Cross-Lingual Unsupervised Sentiment Classification with Multi-View Transfer Learning

Hongliang Fei, Ping Li
Cognitive Computing Lab
Baidu Research
1195 Bordeaux Dr, Sunnyvale, CA 94089, USA
10900 NE 8th St, Bellevue, WA 98004, USA
{hongliangfei,liping11}@baidu.com

Abstract

Recent neural network models have achieved impressive performance on sentiment classification in English as well as other languages. Their success heavily depends on the availability of a large amount of labeled data or parallel corpus. In this paper, we investigate an extreme scenario of cross-lingual sentiment classification, in which the low-resource language does not have any labels or parallel corpus. We propose an unsupervised cross-lingual sentiment classification model named multi-view encoder-classifier (MVEC) that leverages an unsupervised machine translation (UMT) system and a language discriminator. Unlike previous language model (LM) based fine-tuning approaches that adjust parameters solely based on the classification error on training data, we employ the encoder-decoder framework of a UMT as a regularization component on the shared network parameters. In particular, the cross-lingual encoder of our model learns a shared representation, which is effective for both reconstructing input sentences of two languages and generating more representative views from the input for classification. Extensive experiments on five language pairs verify that our model significantly outperforms other models for 8/11 sentiment classification tasks.

1 Introduction

Recent neural network models have achieved remarkable performance on sentiment classification in English and other languages (Conneau et al., 2017; Chen et al., 2018; He et al., 2019; Chen and Qian, 2019). However, their success heavily depends on the availability of a large amount of labeled data or parallel corpus. In reality, some low-resource languages or applications have limited labeled data or even without any labels or parallel corpus, which may hinder us from training a robust and accurate sentiment classifier.

To build sentiment classification models for low-resource languages, recent researchers developed *cross-lingual text classification* (CLTC) models (Xu and Yang, 2017; Eriguchi et al., 2018), which transfers knowledge from a resource-rich (source) language to a low-resource (target) language. The core of those models is to learn a shared language-invariant feature space that is indicative of classification for both languages. Therefore a model trained from the source language can be applied to the target language. Based on how the shared feature space is learned, there are three categories, namely word-level alignments (Andrade et al., 2015), sentence-level alignments (Eriguchi et al., 2018) and document level alignments (Zhou et al., 2016). Those models can well capture the semantic similarity between two languages. They, however, require parallel resources such as a bilingual dictionary, parallel sentences, and parallel Wikipedia articles. Such a limitation may prevent these models from being applicable in languages without any parallel resources.

Recently, there have been several attempts at developing "zero-resource" models (Ziser and Reichart, 2018; Chen et al., 2018; Chen and Qian, 2019). Most notably, Ziser and Reichart (2018) proposed a cross-lingual & cross-domain (CLCD) model that builds on pivot based learning and bilingual word embedding. Although CLCD does not directly need labeled data or parallel corpus, it requires *bilingual word embeddings* (BWEs) (Smith et al., 2017) that requires thousands of translated words as a supervised signal. Chen et al. (2018) developed an adversarial deep averaging network to learn latent sentence representations for classification, but it had an implicit dependency on BWEs (Zou et al., 2013) that requires pretraining on a large bilingual parallel corpus. Chen and Qian (2019) extended the

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5759–5771
July 5 - 10, 2020.

cross-lingual model in Chen et al. (2018) to multiple source languages by using the unsupervised BWEs (Lample et al., 2018b) and adding individual feature extractor for each source language, which eliminated the dependency on a parallel corpus. Nevertheless, their model is very sensitive to the quality of BWEs and performs poorly on distant language pairs such as English-Japanese, as illustrated in their experimental study.

In parallel, cross-lingual language models (LMs) trained from raw Wikipedia texts, such as multilingual BERT[1] (Devlin et al., 2019) and XLM (Conneau and Lample, 2019), have been prevalent in solving zero-shot classification problems (Wu and Dredze, 2019). Those models use the BERT-style Transformer (Vaswani et al., 2017) architecture simultaneously trained from multiple languages to construct a sentence encoder, and fine-tune the encoder and a classifier on labeled training data from the source language. Then the fine-tuned model is applied to the target language. The whole process does not require any labeled data or parallel corpus. However, under the "zero parallel resource" setting, the encoder trained from self-supervised masked language modelling within each language may not well capture the semantic similarity among languages, which could harm the generalization performance of fine-tuned models.

In this paper, we propose a sentiment classification model called *multi-view encoder-classifier* (MVEC) in an unsupervised setting, in which we only have monolingual corpora from two languages and labels in the source language. Different from previous language model (LM) based fine-tuning approaches (Devlin et al., 2019; Conneau and Lample, 2019) that adjust parameters solely based on the classification error of training data, we utilize the encoder-decoder network from unsupervised machine translation (UMT) (Lample et al., 2018a) to regularize and refine the shared latent space. In particular, the transformer-based encoder regularized by a language discriminator learns shared but more refined language-invariant representations, which are effective for both reconstructing sentences from two languages by the decoder and generating multi-view feature representations for classification from input documents. In our model, we construct two views from the encoder: (i) the encoded sentences in the source language; (ii) the encoded translations of the source sentences in the target language.

Our proposed MVEC is partially initialized by pretrained LMs (Conneau and Lample, 2019) but further fine-tuned to align sentences from two languages better, accurately predict labeled data in the source language and encourage consensus between the predictions from the two views. The full model is trained in an end-to-end manner to update parameters for the encoder-decoder, the language discriminator, and the classifier at each iteration.

Our contributions in this paper are as follows:

- We present an unsupervised sentiment classification model without any labels or parallel resource requirements for the target language. By designing a multi-view classifier and integrating it with pretrained LMs and UMT (Lample et al., 2018a), we build our model (MVEC) on a more refined latent space that is robust to language shift with better model interpretation compared to previous zero-shot classification works (Chen et al., 2018; Conneau and Lample, 2019).

