Explicit Aspect Annotation via Transfer and Active Learning

Maroua Boudabous, Anna Pappa

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Explicit Aspect Annotation via Transfer and Active Learning

Maroua Boudabous[a], Anna Pappa[a]

[a] *University Paris 8, 2 rue de la Liberté, Saint-Denis 93526 cedex, France*

Abstract

We present a semi-supervised annotation process for identifying and labelling explicit aspects of an initially unlabelled corpus. Firstly, we employ cross-domain learning to pre-annotate the initial data, deliberately excluding domain-related input features to ensure effective learning transfer. Then, we apply an active learning strategy to enhance the pre-annotation performance and enrich the learning data. We adjust the strategy to sequence labeling and address class imbalance. We evaluate this process using two unlabelled datasets in French, consisting of user opinions on beauty products and electronic devices, respectively. The results show an improved F1-score achieved by increasing and correcting 30% of the training dataset.

Keywords: Automatic annotation; Transfer learning; Active learning; Aspect term extraction.

1. Introduction

Data annotation is crucial for refining or improving the performance of language models in natural language processing (NLP). Leech et al. [14] described, in addition to the first tools and methods for corpora annotating, the need for annotation to extract information from corpora by adding labels containing explicit information about the content. Annotation helps to emphasize the meaning and remove ambiguities. It also allows capturing information valuable to machine learning algorithms, improving their ability to understand and analyze natural languages. Several works describe methods and tools for different types of annotation, such as linguistic annotation of corpora [7] or semantic textual annotation [18].

Aspect annotation is a fine semantic task, often related to sentiment analysis [37]. It is also denoted as Aspect Terms Extraction (ATE) and is defined as the task of detecting aspects (the terms used for designating product or

* Boudabous Maroua. Tel.: +3-362-044-4780.
E-mail address: mboudabous@ai.univ-paris8.fr

service attributes and features) from text. It is the first step of the Aspect-Based opinion or Sentiment Analysis (ABSA) process.

This task can be challenging and has received substantial attention from the NLP scientific community. Two types of aspects are mainly distinguished: explicit and implicit or hidden ones (aspects which are implied in the sentence).

Aspect Term Extraction (ATE) is often considered as a supervised sequence tagging task. However, gathering annotated data from each new domain for model training is costly and time-consuming. Therefore, it requires implementing a learning strategy that helps minimize the human efforts for annotation and adapt the aspect extraction model for low-resource languages.

In this paper, we propose an end-to-end process for explicit aspect term extraction where no in-domain labeled data is available. We designed three main steps for pseudo-labelling, model learning, and performance boosting with active learning, respectively. This paper presents three main contributions:

- We adapt features selection for the pseudo labeler to cross-domain learning by omitting domain-related ones.
- We define the process as model-agnostic so that any sequence labelling learning model can be plugged into it.
- We adjust the pool query strategy for active learning to account for sequence labelling and class imbalance.

To evaluate our process, we consider the use case of French unlabelled data from Beauty and Hightech product reviews for performance evaluation.

This paper is organized as follows: section 2 gives an overview of existing works in the fields of ATE and AL in Natural Language Processing. Section 3 presents formally the considered problem. Section 4 describes in details the proposed end-to-end process. In section 5, we describe the experiment setting and discuss the performance results at each step. At last, section 6 concludes the work and presents future perspectives.

2. Related works

Aspect terms extraction task, applied to textual data in different languages and domains, is widely considered by the research community and remains a highly challenging field. We observe two main approaches: supervised and unsupervised ones. They include three main categories: rule-based extractors, topic modeling algorithms and supervised machine learning models.

Among unsupervised approaches, Bu et al. [24] use double propagation along with the syntactic relationships between attributes (for example ”big”) and nouns (for example ”screen”) to deal with aspect extraction. This approach does not require human-labeled data. However, it lacks precision due to the complexity of setting all the rules necessary to appoint the syntactic relations of natural language’s flexibility.

