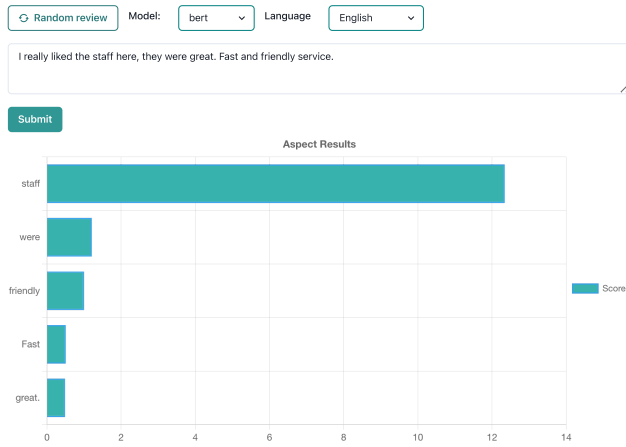
Figure 7: LADy's web user interface. The figure shows `bert-tfm`'s inference for a sample input review. The model has trained on the original dataset without backtranslation.

comparison functionality within the application, allowing for a direct comparison between models with and without backtranslation. Figure 7 presents a sample demonstration for `bert-tfm` baseline trained on the original dataset without backtranslation.

5 CONCLUSIONS

We introduced LADy 🐞, an open-source Python-based toolkit designed for conducting reproducible research in aspect-based sentiment analysis, with a particular focus on latent aspect detection. LADy incorporates various aspect detection models and diverse datasets, including unsolicited reviews from online social platforms. It also features backtranslation augmentation during the training phase, all within a flexible and adaptable integrated structure. Our future work will involve expanding LADy's capabilities to encompass sentiment analysis and opinion mining, allowing its ability to handle a broader spectrum of review analysis tasks.

REFERENCES

[1] Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. 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 (Volume 2: Short Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 759–766. https://doi.org/10.18653/v1/2021.acl-short.96

[2] Samuel Brody and Noemie Elhadad. 2010. An Unsupervised Aspect-Sentiment Model for Online Reviews. In *NAACL 2010*. 804–812. https://aclanthology.org/N10-1122/

[3] Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 2974–2985. https://doi.org/10.18653/v1/2022.acl-long.212

[4] Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, and et al. 2022. No Language Left Behind: Scaling Human-Centered Machine Translation. *CoRR* abs/2207.04672 (2022). arXiv:2207.04672

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[6] Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Kristina Toutanova and Hua Wu (Eds.). Association for Computational Linguistics, Baltimore, Maryland, 49–54. https://doi.org/10.3115/v1/P14-2009

[7] John Giorgi, Osvald Nitski, Bo Wang, and Gary Bader. 2021. DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 879–895. https://doi.org/10.18653/v1/2021.acl-long.72

[8] Zhibin Gou, Qingyan Guo, and Yujiu Yang. 2023. MvP: Multi-view Prompting Improves Aspect Sentiment Tuple Prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 4380–4397. https://doi.org/10.18653/v1/2023.acl-long.240

[9] Farinam Hemmatizadeh, Christine Wong, Alice Yu, and Hossein Fani. 2023. Latent Aspect Detection via Backtranslation Augmentation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 3943–3947. https://doi.org/10.1145/3583780.3615205

[10] Ehsan Hosseini-Asl, Wenhao Liu, and Caiming Xiong. 2022. A Generative Language Model for Few-shot Aspect-Based Sentiment Analysis. In *Findings of the Association for Computational Linguistics: NAACL 2022*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 770–787. https://doi.org/10.18653/v1/2022.findings-naacl.58

[11] Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. SimAlign: High Quality Word Alignments Without Parallel Training Data Using Static and Contextualized Embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 1627–1643. https://doi.org/10.18653/v1/2020.findings-emnlp.147

[12] Baoxing Jiang, Shehui Liang, Peiyu Liu, Kaifang Dong, and Hongye Li. 2023. A semantically enhanced dual encoder for aspect sentiment triplet extraction. *Neurocomputing* 562 (2023), 126917. https://doi.org/10.1016/j.neucom.2023.126917

[13] Ning Li, Chi-Yin Chow, and Jia-Dong Zhang. 2019. Seeded-BTM: Enabling Biterm Topic Model with Seeds for Product Aspect Mining. In *21st IEEE International Conference on High Performance Computing and Communications*. IEEE, 2751–2758.