- We extensively evaluate our model in 5 language pairs involving 11 sentiment classification tasks. Our full model outperforms state-of-the-art unsupervised fine-tuning approaches and partially supervised approaches using cross-lingual resources in 8/11 tasks. Therefore, our results provide a strong lower bound performance on what future semi-supervised or supervised approaches are expected to produce.

2 Related Work

2.1 Cross-Lingual Text Classification (CLTC)

CLTC aims to learn a universal classifier that can be applied to languages with limited labeled data (Bel et al., 2003; Dong and de Melo, 2019; Keung et al., 2019), which is naturally applicable for sentiment analysis. Traditional supervised methods utilize cross-lingual tools such as machine translation systems and train a classifier on the source language (Prettenhofer and Stein, 2010). The latest models used parallel corpus either to learn a bilingual document representation (Zhou et al., 2016) or to conduct cross-lingual model distillation (Xu and Yang, 2017).

In the unsupervised setting, Chen et al. (2018) learned language-invariant latent cross-lingual representations with adversarial training. Ziser

[1] https://github.com/google-research/BERT/blob/master/multilingual.md

and Reichart (2018) used pivot based learning and structure-aware DNN to transfer knowledge to low-resourced languages. In both papers, however, they have an implicit dependency on BWEs, which requires a bilingual dictionary to train. Chen and Qian (2019) was the first fully unsupervised approach using the unsupervised BWEs (Lample et al., 2018b) and multi-source languages with adversarial training. In contrast, our model is a multi-view classification model that is seamlessly integrated pretrained LMs (Conneau and Lample, 2019) and the encoder-decoder from UMT (Lample et al., 2018a) with adversarial training. Hence we learn a more fine-tuned latent space to better capture document-level semantics and generate multiple views to represent the input.

2.2 Unsupervised Machine Translation

UMT *does not* rely on any parallel corpus to perform translation, which lays a foundation for our approach. At the word-level, Lample et al. (2018b) built a bilingual dictionary between two languages by aligning monolingual word embeddings in an unsupervised way. At the sentence and document level, Lample et al. (2018a) proposed a UMT model by learning an autoencoder that can reconstruct two languages under both within-domain and cross-domain settings. Lample et al. (2018c) extended Lample et al. (2018a) with a phrase-based approach. Since we aim to learn more refined language-invariant representations for classification, it is natural to employ the encoder from a UMT system to generate multiple views of the input and enable knowledge transfer.

2.3 Multi-View Transfer Learning

The task of multi-view transfer learning is to simultaneously learn multiple representations and transfer the learned knowledge from source domains to target domains, which have fewer training samples. Generally, data from different views contains complementary information and multi-view learning exploits the consistency from multiple views (Li et al., 2019).

Our work is particularly inspired by Fu et al. (2015) and Zhang et al. (2019), both of which exploit the complementarity of multiple semantic representations with semantic space alignment. The difference is that we use an encoder-decoder framework to generate multiple views for input from the source language and enforce a consensus between their predictions. Furthermore, we introduce a language discriminator (Lample et al., 2018a) to encourage the encoder to generate language-invariant representations from the input.

3 Methodology

In this section, we will introduce our model's general workflow, including the details of each component and our training algorithm.

3.1 Problem Setup

Given monolingual text data $\{\boldsymbol{D}_{src},\ \boldsymbol{D}_{tgt}\}$ from both the source and target language with a subset of labeled samples $\{\boldsymbol{D}^L_{src},\ \boldsymbol{y}^L_{src}\}$ in the source language where $\boldsymbol{y}^L_{src}$ is a vector of class labels and $\boldsymbol{D}^L_{src} \subset \boldsymbol{D}_{src}$, the task aims to build a universal classification model $f(X;\boldsymbol{\theta}) \rightarrow y$ parameterized by $\boldsymbol{\theta}$ that can be directly applicable to unlabeled data in the target language, where X is an input document from any language and y is its class label. Note that in this paper we assume two languages share the same class types.

3.2 Model Architecture

Our proposed approach multi-view encoder classifier (MVEC) is composed of three components: an encoder-decoder, a language discriminator, and a classifier. Motivated by the success of unsupervised machine translation (UMT) in Lample et al. (2018a) and reconstruction regularization by an autoencoder in Sabour et al. (2017), we adopt the encoder-decoder framework from UMT (Lample et al., 2018a) and introduce self-reconstruction loss within one language and back-translation reconstruction loss across languages together with the normal loss from classification. For simplicity, we denote self-reconstruction loss as "within-domain loss" and back-translation reconstruction loss as "cross-domain loss" throughout the paper.

Although the encoder from UMT can generate a latent representation for input sentences/documents, there is still a semantic gap between the source and target language. Following Lample et al. (2018a); Chen et al. (2018), we enrich the encoder-decoder framework with a language discriminator that can produce fine-tuned latent representations to align latent representations from two languages better. Such representations are necessary to train a language-invariant classifier that is robust to the shift in languages.

In particular, as illustrated in Figure 1, the encoder is used to encode source and target docu-

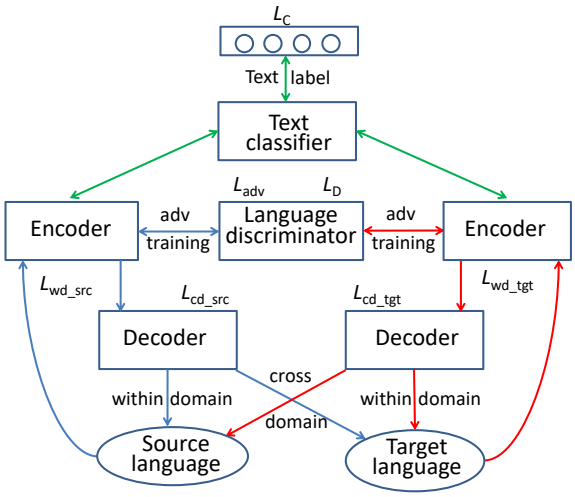

Figure 1: Multi-view encoder classifier (MVEC) architecture. Blue (red) lines indicate the message flow within the source/target language (across languages), respectively. Green lines indicate the message flow from the encoder to the text classifier. The encoder and decoder share the same parameters.

ments (a sequence of sentences) into a shared latent space, while the decoder is responsible for decoding the documents from the latent space to the source or the target language. Following Lample et al. (2018a), the encoder-decoder is shared for both languages (domains) and trained within-domain and cross-domain. The language discriminator aims to predict the language source for each document, and the classifier is trained to classify each document into predefined class labels.