Rule-based methods consist of identifying linguistic feature patterns in opinionated texts based on the grammatical classes of the terms and the syntactic relations between them. The work of Hu and Liu [8] is one of the pioneers works in this category considering frequent nouns and noun phrases as aspects. Their work has been improved by Popescu and Etzioni [23] using the product class as prior information, and by Blair-Goldensohn et al. [3] who defined filtering rules to identify only nouns and noun phrases in sentiment bearing sentences. Aspect and sentiment relations have also been considered in [38] and [10]. The main problem of such approaches is the lack of adaptation requiring changing the rules whenever the domain of the application changes. They are also highly dependent on human expertise.

In the last decades, a growing interest has been raised in using supervised machine learning models for aspect extraction, where the problem is modelled as a sequence labelling one. For that, authors in [9] adopted the Conditional Random Fields (CRF) model. Recently, different deep network architectures have been defined based on LSTM [19], Attention [17], CNN [35], transformers [28], soft retrieval process [39] and constituency parsing [36]. The performance of those models depends on the availability of a sufficiently large labelled training dataset, which may limit their application to low-resource languages.

The lack of labelled data to train the supervised ATE model has been addressed in [16] through data augmentation. Chen and Qian in [5] used soft prototypes learned on internal or external data. In this work, we consider the extreme situation where no in-domain labeled data is available. We rely on a pseudo labelling step to enable supervised model training and an active learning process to enhance the model via label correction and data enrichment iteratively.

Active Learning (AL) represents also an alternative to improve the performance of learning models when manual annotation becomes too costly. It aims to guide the labelling process by selecting the most useful instances to increase performance. These instances are then exposed to the oracle for annotations. Thus, the number of instances needed for learning to maintain high performance is minimized.

Active learning (AL) was largely adopted in many machine learning areas, such as classification, clustering and regression. While there have been more efforts on active learning for text classification, in sequence labelling, AL has been considered mainly for Named Entity Recognition (NER) [32] and Part-Of-Speech tagging (PoS) [27] tasks.

In this context, AL was first proposed with classic machine learning models for training e.g. [31, 21]. Then recently, AL became more attractive when combined with deep learning models [30, 26]. Authors in [33] demonstrated that when deep learning models are used in AL, they are more efficient computationally and reduce the annotation cost.

Most of the previous works belong to pool-based active learning. In this setting, the initial annotated dataset is iteratively enriched with a subset of samples selected from an available pool of unlabelled data where an oracle is asked for labels. In this paper, we consider a slightly different setting where active learning is adopted for data enrichment and label correction. In the latter, an oracle is asked to check the correctness of uncertain labels of samples selected from the initial dataset.

The core of active learning is the query strategy that defines the instance selection criteria. Uncertain sampling is among the most used strategies in the pool-based AL setting [29, 6], where instances to be labelled are selected from the set of the most uncertain ones. We adapt this instance selection criteria to fit the sequence labelling setting inherent to ATE and handle label imbalance.

Prodigy [1] is an annotation tool that trains and evaluates models for specific automatic natural language processing tasks, namely named entity recognition, text classification and PoS annotation. It employs active learning to limit annotation costs and uses user interactivity to facilitate annotation and learning. However, it is a non-open-source commercial tool.

3. Problem Definition

Given a fixed input vocabulary of words σ, an output set of K tags T and an input sequence $x(x_1, x_2, ..., x_M)$ consisting of M words from σ, in the sequence labelling task, our aim is to learn a function $h : \sigma \rightarrow T$ that assigns a tag from T to each element of the input sequence x. The function h maps the input sequence to the most likely sequence of tags $t^* = (t_1^*, ..., t_M^*)$ where $t \in K$. That is, we seek to find::

$$t^* = argmax_{t \in T} P(t|x) \tag{1}$$

In a supervised setting, the mapping function h is learned by training a learning model over a dataset of N labeled sequences in the form of pairs of sequences $(x^{(i)}, t^{(i)})$ for $i \in 1, ..., N$.

We consider aspect terms extraction (ATE) as a sequence labelling problem where each token (word) is tagged as an aspect or not. Note that we consider both multi-word and single-word aspects and use the IOB2 tagging format [25] for this purpose. In this format, three tags are defined: "I" denotes that the token is inside the aspect expression, "O" means that the token is not an aspect and "B" expresses that the token is the beginning of an aspect expression.

