Latent Aspect Detection via Backtranslation Augmentation

Farinam Hemmatizadeh
University of Windsor, Canada
hemmatif@uwindsor.ca

Christine Wong
University of Windsor, Canada
wong93@uwindsor.ca

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

Hossein Fani
University of Windsor, Canada
hfani@uwindsor.ca

ABSTRACT

Within the context of review analytics, aspects are the features of products and services at which customers target their opinions and sentiments. Aspect detection helps product owners and service providers identify shortcomings and prioritize customers' needs. Existing methods focus on detecting the surface form of an aspect falling short when aspects are latent in reviews, especially in an informal context like in social posts. In this paper, we propose data augmentation via natural language backtranslation to extract latent occurrences of aspects. Specifically, we presume that backtranslation *i*) can reveal latent aspects because they may not be commonly known in the target language and can be generated through backtranslation; *ii*) augments context-aware synonymous aspects from a target language to the original language, hence addressing the out-of-vocabulary issue; and *iii*) helps with the semantic disambiguation of polysemous words and collocations. Through our experiments on well-known aspect detection methods across semeval datasets of restaurant and laptop reviews, we demonstrate that review augmentation via backtranslation yields a steady performance boost in baselines in all datasets. We further contribute LADy[1] (💃), a benchmark library to support the reproducibility of our research.

1 INTRODUCTION

The key characteristic of customers' opinions is to target aspects of a product or service to convey an opinion. For example, the review *"forced us to buy pricey dresses by cheap behavior"* entails the aspects *'dress'* and *'seller'* toward which opinions of *'pricey'* and *'cheap behavior'* is expressed with *'negative'* sentiments, respectively. Aspect detection is crucial in customers' review analysis in e-commerce and social platforms [19]; it helps product owners and service providers to identify shortcomings and prioritize improvements according to customers' needs, hence maintaining revenues and preventing customer churn [1, 14, 33].

Existing methods to detect aspects in reviews and their respective opinions and sentiments have primarily expanded into two subtasks: *i*) *aspect term* sentiment analysis which extracts explicit term(s) that point to an aspect either by tagging the terms [16, 17] or delineating the span of the terms [31, 36] in the review's text (fine-grained terms), and *ii*) *aspect category* sentiment analysis which maps a review into a set of predefined course-grained (higher-order) categories of aspects [13, 25, 28]. In our example, an aspect term extraction method would tag *'dresses'* as an aspect. However, it falls short of detecting a latent or implicit occurrence of an aspect like *'seller'*. An aspect category detection method would detect the categories *'clothing'* (as *'dress'* is-a *'clothing'*), along with *'staff'* (as *'seller'* is-a *'staff'*) irrespective of explicit or latent occurrences of aspects.

[1] `anonymous.4open.science/r/LADy/`

Table 1: Sample reviews: [latent] vs. explicit aspects.

original	*"[seller] forced us to buy pricey dresses through cheap behavior"*
`farsi`	"فروشنده با رفتار زننده، ما را مجبور به خرید لباس‌های گران می کرد"
backtranslated	*"with cheap behavior, seller forced us to buy expensive clothes"*
original	*"served me the wrong dish!"*
`spanish`	*"me sirvió la comida equivocada!"*
backtranslated	*"served me the wrong food!"*

On the one hand, aspect term sentiment analysis approaches forego the latent occurrences of aspects in reviews, as seen earlier in the example for *'seller'*. Indeed, when writing a review, the customer may write about her opinion and overall rating while overlooking the aspects' terms for a product or service due to being part of common knowledge. As shown in the literature, 35% of reviews on restaurants and electronics include latent aspects [5, 30]. The need for latent aspects detection is even more pressing in online social review platforms where reviews are unsolicited, short, noisy, informal, and mostly rely on social background knowledge and context. Furthermore, aspect term extraction methods fall short in the presence of out-of-vocabulary (oov) aspects, i.e., aspects that have been *un*seen in the training dataset. On the other hand, although aspect category sentiment analysis approaches are agnostic to the occurrence of aspect terms, they heavily rely on occurrences of other parts of a review, such as opinion terms whose latency renders aspect category detection challenging.