Under the unsupervised setting, MVEC only observes unlabeled monolingual corpora from two languages and some labeled documents in the source language. The unlabeled monolingual data is normally sampled from the application domain, i.e., unlabeled product reviews or social media posts, which is used in both adopting pretrained LMs in the target domain and training UMT. As shown in Figure 1, unlabeled source and target data only pass through encoder-decoder and language discriminator, while the labeled source data pass all components in the system, including the sentiment classifier. For evaluation purposes, we may have labeled documents in the target language. However, they are only used during the test period. In the following subsections, we introduce each component of MVEC in detail.

3.3 Encoder-Decoder

Let $x^{(l)} = (x_1^{(l)}, x_2^{(l)}, \cdots, x_n^{(l)})$ denote the input document of n words from a particular language l, where $l \in \{src, tgt\}$. The encoder is a neural network $e_{\boldsymbol{\theta}_{enc}}(x^{(l)})$ parameterized by $\boldsymbol{\theta}_{enc}$ that produces a sequence of n hidden states $\mathbf{Z}^{(l)} = (\mathbf{z}_1^{(l)}, \mathbf{z}_2^{(l)}, \cdots, \mathbf{z}_n^{(l)})$ by using the corresponding word embedding for $x_i^{(l)}$, where $\mathbf{z}_i^{(l)}$ is the latent representation of $x_i^{(l)}$ in the shared latent space and $\boldsymbol{\theta}_{enc}$ are parameters of the encoder shared between two languages. The encoder could be a BiLSTM or a transformer (Vaswani et al., 2017). In this paper, we adopt the transformer, which has achieved enormous success in (e.g.,) recent text representation learning tasks (Devlin et al., 2019; Conneau and Lample, 2019).

Given $\mathbf{Z}^{(l)}$ as the input, the decoder $d_{\boldsymbol{\theta}_{dec}}(\mathbf{Z}^{(l)})$ generates the output sequence $y^{(l)} = (y_1^{(l)}, y_2^{(l)}, \cdots, y_k^{(l)})$. We use the same transformer based decoder as in Conneau and Lample (2019), parameterized by $\boldsymbol{\theta}_{dec}$. For simplicity, we will denote the encoder and decoder by $e(x^{(l)})$ and $d(\mathbf{Z}^{(l)})$ respectively instead of $e_{\theta_{enc}}(x^{(l)})$ and $d_{\theta_{dec}}(\mathbf{Z}^{(l)})$.

It is more likely for the encoder-decoder to merely memorize every input word one by one if there are no imposed constraints. To improve the robustness of encoder-decoder, we follow Lample et al. (2018a) to adopt the Denoising Auto-encoders (DAE) (Vincent et al., 2008), which recovers input from its corrupted version.

There are three ways to inject noise into the document including shuffle, dropout, and replacement by special words. In our model, we drop and replace every word with probabilities of p_d and p_b, respectively, and we slightly shuffle the input document by implementing random permutation σ on the input document, where p_d and p_b can be viewed as hyper-parameters for controlling noise levels. In our design, the permutation σ satisfies the condition $|\sigma(i) - i| \leq k, \ \forall i \in \{1, \cdots, n\}$, where n is the length of input document and k is another hyper-parameter.

Note that the noise model is only applied to unlabeled data used for training the encoder-decoder and the discriminator, while labeled data will keep its originality for all components training. We use $G(.)$ to denote a stochastic noise model, which takes input document $x^{(l)}$ and generates $G(x^{(l)})$ as a randomly sampled noisy version of $x^{(l)}$.

To incorporate the encoder-decoder as regularization components, we follow Lample et al. (2018a) to consider both within-domain and cross-domain objective functions. The first objective

function aims to reconstruct a document from a noisy version of itself within a language, whereas the second (cross-domain) objective function targets to teach the model to translate an input document across languages. Specifically, given a language $l \in \{src, tgt\}$, the within-domain objective function can be written as:

$$R_{wd}(\boldsymbol{\theta}_{ed}, l) = \mathbb{E}_{x \sim \boldsymbol{D}_l, \hat{x} \sim d(e(G(x)))}[\Delta(x, \hat{x})] \quad (1)$$

where $\boldsymbol{\theta}_{ed} = [\boldsymbol{\theta}_{enc}, \boldsymbol{\theta}_{dec}]$, $\hat{x} \sim d(e(G(x)))$ is a reconstruction of the corrupted version of x sampled from the monolingual dataset $\boldsymbol{D}_l$, and Δ is the sum of token-level cross-entropy loss to measure discrepancy between two sequences.

Similarly, we consider teaching the encoder-decoder to reconstruct x in one language from a translation of x in the other language, leading to the following cross-domain objective function:

$$R_{cd}(\boldsymbol{\theta}_{ed}, l_1, l_2) = \mathbb{E}_{x \sim \boldsymbol{D}_{l_1}, \hat{x} \sim d(e(T(x)))}[\Delta(x, \hat{x})] \quad (2)$$

where $(l_1, l_2) \in \{(src, tgt), (tgt, src)\}$ and $T(.)$ is the current UMT model applied to input document x from language l_1 to language l_2.