4. Aspect Extraction from French Reviews

We address the problem of data scarcity in fine-grained explicit aspect extraction from online product reviews, using French as an example of a low-resource language for this task where only a few labeled datasets are available.

As shown in figure 1, we propose a three-step process. First, we pseudo-label the data through cross-domain learning to help avoid inconsistency of aspects from target and source domains using a lightweight model. Then, we train an ATE deep learning model on the previously labelled data under a supervised setting. In the last step, we use active learning to deal with label uncertainty.

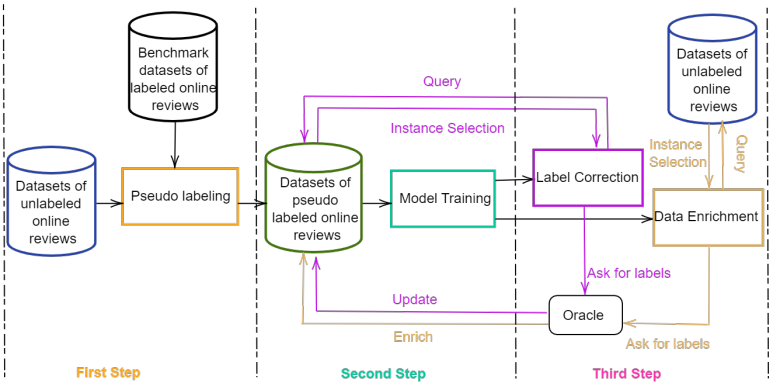

Fig. 1. Overview of the aspect extraction process consisting of a pseudo-labelling step (in orange), a model training step (in green) and a last iterative active learning step composed of two subtasks related to label correction (in violet) and data enrichment (in brown).

4.1. Lightweight Data Pseudo-labelling

Pseudo-labelling has been proposed as a preprocessing step for data annotation in low-resource scenarios where a large annotated dataset for training is not available [13]. To this end, we rely on Conditional Random Fields (CRF) [12].

CRF is a probabilistic framework for segmenting and labelling sequential data such as text. The CRF model predicts the label of each token in the context of its neighbouring tokens. In NLP tasks, the CRF model is graphically represented as a linear chain that captures dependencies with immediate neighbouring tokens.

Formally, given an input sequence $t = t_1, , t_n$ where t_i is the feature vector of the i^{th} token (word) and its corresponding sequence of labels $l = l_1, , l_n$. The possible label sequences for t is denoted by $\lambda(t)$. The probabilistic model of CRF defines a family of conditional probability $p(l|t; W, b)$ over all possible label sequences l given t as follows:

$$p(l|t; W, b) = \frac{\prod_{i=1}^{n} \psi_i(l_{i-1}, l_i, t)}{\sum_{l' \in \lambda(t)} \prod_{i=1}^{n} \psi_i(l'_{i-1}, l_i, t)} \tag{2}$$

where $\psi_i(l', l', t) = exp(W^T_{l',l} t_i + b_{l',l})$ are potential functions, $W^T_{l',l}$ and and $b_{l',l}$ are the weight vector and bias corresponding to label pair (l', l) respectively.

Inputs to the CRF model are vectors of features representing different characteristics of the token. These features include the word text, whether it is capitalized, its length, punctuation, or whether it is a digit token. We also included the Part of Speech (PoS) tag. Moreover, we enriched the feature vector with the same information about the immediate k-hop neighbour tokens.

It is worth noting that CRF is a supervised learning model, and thus labels are required for training. To handle label scarcity, we train the CRF through cross-domain learning, where predictions are made in a different domain from the training one. Furthermore, to make our CRF model more domain-agnostic, we omitted the word text feature for tokens of type "Noun," "PropN," and "Verb" since we consider that they are most likely to be domain-dependent.

4.2. Aspect Extraction using deep learning model

We used the BiLSTM-CNN-CRF deep learning model for aspect extraction, which has shown state-of-the-art performance in sequence tagging tasks [20]. The model consists of two embedding layers, a CNN layer, a bidirectional LSTM layer, and a CRF layer for outputting predicted labels (see Figure 2). We trained the model on the labeled dataset obtained from the pseudo labelling step using a split of 80/20 for training and validation. We optimized the model using the Adam optimizer and applied early stopping to prevent overfitting. The model achieved competitive

results compared to existing methods and demonstrated potential for adapting to low-resource languages with the use of pseudo labelling.