In this paper, we propose review augmentation through natural language backtranslation to address the latent occurrence and out-of-vocabulary aspects in aspect term extraction methods while overcoming the performance drain caused by the latency of fine-grained terms in aspect category detection methods. Specifically, we translate a review from its original language (e.g., `english`) to a target language (e.g., `french`), and then translate it back to the original language using a machine translator (e.g., Meta's `nllb` [6]). We presume this round-trip translation generates diverse paraphrases of a review while withholding semantic context [34], as a result of which:

(1) Backtranslation can reveal latent aspects as they may not be commonly known in the target language. For instance, when *"forced us to buy pricey dresses by cheap behavior"* is translated to `farsi` as: "فروشنده با رفتار زننده ما را مجبور به خرید لباس‌های گران می کرد", followed by a backtranslation to `english`, *"with cheap behavior, seller forced us to buy expensive clothes"*, it brings up *'seller'*;

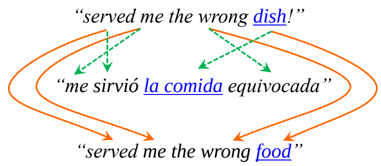

Figure 1: Term alignment.

(2) Backtranslation can address out-of-vocabulary aspects by augmenting *context-aware* synonymous aspects from the target language to the original language, as opposed to simple synonym replacement [7, 35]. For example, when *"... served me the wrong dish!"* is translated to `spanish` as *"... me sirvió la comida equivocada! ..."*, followed by a backtranslation to `english` *"... served me the wrong food!"*, *'food'* appears for *'dish'*, compared to *'bowl'* or *'plate'*; **(3)** Backtranslation can disambiguate polysemous terms and collocations, leading to the detection of latent aspects. For instance, translating *"... through cheap behavior"* to `spanish` *"... a través de su comportamiento insignificante"*, and backtranslating to `english` *"... through her petty behavior"* maps the term *'cheap'* to *'petty'*, which is more semantically related to behavior of a person, leading to the detection of the latent aspect *'staff'*, as opposed to other semantics like *'inexpensive'* for *'dresses'*.

For similar reasons, backtranslation has been employed in opinion mining [10, 29] and other various natural language processing tasks [9, 18, 24]. However, there has been no study on its synergistic impacts on aspect detection, to the best of our knowledge. In this paper, we systematically benchmark aspect detection models when their training sets of reviews in `english` are augmented with backtranslated versions through various languages from different language families and study the effects on the performance in terms of information retrieval metrics at top-k predicted aspects. We further contribute `LADy`[1] (💃), an open-source, extensible, and standard benchmark library, to support the reproducibility of our research. Through our experiments on well-known supervised and unsupervised aspect detection baselines across semeval datasets of 2014, 2015, and 2016, on restaurant and laptop domains, we demonstrate that review augmentation via backtranslation has led to a steady performance boost in baselines in different domains.

2 PROBLEM DEFINITION

Our goal is to explore the synergistic impact of natural language backtranslation as an augmentation technique on the aspect detection task, especially when aspects are *latent*.

Definition 1 (**Aspect Detection**). *Given a review after normalization of its raw text in a natural language l as a bag (set) of terms $r_p = \{i\}$ about a product or a service p concerning an aspect $a \in A_p$ which <u>may not be</u> in r_p, i.e., $a \notin r_p$ or $a \in r_p$, the aspect detection aims at identifying aspect a in r_p. A_p is the set of all aspects of the product or service p and $\mathcal{R}_p$ is the set of all p's reviews.*