3.4 Language Discriminator

Cross-lingual classifiers work well when their input produced by the encoder is language-invariant, as studied in Chen et al. (2018). Thus, we prefer our encoder to map input documents from both languages into a shared feature space independent of languages. To achieve this goal, we follow Chen et al. (2018); Lample et al. (2018a) and introduce a language discriminator into our model, which is a feed-forward neural network with two hidden layers and one softmax layer to identify the language source from the encoder's output. In particular, we minimize the following cross-entropy loss function:

$$L_D(\boldsymbol{\theta}_D | \boldsymbol{\theta}_{enc}) = -\mathbb{E}_{(l, x^{(l)})}[\log P_D(l | e(x^{(l)})] \quad (3)$$

where $\boldsymbol{\theta}_D$ denotes parameters of the discriminator, $(l, x^{(l)})$ corresponds to language and document pairs uniformly sampled from monolingual datasets, and $P_D(.)$ is the output from the softmax layer. Meanwhile, the encoder is trained to "fool" the discriminator:

$$L_{adv}(\boldsymbol{\theta}_{enc} | \boldsymbol{\theta}_D) = -\mathbb{E}_{x^{(l_i)} \sim D_{l_i}}[\log P_D(l_j | e(x^{(l_i)})] \quad (4)$$

with $l_j = l_1$ if $l_i = l_2$, and vice versa.

3.5 Multi-view Classifier

Thus far, we have described how we obtain a language-invariant latent space to encode two languages, which may not be sufficient to generalize well across languages if we simply train a classifier on the encoder's output for the source language (Chen et al., 2018). One key difference between Chen et al. (2018) and our work is that we use UMT (Lample et al., 2018a), which can generate multiple views for the input labeled documents from the source language. We can thereby benefit from multi-view learning's superior generalization capability over single-view learning (Zhao et al., 2017).

Particularly, we consider two views of input: (i) the encoded labeled documents from the source language; (ii) the encoded back-translations of the source documents from the target language. Our learning objective is to train the classifier to match predicted document labels with ground truth from the source language and to encourage two predictive distributions on the two views to be as similar as possible. We consider the following objective function:

$$L_C(\boldsymbol{\theta}_C, \boldsymbol{\theta}_{ed}) = \mathbb{E}_{(x,y)}[\Delta(y, P_{\boldsymbol{\theta}_c}(e(x))) + \underbrace{D_{KL}(P_{\boldsymbol{\theta}_c}(e(x)) \,||\, P_{\boldsymbol{\theta}_c}(e(T(x))))]}_{\text{Two views' consensus}} \quad (5)$$

where $(x, y) \sim \{\boldsymbol{D}^L_{src}, \boldsymbol{y}^L_{src}\}$, $D_{KL}(. \,||\, .)$ is KL Divergence to measure the difference between two distributions, y is the class label of input document x and θ_c are parameters of classifier. Following previous studies in text classification (Devlin et al., 2019), we use the first token's representation in the last hidden layer from the transformer encoder as the document representation vector. The classifier is a feed-forward neural network with two hidden layers and a softmax layer.

The final objective function at one iteration of our learning algorithm is to minimize the following loss function:

$$L_{all} = L_C + \lambda_{wd} \times (R_{wd\_src} + R_{wd\_tgt}) + \lambda_{cd} \times (R_{cd\_src} + R_{cd\_tgt}) + \lambda_{adv} \times L_{adv} \quad (6)$$

where $\lambda_{wd}, \lambda_{cd}, \lambda_{adv}$ are hyper-parameters to trade-off among within-domain loss, the cross-domain loss and the adversarial loss, respectively.

3.6 Training Algorithm

Our model relies on an initial translation machine $T^{(0)}$, which provides a translation from one lan-

guage to another for calculating the cross-domain loss in Eq. (2) and classifier loss in Eq. (5).

To accelerate the training, we initialize $T^{(0)}$ by pretraining a transformer-based UMT (Conneau and Lample, 2019) for certain steps with the same encoder-decoder architecture as our model on monolingual Wikipedia text. After pretraining, we use the pretrained encoder-decoder network to initialize our model and start training the classifier and the discriminator. Meanwhile, we refine the encoder and the decoder on monolingual data and labeled data from the source language.

During each training step, the optimization iterates from updating $\boldsymbol{\theta}_D$ in Eq. (3) to updating $\boldsymbol{\theta}_{ed}$ and $\boldsymbol{\theta}_C$ in Eq. (6). Note that if a batch of documents drawn from monolingual data are all unlabeled, then we suspend updating classifier parameters and only update the parameters of the language discriminator and encoder-decoder. In Algorithm 1, we provide a detailed procedure.

Algorithm 1 The proposed MVEC algorithm.

procedure TRAINING($\boldsymbol{D}_{src}$, $\boldsymbol{D}_{tgt}$, $\boldsymbol{y}^L_{src}$)
$\boldsymbol{D}_{src}$ and $\boldsymbol{D}_{tgt}$: monolingual datasets, $\boldsymbol{y}^L_{src}$: labels in the source language.
  $T^{(0)} \leftarrow$ pretrain a transformer based UMT using (Conneau and Lample, 2019);
  for $t = 0, \cdots, max\_epoch$ **do**
    Using $T^{(t)}$ to translate each document in a batch;
    $\boldsymbol{\theta}_D \leftarrow \text{argmin}\ L_D$ in Eq. (3) while fixing $\boldsymbol{\theta}_C$, $\boldsymbol{\theta}_{ed}$;
    $\boldsymbol{\theta}_C$, $\boldsymbol{\theta}_{ed} \leftarrow \text{argmin}\ L_{all}$ in Eq. (6) while fixing $\boldsymbol{\theta}_D$;
    Update $T^{(t+1)} \leftarrow \{e^{(t)}, d^{(t)}\}$;
  return θ_C, θ_{enc}
End procedure

4 Experiment

We conduct experiments on cross-lingual multi-class and binary sentiment classification using five language pairs involving 11 tasks. More specifically, English is always the source language, and the target languages are French, German, Japanese, Chinese, and Arabic, respectively.