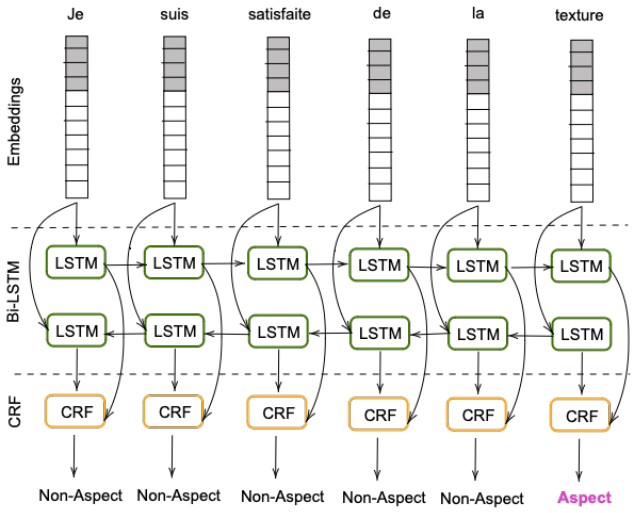

Fig. 2. BiLSTM-CNN-CRF Architecture.

4.3. Labelling Correction through Active Learning

Active learning reduces label uncertainty from the pseudo-labelling step by selectively querying the most informative instances for human annotation. The process is iterative and takes an initial, labeled dataset and large unlabelled dataset. A query strategy function is used to select instances from the unlabelled dataset, such as uncertainty sampling or diversity sampling. The selected instances are labeled by the oracle and added to the labeled dataset. This process is repeated iteratively until the required level of performance improvement is achieved, as determined by a stopping criterion. Many strategies for querying the unlabelled data instances are explained in [11], including entropy-based sampling, margin sampling, and committee-based sampling. (see Algorithm 1 and Algorithm 2).

In this work, we adapted the standard active learning process in two ways:

- In addition to pool querying in each iteration, we enable label correction from the training set using the defined query strategy. The idea is to use the oracle annotations to assess detected uncertainties resulting from erroneous labels due to pseudo-labelling or model misclassification.
- The instance selection process is implemented to fit the sequence labelling task and handle class imbalance. We used uncertainty sampling based on the Least Confidence (LC) strategy [15], where instances with the least confidence in their most likely label are selected. The uncertainty is defined as $1 - max(p)$, where p is the categorical probability distribution over labels. Because text instances consist of sequences of tokens, the formulation is updated by computing the average of token-level uncertainties, as shown in equation 3.

$$LC = 1 - \frac{1}{N}\sum_{i=1}^{N} max(p_i) \tag{3}$$

The uncertainty score is further adjusted as in equation 4 by omitting all tokens which have not been predicted as an aspect (i.e. with an ”O” tag) to reduce bias related to the dominance of non-aspect class and thus improve

the informativeness of selected samples towards aspect tokens since the primary goal of the AL process is to accelerate model learning and to minimize annotation effort.

$$LC = 1 - \frac{1}{N_{BI}} \sum_{i \in C_{BI}} max(p_i) \tag{4}$$

where $C_{BI} = \{i/\hat{y}_i \in \{'Aspect'\}, 1 \leq i \leq N\}$, $N_{BI} = |C_{BI}|$ and $\hat{y}_i$ is the label predicted for the i^{th} token by the learning model.