3 AUGMENTATION VIA BACKTRANSLATION

While languages share underlying commonalities referred to as linguistic *universals* or cross-linguistic generalizations due to the common neurobiological basis of the human brain [11], they carry differences on the surface, including phonetics, morphological units (terms), syntax, and semantics, especially in an informal context like in social posts. Particular examples are ellipsis in writing when we omit terms that are superfluous or can be understood from the context in one language but need to be explicitly mentioned in another language like in *"[seller] forced us to buy pricey dresses through cheap behavior"* in which ellipsis occurred in the aspect *'seller'* in `english` as opposed to its translation in `farsi`, as shown in Table 1. In this paper, we aim to employ the textual ellipsis differences in various natural languages to uncover latent aspects in reviews through language backtranslation as well as addressing out-of-vocabulary aspects. Our research includes two main pipeline components: (1) review backtranslation, which is further divided into two subcomponents of backtranslation and semantic alignment, and (2) review aspect detection.

3.1 Review Backtranslation

Let $\mathcal{L}$ be the set of languages. Given a review $r_p = \{i\} \in l \in \mathcal{L}$, we translate it to language $l' \in \mathcal{L}$ resulting in review $r'_p \in l'$ and backtranslate it to l which results in $r^+_p = \{i^+\}$. We create an augmented review set $\mathcal{R}^+_p$ by adding augmented reviews to the original set $\mathcal{R}_p$, which is used for the training phase. The original set $\mathcal{R}_p$ can also be augmented with multiple or *all* languages in $\mathcal{L}$.

3.1.1 Backtranslation. In our study, without loss of generality to any machine translation models, we apply Meta's *'no language left behind'* (`nllb`) [6], an open-source neural machine translator capable of providing high-quality translations between `200` languages. We deliberately chose `nllb` for its particular focus on realizing a universal translation system while prioritizing the needs of underserved communities for low-resource natural languages, as opposed to a small dominant subset of natural languages; it enables review backtranslation augmentation via a vast variety of natural languages with distinct properties. Further, `nllb` is open-sourced to foster transparency and can be smoothly integrated into any pipeline with few lines of code. For this paper, we translated `english` reviews into `french`, `german`, `spanish`, and `farsi` from indo-european language family, `chinese` from sino-tibetan language family, and `arabic` from afro-asiatic language family.

We employ the same `nllb` translator to bring the translated reviews back to `english`. As shown in Table 1, a backtranslated version of a review may carry term replacement (e.g., *'food'* for *'dish'*) and/or new terms (e.g., *'seller'*), among other changes. Although Yu et al. [34] have shown that natural language backtranslation preserves the semantic context, i.e., an original piece of text and its backtranslated versions are about the same topic, we have observed otherwise in review analysis, especially when using languages whose family is different from `english`'s like `chinese` or `arabic`. For example, the review *"the duck confit is always amazing"* became *"the duck crib is always fantastic"* after backtranslation via `chinese` where the aspect *'duck confit'* (*'food'*) has semantically drifted to *'duck crib'* (*'furniture'*). To preserve semantic context consistency during backtranslation, we filter out backtranslated versions of original reviews that are semantically drifted, as explained in the next section.

3.1.2 Semantic Alignment. To alleviate semantic drift in backtranslation augmentation, we perform pairwise term

Table 2: Statistics on original and backtranslated reviews.

dataset	#reviews	avg #aspects	exact match					
			chinese	farsi	arabic	french	german	spanish
semeval-14-laptop	1,488	1.5846	0.1763	0.2178	0.2727	<u>0.3309</u>	0.3214	**0.3702**
semeval-14-restau	2,023	1.8284	0.1831	0.2236	0.2929	0.3645	<u>0.3724</u>	**0.4088**
semeval-15-restau	0,833	1.5354	0.2034	0.2312	0.3021	0.3587	<u>0.3907</u>	**0.4128**
semeval-16-restau	1,234	1.5235	0.2023	0.2331	0.2991	0.3556	<u>0.3834</u>	**0.4034**

alignment [27] between the original review and backtranslated versions to ensure accurate and meaningful machine translation. Term alignment matches the aspect terms of an original review with a corresponding term in the backtranslated review. Moreover, it matches the aspect terms of the original review with its corresponding *aligned* backtranslated aspect terms. Figure 1 displays an example where *'dish'* in the original review has matched with *'food'* in the spanish-backtranslated version. For term alignment, we use simalign-itermax by Sabet et al. [27], an unsupervised approach to align terms (also phrases) in a pair of parallel texts by calculating the pairwise cosine similarities between the embeddings of words in the source and target texts.