4.1 Datasets

Amazon Review (French, German, Japanese). This is a multilingual sentiment classification dataset (Duh et al., 2011) in four languages, including English (en), French (fr), German (de), and Japanese (ja), covering three products (book, DVD, and music). For each product in each language, there are 2000 documents in each of the training and test sets. Each document contains a title, a category label, a review, and a 5-point scale star rating. Following Xu and Yang (2017); Chen and Qian (2019), we convert multi-class ratings to binary ratings by thresholding at 3-point. For each product, since the test set in English is not used, we combine the English training and test sets and randomly sample 20% (800) documents as the validation set to tune hyper-parameters, and use the rest 3200 samples for training. For each target language, we use the original 2000 test samples for comparison with previous methods. Unlike Chen et al. (2018); Chen and Qian (2019) that used labeled data in the target language for model selection, we only use the labels of reviews in the target language for testing. There are 105k, 58k, 317k, 300k unlabeled reviews for English, French, German and Japanese, respectively, which can be used as monolingual data to train the encoder-decoder of our model.

Yelp and Hotel Review (Chinese). This dataset is from two sources: (i) 700k Yelp reviews in English with five classes from Zhang et al. (2015), and (ii) 170k hotel reviews in Chinese segmented and annotated with five classes from Lin et al. (2015). Following the same setup in Chen et al. (2018), we split all Yelp reviews into a training set with 650k reviews and validation set with 50k reviews. The 650k review contents are also served as the monolingual training data for English. For Chinese hotel review data, we sample 150k reviews as the monolingual training set. The rest 20k reviews are treated as the test set.

Social Media Posts (Arabic). The BBN Arabic Sentiment dataset is from Mohammad et al. (2016). There are 1200 documents from social media posts annotated with three labels (negative, neutral, positive) in the data. The original dataset was split into half as training and the other half as testing. Since we do not need validation data in the target language to tune the model, we randomly sample 1000 documents as test data. For English resource, we still use Yelp reviews and follow the same split as the Chinese case, but convert 5 level reviews into 3 levels[2]. Also, we randomly sample

[2] 1,2 → negative, 3 → neutral, 4,5 → positive

161k sentences from the United Nations Corpus Arab subset (Ziemski et al., 2016) as unlabeled monolingual data for our model training.

4.2 Experiment Setting

For French, German and Japanese, we perform binary classification. For Chinese and Arabic, we perform multi-class classification.

Data Preprocessing. Following Lample et al. (2018c), we extract and tokenize monolingual data of each language using Moses (Koehn et al., 2007). Then we use the neural machine translation for rare words with subword units, named fastBPE (Sennrich et al., 2016) in three steps. In detail, BPE code is collected from the pretrained XLM-100 models (Conneau and Lample, 2019), then applied to all tokenized data and used to extract the training vocabulary. To constrain our model size, we only keep the top 60k most frequent subword units in our training set. Finally, we binarize monolingual data and labeled data for model training, validation and testing.

Pretraining Details. As mentioned earlier, our model depends on an initial translation machine to compute reconstruction loss and classifier loss. We leverage pretrained language models (Conneau and Lample, 2019) to initialize a transformer-based UMT (Lample et al., 2018a) and train it on Wikipedia text[3]. In particular, we sample 10 million sentences from each language pairs and use the XLM library[4] to train a UMT (Lample et al., 2018a) for 200K steps. The resulting encoder-decoder are used to initialize our model.

Regarding word embedding initialization, we use the embeddings obtained from the 1st layer of pretrained language models (Conneau and Lample, 2019), which has demonstrated better cross-lingual performance in a number of evaluation metrics over MUSE (Lample et al., 2018b).

Training Details. In our experiment, both encoder and decoder are 6 layer transformers with 8-head self-attention. We set both subword embedding and hidden state dimension to 1024 and use greedy decoding to generate a sequence of tokens. The encoder-decoder and classifier are trained using Adam optimizer (Kingma and Ba, 2015) with a learning rate of 10^{-5} and a mini-batch size of 32. We set the hidden dimension to 128 for both classifier and discriminator. For parameters of denoising auto-encoder, we set $p_d = 0.1$, $p_b = 0.2$ and $k = 3$ following Lample et al. (2018a). Finally, we perform a grid search for hyper-parameters on $\{0.5,1,2,4,8\}$ and set λ_{wd}, λ_{cd} to 1 and λ_{adv} to 4. To prevent gradient explosion, we clip the gradient L_2 norm by 5.0. Our approach is implemented in PaddlePaddle[5] and all experiments are conducted on an NVIDIA Tesla M40 (24GB) GPU.

Competing Methods. We have compared our method with several recently published results. Due to the space limit, we briefly introduce several representative baselines: LR+MT translated the bag of words from target language to source language via machine translation and then built a logistic regression model. BWE baselines rely on Bilingual Word Embeddings (BWEs), wherein 1-to-1 indicates that we are only transferring from English, while 3-to-1 means the training data from all other three languages. CLDFA (Xu and Yang, 2017) was built on model distillation on parallel corpora with adversarial feature adaptation technique. PBLM (Ziser and Reichart, 2018) used bilingual word embeddings and pivot-based language modeling for cross-domain & cross-lingual classification. MBERT (Devlin et al., 2019) and XLM-FT (Conneau and Lample, 2019) directly fine-tuned a single layer classifier based on pretrained LM multilingual BERT and XLM.

4.3 Experiment Results

In Table 1 and Table 2, we compare our method with others based on their published results or our reproduced results from their code. Our results are averaged based on 5 rounds of experiment with the standard deviation around 1%-1.5%. Following previous baselines, we do not report them here.

Our first observation from Table 1 is that our model and the fine-tuned multilingual LM MBERT (Devlin et al., 2019) and XLM-FT (Conneau and Lample, 2019) outperform all previous methods including the methods with cross-lingual resources for 8/9 tasks by a large margin, which indicates the huge benefit from pretrained LMs in the zero-shot setting. Compared with MBERT and XLM-FT, our model obtains better performance when the target language is more similar to the source language, for example, German and French, and one task in Japanese.