Algorithm 1 Custom Active Learning Algorithm

Data:
L – Training set of labeled sequences
U – Pool set of unlabelled sequences
$M(.,\theta_M)$ – Learning model
θ_{LC}– Query strategy function
B– query batch size
repeat
//Train model M using the training set L
$\theta_M \leftarrow train(M, L)$

//Labels correction
//Select the least confident from L and ask for labels
$I_B, L \leftarrow query\_and\_annotate(\theta_{LC}, L, M, B)$
//Update L with labeled instances I_B
$L \leftarrow L \cup I_B$

//Dataset enrich:
//select the least confident from U and ask for labels
$I_B, U \leftarrow query\_and\_annotate(\theta_{LC}, U, M, B)$
//Add the labeled batch of instances I_B *to L*
$L \leftarrow L \cup I_B$
until Stop criteria is met

Algorithm 2 *query_and_annotate* Function

Data:
X – Set of sequences
θ_{LC} – Query strategy function
$M(.,\theta_M)$ – Learning model
B– Query batch size

//Predict sequences of labels to X with M
$\hat{Y} \leftarrow predict(X; M, \theta_M)$

$I_B \leftarrow \oslash$
for b in B :
//Query for the least confident samples from X
$x_b \leftarrow argmax_{x \in X} \theta_{LC}(x)$
//Ask annotator for label
$y_b^* \leftarrow label(x_b)$
//Update samples batch
$I_B \leftarrow I_B \cup \{(x_b, y_b^*)\}$
//Update the input set
$X \leftarrow X - \{x_b\}$
end for

return I_B, X

5. Experiments and Evaluation

5.1. Data description

We evaluated our proposed process using two datasets: a French reviews dataset with aspect-level annotations from the SemEval-2016 Task 5 "Aspect-Based Sentiment Analysis" competition, and a large unlabelled dataset scraped from the web. The SemEval dataset is the only known annotated resource for aspect extraction from French reviews and was annotated with relevant entities, aspects, and polarity values [2]. The SemEval benchmark dataset consists of

	Restaurants	Musems
Number of reviews	Training Data	
	337	-
Number of reviews	Test Data	
	120	162

Table 1. SemEval 2016 French datasets Statistics

	Corpus	
	Hightech Products	Beauty Products
Training	4000	4000
Validation	200	190
Pool	2000	2000
Test	303	300

Table 2. Unlabelled web-scraped datasets description

one training dataset of 337 French restaurant customer reviews and two validation datasets for in-domain validation (120 reviews from the restaurant domain). The second dataset is used for out-of-domain validation and is composed of 162 reviews from the museum's domain. Table 1 summarizes the information about the SemEval data.

5.1.1. Web-scrapped dataset of online reviews

To evaluate the performance of our proposed method in the Hightech and Beauty domains, we manually annotated a set of 300 and 303 reviews, respectively. These reviews were collected by scraping French customers' reviews from

different online shopping websites as described in [4], and served to construct both the initial training dataset for the BiLSTM-CNN-CRF model and the pool set used in the active learning set.

5.2. Evaluations metrics

The performance at each step of the proposed model was measured using the precision, the recall, and the F1 score [34]. The scores were computed as follows for each class:

$$Prec = \frac{TP}{(TP + FP)}, Recall = \frac{TP}{(TP + FN)}, F1 = 2.\frac{PrecRecall}{(Prec + Recall)}$$

where TP, FP, TN and FN be the true positives, false positives, true negatives and false negatives. We merged classes 'I' and 'B' (i.e. no distinction was made between 'B' and 'I' classes) to get global performance scores for aspect extraction.

5.3. Pseudo-labelling Evaluation

As described in section 4.1, we defined a lightweight CRF to pseudo-label unlabelled data using cross-domain learning. The CRF model was first trained and got its parameters tuned on the labelled benchmark datasets from the SemEval competition. The training is done with the L1 regularization coefficient set to 0.5, the coefficient for L2 set to 0.1 regularization, the maximum number of iterations equal to 50 and the nonlinear optimization algorithm "lbfgs" as tuned parameters.

	Precision		Recall		F1-score	
Source → *Target*	Version 1	Version 2	Version 1	Version 2	Version 1	Version 2
Restaurants → *Restaurants*	0.714	0.428	0.563	0.632	0.63	0.51
Restaurants → *Musems*	0.597	0.448	0.252	0.326	0.354	0.377
Restaurants → *HighTech*	0.361	0.531	0.249	0.447	0.29	0.40
Restaurants → *Beauty*	0.509	0.602	0.334	0.463	0.432	0.519

Table 3. Pseudo-labelling Evaluation

Table 3 illustrates the performance of the applied CRF to three different configurations using as target domains the datasets of Museums reviews, Hightech products reviews and Beauty products reviews, respectively. We compare the approach using all the descriptors of CRF (Version 1) to the approach that omits the "literal value" descriptor from terms having as morphosyntactic tag "Noun" or "Verb" or "PropN"(version 2).