[14] Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. Aspect Term Extraction with History Attention and Selective Transformation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 4194–4200. https://doi.org/10.24963/ijcai.2018/583

[15] Xin Li, Lidong Bing, Wenxuan Zhang, and Wai Lam. 2019. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*. Association for Computational Linguistics, Hong Kong, China, 34–41.

[16] Joseph Peper and Lu Wang. 2022. Generative Aspect-Based Sentiment Analysis with Contrastive Learning and Expressive Structure. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 6089–6095. https://doi.org/10.18653/v1/2022.findings-emnlp.451

[17] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, and et al. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch (Eds.). Association for Computational Linguistics, San Diego, California, 19–30. https://doi.org/10.18653/v1/S16-1002

[18] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015*. The Association for Computer Linguistics, 486–495.

[19] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING*. The Association for Computer Linguistics, 27–35.

[20] Tian Shi, Liuqing Li, Ping Wang, and Chandan K Reddy. 2021. A simple and effective self-supervised contrastive learning framework for aspect detection. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 13815–13824.

[21] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Targeted sentiment classification with attentional encoder network. In *Artificial Neural Networks and Machine Learning–ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part IV 28*. Springer, 93–103.

[22] Akash Srivastava and Charles Sutton. 2017. Autoencoding Variational Inference For Topic Models. In *5th International Conference on Learning Representations, ICLR 2017*. OpenReview.net. https://openreview.net/forum?id=BybtVK9lg

[23] Mohammad Tubishat, Norisma Idris, and Mohammad Abushariah. 2021. Explicit aspects extraction in sentiment analysis using optimal rules combination. *Future Generation Computer Systems* 114 (2021), 448–480. https://doi.org/10.1016/j.future.2020.08.019

[24] Stéphan Tulkens and Andreas van Cranenburgh. 2020. Embarrassingly Simple Unsupervised Aspect Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 3182–3187. https://doi.org/10.18653/v1/2020.acl-main.290

[25] Manju Venugopalan and Deepa Gupta. 2022. An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. *Knowledge-Based Systems* 246 (2022), 108668. https://doi.org/10.1016/j.knosys.2022.108668

[26] Manju Venugopalan and Deepa Gupta. 2022. A reinforced active learning approach for optimal sampling in aspect term extraction for sentiment analysis. *Expert Systems with Applications* 209 (2022), 118228. https://doi.org/10.1016/j.eswa.2022.118228

[27] An Wang, Junfeng Jiang, Youmi Ma, Ao Liu, and Naoaki Okazaki. 2023. Generative Data Augmentation for Aspect Sentiment Quad Prediction. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, Alexis Palmer and Jose Camacho-collados (Eds.). Association for Computational Linguistics, Toronto, Canada, 128–140. https://doi.org/10.18653/v1/2023.starsem-1.12

[28] Qianlong Wang, Zhiyuan Wen, Qin Zhao, Min Yang, and Ruifeng Xu. 2021. Progressive Self-Training with Discriminator for Aspect Term Extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 257–268. https://doi.org/10.18653/v1/2021.emnlp-main.23

[29] Zengzhi Wang, Qiming Xie, and Rui Xia. 2023. A Simple yet Effective Framework for Few-Shot Aspect-Based Sentiment Analysis. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1765–1770.

[30] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*. International World Wide Web Conferences Steering Committee / ACM, 1445–1456.

[31] Heng Yang, Chen Zhang, and Ke Li. 2022. PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis. *arXiv preprint arXiv:2208.01368* (2022).

[32] Yunyi Yang, Kun Li, Xiaojun Quan, Weizhou Shen, and Qinliang Su. 2020. Constituency Lattice Encoding for Aspect Term Extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), 844–855. https://doi.org/10.18653/v1/2020.coling-main.73

[33] Guoxin Yu, Jiwei Li, Ling Luo, Yuxian Meng, Xiang Ao, and Qing He. 2021. Self Question-answering: Aspect-based Sentiment Analysis by Role Flipped Machine Reading Comprehension. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 1331–1342. https://doi.org/10.18653/v1/2021.findings-emnlp.115