[3] http://dumps.wikimedia.org/

[4] www.github.com/facebookresearch/XLM

[5] http://www.paddlepaddle.org/

Approach	German (2) books	German (2) DVD	German (2) music	German (2) avg	French (2) books	French (2) DVD	French (2) music	French (2) avg	Japanese (2) books	Japanese (2) DVD	Japanese (2) music	Japanese (2) avg
With cross-lingual resources												
LR+MT	79.68	77.92	77.22	78.27	80.76	78.83	75.78	78.46	70.22	71.30	72.02	71.18
CR-RL[1]	79.89	77.14	77.27	78.10	78.25	74.83	78.71	77.26	71.11	73.12	74.38	72.87
Bi-PV[2]	79.51	78.60	82.45	80.19	84.25	79.60	80.09	81.31	71.75	75.40	75.45	74.20
CLDFA[3]	83.95	83.14	79.02	82.04	83.37	82.56	83.31	83.08	77.36	**80.52**	76.46	78.11
With implicit cross-lingual resources												
UMM[4]	81.65	81.27	81.32	81.41	80.27	80.27	79.41	79.98	71.23	72.55	75.38	73.05
PBLM[5]	78.65	79.90	80.10	79.50	77.90	75.65	75.95	76.50	-	-	-	-
Without cross-lingual resources												
BWE (1-to-1)	76.00	76.30	73.50	75.27	77.80	78.60	78.10	78.17	55.93	57.55	54.35	55.94
BWE (3-to-1)	78.35	77.45	76.70	77.50	77.95	79.25	79.95	79.05	54.78	54.20	51.30	53.43
MAN-MoE[6]	82.40	78.80	77.15	79.45	81.10	84.25	80.90	82.08	62.78	69.10	72.60	68.16
MBERT[7]	84.35	82.85	83.85	83.68	84.55	85.85	83.65	84.68	73.35	74.80	76.10	74.75
XLM-FT[8]	86.85	84.20	85.90	85.65	88.1	86.95	86.20	87.08	**80.95**	79.20	78.02	**79.39**
MVEC (Ours)	**88.41**	**87.32**	**89.97**	**88.61**	**89.08**	**88.28**	**88.50**	**88.62**	79.15	77.15	**79.70**	78.67

[1] Xiao and Guo (2013) [2] Pham et al. (2015) [3] Xu and Yang (2017) [4] Xu and Wan (2017)
[5] Ziser and Reichart (2018) [6] Chen and Qian (2019) [7] Devlin et al. (2019) [8] Conneau and Lample (2019)

Table 1: Prediction accuracy of binary classification in the test set for three language pairs. The highest performance is in bold, while the highest performance within the method group is underlined.

Approach	Chinese (5)	Arabic (3)
LR+MT	34.01	51.67
DAN	29.11	48.00
mSDA	31.44	48.33
ADAN	42.49	**52.54**
MBERT	38.85	50.40
XLM-FT	42.22	49.50
MVEC (Ours)	**43.36**	49.70

Table 2: Prediction accuracy of 5-class and 3-class classification tasks on the test set.

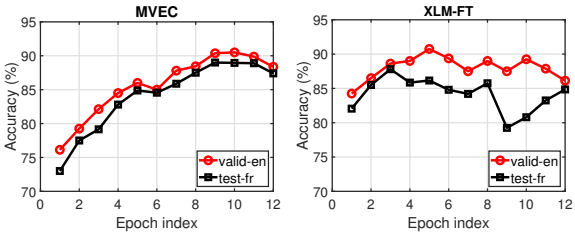

Figure 2: Validation and test accuracy w.r.t. training epochs for Amazon book review in French. Left: our method (MVEC). Right: XLM-FT.

In Table 2, we show the comparison between our method and a few other published results, including ADAN (Chen et al., 2018) and mSDA (Chen et al., 2012) for Chinese and Arabic languages in multi-class setting. Similarly, our model obtains slightly better accuracy in Chinese. Overall, built on top of the pretrained LMs and UMT, our full model achieves the state-of-the-art performance on 8/11 sentiment classification tasks, especially when the target language is more similar to the source language.

Moreover, we illustrate the effectiveness of encoder-decoder based regularization in reducing the language shift in the shared latent space. Intuitively, if the fine-tuned latent space is less sensitive to the language shift, the performance on validation sets and test sets should be highly correlated during training. In Figure 2, we report the average accuracy of both validation and test set w.r.t. training epochs over five runs on Amazon book review data in French.

From Figure 2, we observe that even though our model's best validation accuracy is lower than XLM-FT (Conneau and Lample, 2019) in English, it has more correlated accuracy curves than XLM-FT across English and French. For example, the validation accuracy of XLM-FT starts decreasing after epoch 10, while the test accuracy is still increasing. Such an observation shows that the latent representation learned solely from self-supervised objectives (e.g., masked language modeling) may not well capture the semantic similarity among languages. Hence the resulting classifier may work well in the source language but may not generalize to the target language. In contrast, our model sacrifices some accuracy in the source language but can select better models for the target language in a cross-lingual setting.

4.4 Ablation Study

To understand the effect of different components in our model on the overall performance, we con-

	German	French	Japanese	Chinese	Arabic
Full model:	**88.61**	**88.62**	**78.67**	**43.36**	**49.70**
w/o cross-domain loss:	83.22	82.40	72.05	35.74	42.80
w/o within-domain loss:	82.90	82.15	71.27	37.21	41.60
w/o adversarial training:	84.85	84.58	73.75	39.36	46.37
w/o two-views consensus:	86.21	86.18	75.25	40.95	46.77

Table 3: Ablation study on five language pairs.

duct an ablation study, as reported in Table 3. Clearly, the encoder-decoder trained either by the within-domain objective or cross-domain objective is the most critical. For Amazon data in three languages (German, French, Japanese), the model without cross-domain loss obtains prediction accuracy of 83.22%, 82.40%, and 72.05%, which gets decreased by 5%−7% compared with the full model. The performance is also significantly degraded when the adversarial training component is removed because the distribution of latent document representations is not similar between two languages. The two-views consensus component also has a significant effect on the performance of our model, with a performance drop up to 5 points for en-jp. Such a result verifies our claim that cross-lingual model benefits from training on multiple views of the input.