The results confirm that the second approach facilitates cross-domain transfer learning. With the first approach, the CRF model performance decreases considerably when the target domain is different from the source domain learning domain. However, the second approach achieves significantly better performances with an F1-score on the configuration "Restaurants → Beauty" comparable to that of "Restaurants → Restaurants".

This approach (Version 2) balances the performance between the source and target domains and offers an improvement of 2.3%, 11%, and 8.7% in terms of F1-score respectively on the corpus of Museums reviews, Hightech products reviews and Beauty products reviews. It should be noted that two subsets of corpora (303 of tech device reviews and 300 device reviews Beauty products) have been manually labelled by an expert to form the "ground truth" necessary to assess labelling, the BiLSTM-CNN-CRF model as well as the performance of Active Learning.

5.4. Active learning Evaluation

We define the BiLSTM-CNN-CRF as the learning model for aspect detection. We initialize the word embeddings using the pre-trained Glove embeddings with 300 dimensions (GloVe.840B.300d) released by [22]. The char-level embeddings are randomly initialized from the uniform distribution $U(-0.5, 0.5)$. As in [20], we set the convolutional layer's number and size to 3 and 30, respectively. The hidden representation of the BiLSTM layer is set to 200, and we applied a dropout on the LSTM outputs. The dropout rate is empirically set as 0.5.

We used stochastic gradient descent (SGD) for parameters optimization with a momentum of 0.9 and a learning rate of 10^{-2}. We set the batch size to 32.

		Hightech Products Corpus			Beauty Products Corpus		
AL Setting	Query Strategy	Precision	Recall	F1-score	Precision	Recall	F1-score
LCR=0% , DER=0%*	—-	0.505	0.372	0.429	0.699	0.405	0.513
LCR=10% , DER=0%	Uncertainty Sampling	**0.465**	**0.452**	**0.458**	0.684	**0.476**	**0.561**
	Random Sampling	0.455	0.438	0.446	**0.686**	0.448	0.542
LCR=20% , DER=0%	Uncertainty Sampling	**0.487**	**0.513**	**0.50**	**0.724**	**0.547**	**0.623**
	Random Sampling	0.483	0.464	0.473	0.714	0.48	0.575
LCR=30% , DER=0%	Uncertainty Sampling	**0.550**	0.495	**0.52**	**0.718**	**0.625**	**0.668**
	Random Sampling	0.474	**0.534**	0.502	0.688	0.493	0.574
LCR=30% , DER=10%	Uncertainty Sampling	**0.552**	**0.634**	**0.551**	**0.748**	**0.60**	**0.666**
	Random Sampling	0.536	0.511	0.523	0.724	0.514	0.601
LCR=30% , DER=20%	Uncertainty Sampling	**0.596**	**0.586**	**0.591**	**0.774**	**0.631**	**0.695**
	Random Sampling	0.586	0.5	0.54	0.751	0.592	0.561

* This setting refers to results before the Active Learning step (No data enrichment and label correction)

Table 4. Active learning results on the Hightech and Beauty test datasets in terms of precision, recall, and F1-score metrics. Different setting are considered with different label correction (LCR) and data enrichment (DER) rates.

We trained the BiLSTM-CNN-CRF model over 30 epochs and performed 10 active learning cycles. We conducted two sets of experiments: (1) focusing on label correction and (2) considering both label correction and data enrichment. We increased the Label Correction Rate (LCR) from 10% to 30% and set the Data Enrichment Rate (DER) to 10% and 20%. After each AL cycle, the model was retrained for five more epochs.

Table 4 summarizes the results of the experiments using the uncertainty sampling and random sampling strategies. The uncertainty sampling strategy outperformed the random strategy in terms of F1-scores for all the experiments. The results also show that active learning with label correction and data enrichment leads to higher F1-scores compared to not using these techniques. Increasing the LCR improved the model's precision, recall, and F1-scores for both Beauty and Hightech datasets.