We further apply semantic similarity on pairs of review and backtranslated versions using declutr by Giorgi et al. [12] and filtered out backtranslationed reviews with less than a threshold (e.g., `0.5`) semantic similarity.

3.2 Review Aspect Detection

To assess the contribution of review augmentation using backtranslation, we train two instances of an aspect detection model: one is trained on an original dataset in language l, i.e., $\mathcal{R}_p$, and another one is trained on augmented datasets via language l', i.e., $\mathcal{R}_p^+$. We cross-compare the performance of two instances of the model on the same test set with no augmentation.

4 EXPERIMENTS

We seek to answer the following research questions:

RQ1: Does augmentation via natural language backtranslation improve the performance of aspect detection models?

RQ2: Does backtranslation augmentation improve the performance of aspect detection models when aspects are *latent*?

RQ3: Is the impact of augmentation consistent across datasets from different review domains?

4.1 Setup

4.1.1 Datasets. Our benchmark includes well-known publicly available semeval datasets of reviews in english for aspect detection. We used training sets of semeval-14's reviews on restaurants and laptops [22], and semeval-15 and semeval-16's reviews on restaurants [20, 21]. Following the literature [17, 28], we segmented reviews into sentences and performed our experiments on each sentence as an individual review assuming each sentence entails one aspect. Table 2 shows the datasets' statistics.

4.1.2 Backtranslation and Semantic Alignment. We randomly divided a dataset into 85% training and 15% test sets of reviews. We augmented the training set of reviews through backtranslation using Meta's nllb [6]. Table 2 shows the average scores of exact match metric between original reviews and their respective backtranslated versions per dataset after filtering out backtranslated reviews with less than `0.5` semantic similarity for semantic drift.

Figure 2: Performance gain in latent aspect detection.

As seen, backtranslated versions of reviews from languages that belong to the same family as english, including spanish, french and german, are more similar compared to chinese and arabic. We further merge all backtranslated reviews from all different languages into one set, referred to as all, to explore the impact of backtranslation from all languages at once.

4.1.3 Baselines. We benchmarked the following supervised and unsupervised aspect detection models.

bert-tfm [17] is a *supervised* tagging-based method that employs contextual embeddings of reviews from bert [8] followed by a self-attention layer.

cat [28] is a simple yet effective *unsupervised* aspect detection method. It forms a candidate set of aspect terms whose part of speech is of type noun. An input review is then transformed into an attention matrix using pretrained vector representations of constituent terms. Finally, it selects an aspect term that has the highest similarity with the weighted vector of the input review.

loclda [4] is an *unsupervised* method following the same assumption as in topic modeling methods; the review's terms are generated based on an aspect (topic). The model uses latent Dirichlet allocation [3].

btm [15, 32] is an *unsupervised* method with a special focus on short texts like reviews. It learns the aspects by directly modeling the generation of term cooccurrence pairs (biterms) in the entire review dataset to address the problem of sparse term occurrences in loclda.

neurallda [26] is an *unsupervised* topic modeling method based on variational autoencoder, which encodes the term occurrence vector representation of reviews onto a continuous latent representation as aspects.

ctm [2] extends neurallda with the concatenation of the term occurrence vector and pretrained dense contextual vector representation from a bert-based transformer, e.g. sentence-bert [23], in the input.

random is a naive method that chooses an arbitrary term as the review's aspect to provide a minimum base for comparison.

For the complete list of baselines' hyperparameters, see LADy (💃) codebase.