4.5 Case Study

To further explore the effectiveness of our approach, we visualize the encoder's output and the last layer before softmax for 10 randomly sampled Amazon reviews in English and their translations in French using Google Translation, as shown in Appendix A.2.

As seen in the lower-left panel of Figure 3, most red circles and black squares with the same indices are very close for our method but are distant for XLM-FT in the top-left. Such an observation implies that our encoder combined UMT and a language discriminator adequately maps the input into a shared language-invariant latent space while preserving semantic similarity. For the last layer before softmax, even though XLM-FT also generates reasonable representations to separate positive and negative reviews, the data points are scattered randomly. On the contrary, our model's output in the lower right panel of Figure 3 shows two more obvious clusters with corresponding labels that can be easily separated. One cluster in the left contains all of the positive documents, while the negative examples only appear on the right side.

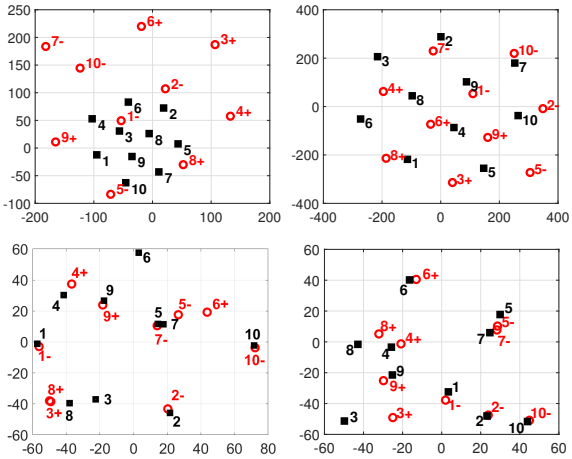

Figure 3: t-SNE visualizations of various layers of XLM-FT and MVEC for en-fr. Red circles and black squares indicate documents from English and their corresponding translations in the target language, respectively. Numbers indicate the document index and have a one-to-one mapping. +/- indicates labels and we only annotate English documents for simplicity. Top left: encoder output of XLM-FT. Top right: the last layer before softmax of XLM-FT. Lower left: encoder output of our method. Lower right: the last layer before softmax of our method.

5 Conclusion

In this paper, we propose a cross-lingual multi-view encoder-classifier (MVEC) that requires neither labeled data in the target language nor cross-lingual resources with the source language. Built upon pretrained language models, our method utilizes the encoder-decoder component with a language discriminator from an unsupervised machine translation system to learn a language-invariant feature space. Our approach departs from previous models that could only make use of the shared language-invariant features or depend on parallel resources. By constructing the fine-tuned latent feature space and two views of input from the encoder-decoder of UMT, our model significantly outperforms previous methods for 8/11 zero-shot sentiment classification tasks.

References

Daniel Andrade, Kunihiko Sadamasa, Akihiro Tamura, and Masaaki Tsuchida. 2015. Cross-lingual text classification using topic-dependent word probabilities. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 1466–1471, Denver, CO.

Núria Bel, Cornelis H. A. Koster, and Marta Villegas. 2003. Cross-lingual text categorization. In *Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL)*, pages 126–139, Trondheim, Norway.

Minmin Chen, Zhixiang Eddie Xu, Kilian Q. Weinberger, and Fei Sha. 2012. Marginalized denoising autoencoders for domain adaptation. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK.

Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Q. Weinberger. 2018. Adversarial deep averaging networks for cross-lingual sentiment classification. *Trans. Assoc. Comput. Linguistics*, 6:557–570.

Zhuang Chen and Tieyun Qian. 2019. Transfer capsule network for aspect level sentiment classification. In *Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL)*, pages 547–556, Florence, Italy.

Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 7057–7067, Vancouver, Canada.

Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann LeCun. 2017. Very deep convolutional networks for text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 1107–1116, Valencia, Spain.

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)*, pages 4171–4186, Minneapolis, MN.

Xin Dong and Gerard de Melo. 2019. A robust self-learning framework for cross-lingual text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6305–6309, Hong Kong, China.

Kevin Duh, Akinori Fujino, and Masaaki Nagata. 2011. Is machine translation ripe for cross-lingual sentiment classification? In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, pages 429–433, Portland, OR.

Akiko Eriguchi, Melvin Johnson, Orhan Firat, Hideto Kazawa, and Wolfgang Macherey. 2018. Zero-shot cross-lingual classification using multilingual neural machine translation. Technical report, arXiv:1809.04686.

Yanwei Fu, Timothy M. Hospedales, Tao Xiang, and Shaogang Gong. 2015. Transductive multi-view zero-shot learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(11):2332–2345.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL)*, Florence, Italy.

Phillip Keung, Yichao Lu, and Vikas Bhardwaj. 2019. Adversarial learning with contextual embeddings for zero-resource cross-lingual classification and NER. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1355–1360, Hong Kong, China.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, and et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, Vancouver, Canada.

Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018b. Word translation without parallel data. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, Vancouver, Canada.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018c. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5039–5049, Brussels, Belgium.

Yingming Li, Ming Yang, and Zhongfei Zhang. 2019. A survey of multi-view representation learning. *IEEE Trans. Knowl. Data Eng.*, 31(10):1863–1883.

Yiou Lin, Hang Lei, Jia Wu, and Xiaoyu Li. 2015. An empirical study on sentiment classification of chinese review using word embedding. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (PACLIC)*, Shanghai, China.

Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *J. Artif. Intell. Res.*, 55:95–130.