The iterative label correction process improved the model's performance, resulting in higher F1-scores for both domains. Overall, active learning with label correction and data enrichment improved the model's performance by around 16.9% and 18.2% for Hightech and Beauty reviews, respectively, compared to not using these techniques. This highlights the effectiveness of active learning for improving the performance of the BiLSTM-CNN-CRF model on low-resource language datasets.

The tables 5 and 6 illustrate the confusion matrices respectively for the Beauty products corpus and the Hightech products corpus using active learning. We observe that 428 out of a total of 678 aspects were correctly labelled by the model on the Beauty products corpus and 450 out of a total of 767 aspects were identified for Hightech products corpus. The difference in performance can be explained by the complexity of the field of Hightech products due to

Beauty products Corpus		True Class	
		Non-aspect	Aspect
Predicted Class	Non-aspect	8495	250
	Aspect	125	428
	Total number of Aspects		678

Table 5. Confusion Matrix on Beauty Corpus

Hightech products Corpus		True Class	
		Non-aspect	Aspect
Predicted Class	Non-aspect	17229	317
	Aspect	304	450
	Total number of Aspects		767

Table 6. Confusion Matrix on Hightech Corpus

the heterogeneity of the terms aspects used according to the category of objects considered (sound, image, battery, storage, accessories, etc.).

6. Conclusion

In this paper, we proposed an end-to-end process using active learning to improve explicit aspect extraction for low-resource language fine-grained labelling. We used a web-scraped dataset of French reviews on Beauty products and electronic devices to run the tests. Our approach consists of a 3-step process, starting with a pseudo-labelling to address the scarcity of annotated data for ATE via cross-domain learning by omitting domain-related features during the learning process. This step is followed by an aspect-detection model training, for instance, the CNN-BilSTM-CRF model and a final active learning step to correct pseudo-labelling misclassification. Active learning reduces manual annotation costs and addresses label imbalance via the proposed adapted query selection process. Our approach is model-agnostic, making it compatible with various sequence labeling learning models. Experiments show that active learning significantly improves the learning model's performance, achieving more than 30% corrections to initial labels. Future work will adapt the proposed end-to-end process to enable aspect terms extraction on multilingual datasets. Moreover, we plan to include an additional processing step to categorize the extracted aspect terms, even with minimal or no prior knowledge.

References

[1] , . Prodigy · an annotation tool for ai, machine learning nlp. https://prodi.gy.

[2] Apidianaki, M., Tannier, X., Richart, C., 2016. Datasets for aspect-based sentiment analysis in french, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 1122–1126.

[3] Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G., Reynar, J., 2008. Building a sentiment summarizer for local service reviews .

[4] Boudabous, M., Pappa, A., 2021. Webt-idc: A web tool for intelligent dataset creation a use case for forums and blogs, in: Procedia Computer Science. Elsevier. volume 192, pp. 1051–1060.

[5] Chen, Z., Qian, T., 2020. Enhancing aspect term extraction with soft prototypes, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 2107–2117.

[6] Culotta, A., McCallum, A., 2005. Reducing labeling effort for structured prediction tasks, in: AAAI, pp. 746–751.

[7] Gries, S., Berez, A., 2017. Linguistic Annotation in/for Corpus Linguistics. pp. 379–409.

[8] Hu, M., Liu, B., 2004. Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168–177.

[9] Jakob, N., Gurevych, I., 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields, in: Proceedings of the 2010 conference on empirical methods in natural language processing, pp. 1035–1045.

[10] Kobayashi, N., Iida, R., Inui, K., Matsumoto, Y., 2006. Opinion mining on the web by extracting subject-aspect-evaluation relations., in: AAAI spring symposium: Computational approaches to analyzing Weblogs, pp. 86–91.

[11] Kumar, P., Gupta, A., 2020. Active learning query strategies for classification, regression, and clustering: A survey. Journal of Computer Science and Technology 35, 913–945.

[12] Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data .