4.1.4 Evaluation Methodology. We performed 5-fold cross-validation on a training set for model training and validation, which results in one trained model per fold. Given a review of the test set, we compared the ranked list of predicted aspect terms by the model of each fold with the observed aspect terms and reported the average performance of models in all folds by information retrieval metrics, including normalized discounted cumulative gain

Table 3: The average performance of 5-fold models with backtranslation augmentation and lack thereof on the test set.

	bert-tfm [17]			cat [28]			loclda [4]			btm [15, 32]			ctm [2]			neuralda [26]			random		
	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5	pr1	rec5	ndcg5
semeval-14-laptop																					
none	**0.6194**	**0.6487**	**0.6235**	0.4591	0.6362	0.5598	0.1188	0.1536	0.1308	0.0705	0.1079	0.0908	0.0286	0.0459	0.0379	0.0116	**0.0179**	**0.0155**	0.0000	0.0000	0.0000
+chinese	0.6018	0.6347	0.6102	**0.5409**	0.6564	0.6082	0.1179	0.1680	0.1407	0.1080	0.1309	0.1173	0.0732	0.1003	0.0844	0.0054	0.0063	0.0059	0.0000	0.0000	0.0000
+farsi	0.6074	0.6314	0.6092	0.5369	0.6685	0.6112	0.0875	0.1390	0.1148	0.1438	0.1321	0.1276	0.0384	0.0593	0.0501	0.0054	0.0086	0.0066	0.0000	0.0000	0.0000
+arabic	0.6018	0.6237	0.6039	0.5154	0.6537	0.5959	0.1000	0.1420	0.1194	0.1107	0.1342	0.1177	0.0464	0.0770	0.0608	0.0063	0.0085	0.0070	**0.0018**	0.0010	0.0012
+french	0.6184	0.6290	0.6112	0.5168	0.6685	0.6040	0.1223	0.1705	0.1462	0.1170	0.1430	0.1263	0.0518	0.0910	0.0733	0.0045	0.0076	0.0061	**0.0018**	**0.0019**	**0.0018**
+german	0.6055	0.6336	0.6096	0.5315	0.6685	0.6103	**0.1411**	**0.1890**	**0.1621**	0.1000	0.1206	0.1068	0.0991	0.1199	0.1066	0.0036	0.0060	0.0049	0.0000	0.0009	0.0004
+spanish	0.6018	0.6291	0.6092	0.5356	**0.6711**	**0.6127**	0.1188	0.1669	0.1414	0.1045	0.1431	0.1225	0.0500	0.0768	0.0638	0.0045	0.0048	0.0047	0.0009	0.0017	0.0015
+all	0.6028	0.6194	0.6025	0.5195	0.6510	0.5966	0.1188	0.1549	0.1336	**0.1339**	**0.1476**	**0.1347**	**0.1652**	**0.2007**	**0.1757**	**0.0134**	0.0149	0.0130	0.0000	0.0001	0.0001
semeval-14-restaurant																					
none	0.6061	0.6564	0.6293	0.3442	0.5478	0.4519	**0.2428**	0.2845	**0.2557**	0.1717	0.2361	0.1995	0.0368	0.0941	0.0682	0.0099	0.0307	0.0221	0.0000	0.0002	0.0001
+chinese	**0.6121**	0.6627	**0.6338**	0.6221	0.8248	0.7395	0.1980	0.2656	0.2270	0.1743	0.2236	0.1945	0.2020	0.1994	0.1863	0.0046	0.0309	0.0203	0.0000	0.0001	0.0001
+farsi	0.5946	0.6390	0.6166	0.6133	0.8159	0.7321	0.1770	0.2803	0.2322	0.1664	0.2238	0.1920	0.1289	0.1912	0.1604	**0.0105**	**0.0493**	**0.0312**	0.0007	0.0013	0.0008
+arabic	0.6054	0.6558	0.6286	0.6221	0.8265	0.7417	0.2408	0.2866	0.2548	0.1678	0.2336	0.1980	0.1724	0.2218	0.1909	0.0026	0.0250	0.0153	0.0000	0.0007	0.0004