Hieu Pham, Thang Luong, and Christopher D. Manning. 2015. Learning distributed representations for multilingual text sequences. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing (VS@NAACL-HLT)*, pages 88–94, Denver, Co.

Peter Prettenhofer and Benno Stein. 2010. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1118–1127, Uppsala, Sweden.

Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. 2017. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3856–3866, Long Beach, CA.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berlin, Germany.

Samuel L. Smith, David H. P. Turban, Steven Hamblin, and Nils Y. Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, Toulon, France.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing (NIPS)*, pages 6000–6010, Long Beach, CA.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML)*, pages 1096–1103, Helsinki, Finland.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 833–844, Hong Kong, China.

Min Xiao and Yuhong Guo. 2013. Semi-supervised representation learning for cross-lingual text classification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1465–1475, Seattle, WA.

Kui Xu and Xiaojun Wan. 2017. Towards a universal sentiment classifier in multiple languages. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 511–520, Copenhagen, Denmark.

Ruochen Xu and Yiming Yang. 2017. Cross-lingual distillation for text classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1415–1425, Vancouver, Canada.

Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5429–5435, Macao, China.

Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems (NIPS)*, pages 649–657, Montreal, Canada.

Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. 2017. Multi-view learning overview: Recent progress and new challenges. *Inf. Fusion*, 38:43–54.

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2016. Cross-lingual sentiment classification with bilingual document representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berlin, Germany.

Michal Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The united nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*, Portorož, Slovenia.

Yftah Ziser and Roi Reichart. 2018. Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 238–249, Brussels, Belgium.

Will Y. Zou, Richard Socher, Daniel M. Cer, and Christopher D. Manning. 2013. Bilingual word embeddings for phrase-based machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1393–1398, Seattle, WA.

A Additional Details on Datasets

A.1 Summary Statistics of Labeled Datasets

For the Amazon Review dataset, we use the same test set as Duh et al. (2011) but transform them into binary labels for comparison with previous works. After transformation, the test set of each product category has equal number of positive and negative ratings (1000 vs 1000).

For the Yelp and Hotel Review dataset, we follow the same split as Chen et al. (2018) and keep the original rating. The test set contains 10k documents in total with around 2000 documents for each rating level.

The Arabic social media dataset contains 1000 test documents sampled from 1200 social media posts with about 400 documents for each rating level. Since Arabic data is not used for tuning parameters as validation set, we use more test samples than Chen et al. (2018).

A.2 Sampled Data for the Case Study

In Section 4.5, we randomly sample 10 Amazon book reviews in English, and translate them into French using Google Translation for case study. The sampled reviews and their *French translations* are as follows:

1. More than mitigated for this tote album that mixes some good ideas (the parodies of works of art) and scenes that only echo the previous albums lazily.

   Plus qu'atténué pour cet album cabas qui mêle quelques bonnes idées (les parodies d'oeuvres d'art) et des scènes qui ne font que faire écho aux albums précédents paresseusement.

2. What a disappointment, so dear for that. After the Gallic comeback, another album story to release an album. Beautiful pictures, some cool stuff (so the picture with all the characters) ... but

   Quelle déception, si chère pour ça. Après le retour des Gaulois, une autre histoire d'album pour sortir un album. De belles photos, des trucs sympas (donc la photo avec tous les personnages) ... mais

3. We obviously believe we know everything about the unspeakable horror of concentration camps. Well no; if it's a man; literally leaves no voice! Any comment seems inappropriate and for all

   Nous pensons évidemment que nous savons tout sur l'horreur indicible des camps de concentration. Eh bien non, si c'est un homme ne laisse littéralement aucune voix! Tout commentaire semble inapproprié et pour tous

4. "We who have survived", said Primo Levi, "are not good witnesses, because we belong to this tiny minority who, by prevarication, by skill or luck, have never touched"

   "Nous qui avons survécu", a déclaré Primo Levi, "ne sommes pas de bons témoins, car nous appartenons à cette minuscule minorité qui, par tergiversation, par habileté ou par chance, n'avons jamais touché"

5. The questions are targeted and you must have the financial means to follow the plan. I would not recommend this document unlike the other book; I do not know how to lose weight, which is useful

   Les questions sont ciblées et vous devez avoir les moyens financiers de suivre le plan. Je ne recommanderais pas ce document contrairement à l'autre livre; Je ne sais pas comment perdre du poids, ce qui est utile

6. I read this book in Spanish, in the native language of the writer. I find the book excellent. Not only because of her passionate story but for her love of books and literature

   J'ai lu ce livre en espagnol, dans la langue maternelle de l'auteur. Je trouve le livre excellent. Pas seulement à cause de son histoire passionnée, mais aussi pour son amour des livres et de la littérature

7. I have been reading Chattam for many years, and this is the first time I have to struggle to finish one of these novels. The bottom of the story is not bad, but the finished product was almost undrinkable.

   Je lis Chattam depuis de nombreuses années et c'est la première fois que je dois lutter pour terminer l'un de ces romans. Le fond de l'histoire n'est pas mauvais, mais le produit fini était presque imbuvable.

8. THIS BOOK IS GREAT! I had seen the movie before I read the book and I was not disappointed!

   CE LIVRE EST SUPER! J'avais vu le film avant de lire le livre et je n'ai pas été déçu!

9. I still love it so much. But I wonder if we will not go around in circles ... We change the scenery, we add endearing characters, but there is already the originality of the original creators.

   Je l'aime toujours tellement. Mais je me demande si nous ne tournerons pas en rond ... On change de décor, on ajoute des personnages attachants, mais il y a déjà l'originalité de la création originale.

10. There are many mysteries in life, including Grangé's! I really do not understand the extraordinary opinions about this author: it's wrong! And I hope it is not broadcast too much abroad.

    Il y a beaucoup de mystères dans la vie, y compris ceux de Grangé! Je ne comprends vraiment pas les opinions extraordinaires sur cet auteur: c'est faux! Et j'espère que ça ne sera pas trop diffusé à l'étranger