[13] Lee, D.H., 2013. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. ICML 2013 Workshop : Challenges in Representation Learning (WREPL) .

[14] Leech, G., 1997. Linguistic Information from Computer Text Corpora. Routledge, London and NY. chapter Introducing corpus annotation. pp. 1 – 19.

[15] Lewis, D.D., Gale, W.A., 1994. A sequential algorithm for training text classifiers, in: SIGIR'94, Springer. pp. 3–12.
[16] Li, K., Chen, C., Quan, X., Ling, Q., Song, Y., 2020. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online.
[17] Li, X., Bing, L., Li, P., Lam, W., Yang, Z., 2018. Aspect term extraction with history attention and selective transformation, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18).
[18] Liao, X., Zhao, Z., 2019. Unsupervised approaches for textual semantic annotation, a survey. ACM Computer. Surv. 52 4, Article 66 (August 2019), 45 pages.
[19] Liu, P., Joty, S., Meng, H., 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings, in: Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 1433–1443.
[20] Ma, X., Hovy, E., 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.
[21] Marcheggiani, D., Artières, T., 2014. An experimental comparison of active learning strategies for partially labeled sequences, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar. pp. 898–906.
[22] Pennington, J., Socher, R., Manning, C.D., 2014. Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543.
[23] Popescu, A.M., Etzioni, O., 2007. Extracting product features and opinions from reviews, in: Natural language processing and text mining. Springer, pp. 9–28.
[24] Qiu, G.; Liu, B.B.J., Chen, C., 2011. Opinion word expansion and target extraction through double propagation, in: Computational Linguistics, p. 9–27.
[25] Ramshaw, L.A., Marcus, M.P., 1999. Text chunking using transformation-based learning, in: Natural language processing using very large corpora. Springer, pp. 157–176.
[26] Ren, P., Xiao, Y., Chang, X., Huang, P.Y., Li, Z., Gupta, B.B., Chen, X., Wang, X., 2021. A survey of deep active learning. ACM Computing Surveys (CSUR) 54, 1–40.
[27] Ringger, E., McClanahan, P., Haertel, R., Busby, G., Carmen, M., Carroll, J., Seppi, K., Lonsdale, D., 2007. Active learning for part-of-speech tagging: Accelerating corpus annotation, in: Proceedings of the Linguistic Annotation Workshop, pp. 101–108.
[28] Santos, B.N.D., Marcacini, R.M., Rezende, S.O., 2021. Multi-domain aspect extraction using bidirectional encoder representations from transformers. IEEE Access 9, 91604–91613.
[29] Scheffer, T., Decomain, C., Wrobel, S., 2001. Active hidden markov models for information extraction, in: International Symposium on Intelligent Data Analysis, Springer. pp. 309–318.
[30] Schröder, C., Niekler, A., 2020. A survey of active learning for text classification using deep neural networks. arXiv:2008.07267 .
[31] Settles, B., 2009. Active learning literature survey .
[32] Shelmanov, A., Liventsev, V., Kireev, D., Khromov, N., Panchenko, A., Fedulova, I., Dylov, D.V., 2019. Active learning with deep pre-trained models for sequence tagging of clinical and biomedical texts, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 482–489.
[33] Shen, Y., Yun, H., Lipton, Z.C., Kronrod, Y., Anandkumar, A., 2018. Deep active learning for named entity recognition, in: International Conference on Learning Representations.
[34] Van Rijsbergen, C.J., 1974. Foundation of evaluation. Journal of documentation 30, 365–373.
[35] Xu, H., Liu, B., Shu, L., Yu, P.S., 2018. Double embeddings and CNN-based sequence labeling for aspect extraction, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 592–598.
[36] Yang, Y., Li, K., Quan, X., Shen, W., Su, Q., 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). pp. 844–855.
[37] Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W., 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering , 1–20.
[38] Zhuang, L., Jing, F., Zhu, X.Y., 2006. Movie review mining and summarization, in: Proceedings of the 15th ACM international conference on Information and knowledge management, pp. 43–50.
[39] Zhuang Chen, T.Q., 2020. Enhancing aspect term extraction with soft prototypes, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics. p. 2107–2117.