+french	0.6040	0.6516	0.6253	0.6159	0.8274	0.7399	0.1645	0.2533	0.2068	0.1572	0.2090	0.1773	0.1947	0.2036	0.1862	0.0086	0.0365	0.0250	0.0000	0.0000	0.0000
+german	0.5987	0.6553	0.6275	0.6257	**0.8363**	0.7487	0.1770	0.2876	0.2329	0.1875	0.2350	0.2056	0.1467	0.1715	0.1474	0.0033	0.0272	0.0167	**0.0013**	**0.0018**	**0.0016**
+spanish	0.6000	**0.6581**	0.6278	**0.6319**	0.8354	**0.7501**	0.1908	0.2853	0.2389	0.1638	0.2266	0.1947	0.1914	0.1922	0.1803	0.0007	0.0157	0.0108	0.0000	0.0003	0.0003
+all	0.5865	0.6566	0.6254	**0.6319**	0.8239	0.7435	0.1993	**0.2919**	0.2445	**0.1921**	**0.2386**	**0.2076**	**0.2487**	**0.2741**	**0.2481**	0.0013	0.0355	0.0197	0.0007	0.0003	0.0004
semeval-15-restaurant																					
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	0.0080	0.0410	0.0244	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	0.0112	0.0575	0.0354	**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	0.0096	0.0400	0.0253	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	0.0271	0.0160	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	0.0176	0.0551	0.0377	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	0.0048	0.0542	0.0324	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	0.0192	0.0314	0.0246	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**	**0.0224**	**0.0731**	**0.0478**	**0.0016**	0.0008	0.0010
semeval-16-restaurant																					
none	0.6844	0.6911	0.6806	0.4193	0.5496	0.4912	0.1699	0.2828	0.2248	0.0828	0.1600	0.1204	0.0226	0.0430	0.0352	0.0097	0.0389	0.0250	**0.0022**	0.0008	0.0009
+chinese	0.6700	**0.7062**	**0.6864**	0.5659	0.6904	0.6371	0.1538	0.2781	0.2173	0.0968	0.1446	0.1189	0.0624	0.0891	0.0769	0.0129	0.0262	0.0196	0.0000	0.0014	0.0008
+farsi	0.6811	0.6915	0.6783	0.5733	0.7259	0.6634	0.1398	0.2716	0.2068	0.0731	0.1425	0.1063	0.0839	0.1230	0.1023	0.0129	**0.0592**	**0.0393**	0.0000	0.0005	0.0003
+arabic	0.6744	0.6849	0.6736	0.5630	0.7378	0.6661	0.1785	0.2879	0.2279	0.0645	0.1456	0.1062	0.0774	0.1118	0.0924	**0.0151**	0.0316	0.0223	0.0000	0.0013	0.0008
+french	0.6811	0.6963	0.6834	**0.5763**	0.7541	**0.6796**	**0.2118**	0.2919	**0.2451**	0.0871	0.1473	0.1159	0.0602	0.1082	0.0868	0.0075	0.0451	0.0287	0.0011	0.0016	0.0011
+german	**0.6856**	0.6891	0.6806	0.5719	**0.7481**	0.6764	0.1312	0.2715	0.2060	0.0860	0.1514	0.1188	0.0602	0.1009	0.0793	0.0086	0.0303	0.0227	0.0011	**0.0022**	**0.0018**
+spanish	0.6656	0.6951	0.6784	0.5600	0.7467	0.6697	0.1957	**0.2935**	0.2408	**0.1054**	**0.1718**	**0.1372**	0.0656	0.1107	0.0879	0.0140	0.0372	0.0279	0.0000	0.0014	0.0007
+all	0.6622	0.6992	0.6774	0.5304	0.7141	0.6375	0.1591	0.2694	0.2130	0.0774	0.1410	0.1070	**0.2097**	**0.2607**	**0.2272**	**0.0151**	**0.0487**	**0.0304**	0.0000	0.0009	0.0006

(ndcg) as well as classification metrics including precision (pr) and recall (rec). To evaluate how backtranslation augmentation helps with our trained model in detecting latent aspects, given a random review from the test set, we removed the aspect terms from the review (synthetically make it latent) and used our model to predict the review's latent aspect. We evaluate the baselines on a randomly increasing percentage of latent aspects in the test set from 0% (all aspects are explicit) to 100% (all aspects are latent).

4.2 Results

In response to **RQ1**, i.e., whether augmentation via language backtranslation improves the performance of aspect detection methods, from Table 3, all baselines, including random, could generally improve upon augmentation across datasets and domains. Specifically, the best results were when the training sets were augmented from *all* languages (+all).

On a per-language basis, we can further observe that backtranslation augmentation via languages in the same family as english like spanish, german, or french generally yielded runner-up results. However, there are inconsistencies like backtranslation augmentation via arabic that obtain the competitor results in semeval-15-laptop using cat baseline. Further, from a row-wise view, baseline methods have different performance gains from review augmentation. For instance, a poor method before augmentation like ctm can boost its performance by augmentation and become one of the best methods among unsupervised baselines as opposed to neurallda's marginal performance gain. With respect to the state-of-the-art methods, bert-tfm's gain is limited since it identifies an aspect term within the elements of an input sequence solely overlooking other possible candidate aspect terms. In contrast, cat is unsupervised and can select aspect terms from a global vocabulary set, hence benefiting substantially from augmentation. In general, unsupervised baselines enhanced their performance through augmentation in any language of choice across all domains and datasets. Last, when it comes to comparing neural vs. traditional topic modelings, neural models like ctm utilized augmentation more effectively compared to non-neural models such as loclda and btm.

To answer **RQ2**, i.e., whether backtranslation augmentation improves the performance of aspect detection methods when aspects are latent. Figure 2 shows the performance gain of models when their training sets were augmented by +all languages compared to no augmentation. As seen, while there is a general declining trend as more aspects in the test set become latent, selected models, specifically unsupervised ones such as cat and ctm, show strong positive performance gain in terms of pr1 and ndcg5 to predict latent aspects. Interestingly, augmentation via backtranslation helps cat and ctm up to +15% improvements in challenging scenarios where 70% of reviews contain latent aspects.

Regarding **RQ3**, i.e., if the impact of augmentation across datasets from different domains is consistent, from Table 3, we see that backtranslation augmentation via all languages (+all) generally improved the performance in restaurant and laptop reviews in *all* semeval datasets, which implies the domain-agnostic synergy of backtranslation augmentation in aspect detection.

5 CONCLUSION AND FUTURE WORK

We presented augmentation via backtranslation for the task of aspect detection. Our experiments on backtranslation via six natural languages from varying language families demonstrate the synergistic impact of backtranslation augmentation across aspect detection methods and domains, including restaurant and laptop, esp., in reviews where aspects happen to be latent. Our future research includes experiments on *i*) datasets of unsolicited reviews from social media, which are short and informal, and *ii*) backtranslation augmentation via low-resource languages.

REFERENCES

[1] Nadia Alboukaey, Ammar Joukhadar, and Nada Ghneim. 2020. Dynamic behavior based churn prediction in mobile telecom. *Expert Syst. Appl.* 162 (2020), 113779.

[2] 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)*. Association for Computational Linguistics, Online, 759–766.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

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

[5] Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions. 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*. Association for Computational Linguistics, 340–350.

[6] 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

[7] Xiang Dai and Heike Adel. 2020. An Analysis of Simple Data Augmentation for Named Entity Recognition. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*. International Committee on Computational Linguistics, 3861–3867.

[8] 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 4171–4186.

[9] Alexander R. Fabbri, Simeng Han, Haoyuan Li, Haoran Li, Marjan Ghazvininejad, Shafiq R. Joty, Dragomir R. Radev, and Yashar Mehdad. 2021. Improving Zero and Few-Shot Abstractive Summarization with Intermediate Fine-tuning and Data Augmentation. 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*. Association for Computational Linguistics, 704–717.

[10] Hao Fei, Yafeng Ren, Shengqiong Wu, Bobo Li, and Donghong Ji. 2021. Latent Target-Opinion as Prior for Document-Level Sentiment Classification: A Variational Approach from Fine-Grained Perspective. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*. ACM / IW3C2, 553–564.

[11] Angela D Friederici. 2017. *Language in our brain: The origins of a uniquely human capacity.* MIT Press.

[12] John M. Giorgi, Osvald Nitski, Bo Wang, and Gary D. 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, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*. Association for Computational Linguistics, 879–895.

[13] Mengting Hu, Shiwan Zhao, Honglei Guo, Chao Xue, Hang Gao, Tiegang Gao, Renhong Cheng, and Zhong Su. 2021. Multi-Label Few-Shot Learning for Aspect Category Detection. 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, 6330–6340.

[14] Hemlata Jain, Ajay Khunteta, and Sumit Srivastava. 2021. Telecom churn prediction and used techniques, datasets and performance measures: a review. *Telecommun. Syst.* 76, 4 (2021), 613–630.

[15] 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.

[16] Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. Aspect Term Extraction with History Attention and Selective Transformation. In *IJCAI 2018*. 4194–4200.

[17] 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.

[18] Yu Li, Xiao Li, Yating Yang, and Rui Dong. 2020. A Diverse Data Augmentation Strategy for Low-Resource Neural Machine Translation. *Inf.* 11, 5 (2020), 255.

[19] Muhammad Marong, Nowshath K Batcha, and Raheem Mafas. 2020. Sentiment Analysis in E-Commerce: A Review on The Techniques and Algorithms. *Journal of Applied Technology and Innovation (e-ISSN: 2600-7304)* 4, 1 (2020), 6.

[20] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, and et al. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, 19–30.

[21] 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.

[22] 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.

[23] 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 2019, Hong Kong, China, November 3-7, 2019*. Association for Computational Linguistics, 3980–3990.

[24] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.

[25] 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.

[26] 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

[27] Steinþór Steingrímsson, Hrafn Loftsson, and Andy Way. 2021. CombAlign: a Tool for Obtaining High-Quality Word Alignments. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics, NoDaLiDa 2021, Reykjavik, Iceland (Online), May 31 - June 2, 2021*. Linköping University Electronic Press, Sweden, 64–73. https://aclanthology.org/2021.nodalida-main.7/

[28] 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*. Association for Computational Linguistics, Online, 3182–3187.

[29] Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2020. Latent Opinions Transfer Network for Target-Oriented Opinion Words Extraction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 9298–9305. https://ojs.aaai.org/index.php/AAAI/article/view/6469

[30] Hua Xu, Fan Zhang, and Wei Wang. 2015. Implicit feature identification in Chinese reviews using explicit topic mining model. *Knowl. Based Syst.* 76 (2015), 166–175.

[31] Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. 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, 4755–4766.

[32] 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.

[33] Carl Yang, Xiaolin Shi, Luo Jie, and Jiawei Han. 2018. I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*. ACM, 914–922.

[34] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. 2018. QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension. In *6th International Conference on Learning Representations, ICLR 2018, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=B14TlG-RW

[35] Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. 2015. Character-level Convolutional Networks for Text Classification. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. 649–657. https://proceedings.neurips.cc/paper/2015/hash/250cf8b51c773f3f8dc8b4be867a9a02-Abstract.html

[36] He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. SpanMlt: A Span-based Multi-Task Learning Framework for Pair-wise Aspect and Opinion Terms Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 3239–3248.