Human Value Detection from Bilingual Sensory Product Reviews

Boyu Niu[1,2], Céline Manetta[2], Frédérique Segond[1,3]

[1] Inalco, Paris, France [2] IFF, Neuilly-sur-Seine, France [3] Inria, Montpellier, France
boyu.niu@{inalco.fr, iff.com},
celine.manetta@iff.com,
frederique.segond@{inalco.fr, inria.fr}

Abstract

We applied text classification methods on a corpus of product reviews we created with the help of a questionnaire. We found that for certain values, "traditional" deep neural networks like CNN can give promising results compared to the baseline. We propose some ideas to improve the results in the future. The bilingual corpus we created which contains more than 16 000 consumer reviews associated to the human value profile of the authors can be used for different marketing purposes.

Introduction

In this paper, we investigate the possibility of detecting human values from consumer reviews about sensory products (perfume and other scented products such as shampoo and detergent). We carried out a series of experiments to detect human values as defined in the Schwartz's theory (1992, 1996, 2003, 2006) in a corpus of consumer reviews about scented products that we created.

These experiments are part of a research project on consumer segmentation based on psychological traits. This is a method widely used in marketing research that allows manufacturers to create products which better meet the expectations of their end users. This is particularly interesting for the fragrance industry, as smells have special links to emotions (Warrenburg 2002) and psychological states and profiles.

There are previous works about the detection of personality traits from texts (Pennebaker et al., 2001; Mairesse et al., 2007; Majumder et al., 2017; Kazameini et al., 2020; Leonardi et al., 2020; Vásquez and Ochoa-Luna, 2021). In these works, a corpus containing texts and the personality traits auto-evaluated by the author of the texts is used – the authors were asked to answer a personality questionnaire, and the result of this questionnaire is considered as the ground truth in this task. The researchers of the previous works applied different methods to this corpus and observed the performance. As there are few existing works on the detection of human values from texts, and personality traits and human values are both psychological traits that describe the psychological profile of an individual[1], we place our work in the field of psycholinguistics and the related works from which we got inspirations are about personality detection from texts. To the best of our knowledge, this is the first work of applying NLP methods to the detection of human values in the fragrance industry.

In this article, we first present the linguistic resources we used and the formalization of the task. We then describe the methods used in the experiments. After that, we present the experiments and the results we obtained. In the

[1] The difference between the two is that personality traits describe an individual, while human values describe what is important for an individual.

Proceedings of Recent Advances in Natural Language Processing, pages 792–802
Varna, Sep 4–6, 2023
https://doi.org/10.26615/978-954-452-092-2_086

end, we propose some ideas that may improve the results in the future.

Linguistic Resources

Corpus

1.1.1 Corpus Collection

We conducted a survey for perfume and other scented product consumers in the United States and in France. The respondents were invited to answer an online questionnaire composed of three parts: a series of questions on human values (PVQ-21, about which we will give more details in the next part), some demographic questions (age group, gender, having children at home or not), and finally some text boxes where the respondents can indicate the name of the products they had recently used (at least two) and write their review as if they were on the Internet. This allows us to have a corpus annotated with the authors' self-evaluated human value profiles, with some meta data such as their age group. Previous studies (Pennebaker et al., 2001; Mairesse et al., 2007; Majumder et al., 2017) have adopted a corpus obtained via the same self-evaluation approach.

The US corpus contains 8502 reviews written in English by 1932 respondents. A review contains 44.63 words (236.48 characters) in average.

The French corpus contains 7895 reviews written in French by 1915 respondents. A review contains 38.82 words (227.09 characters) in average.

To the best of our knowledge, this is the first bilingual corpus about fragrance products aligned with its authors' answers to a human value questionnaire.

1.1.2 Human Values and the Attribution

Human values describe what is important for an individual in his or her life. The values of the Schwartz's model and their abbreviation used in this article are as follows:

- Power: is this someone who likes to have the control over other people and the resources? (POW)

- Achievement: is this someone who likes to demonstrate his or her skills? (ACH)

- Stimulation: is this someone who is looking for novelty and challenges in life? (STI)

- Hedonism: is this someone who is motivated by personal and sensual pleasure? (HED)

- Self-direction: is this someone who likes to think and act in an original way? (SEL)

- Universalism: is the protection of the well-being of all human beings and nature important for this individual? (UNI)

- Benevolence: is the well-being of close others (such as family members and friends) important? (BEN)

- Tradition: is this someone characterized by the respect for tradition? (TRA)

- Conformity: is this someone who considers self-restraint in everyday life to be important? (CON)

- Security: is the safety, harmony, and stability of society, of relationships and of herself or himself important to this individual? (SEC)

We attribute the human values to the respondents of the survey with the help of a questionnaire based on PVQ-21 (Portrait Value Questionnaire, that contains 21 questions) published by Schwartz (2003).

We transformed the answers to PVQ-21 into a binary classification for each of the values as what was done in the previous works about personality trait detection (Pennebaker et al., 2001; Mairesse et al., 2007; Majumder et al., 2017).

Let $\overline{X_{all}}$ represent the average of the answers to all the 21 questions, and let $\overline{X_v}$ represent the average of the answers to the questions related to the value v.[2] If $\overline{X_v} - \overline{X_{all}} > 0$, then class 1 is assigned to the value v; otherwise, class 0 is assigned to this value. Class 1 means the value is important to this respondent, class 0 means the opposite.

An extraction of the corpus can be found in the appendix.

[2] In this questionnaire, each value has 2 or 3 corresponding questions. If an individual answers a question with a higher score, then there is a greater probability that this individual considers this value as important in his or her life.

Below is the distribution of classes in the corpus collected in the two countries:

	US Corpus		French Corpus	
Value	Class 1	Class 0	Class	Class
Pow	0.133	0.867	0.111	0.889
Ach	0.322	0.678	0.242	0.758
Sti	0.526	0.474	0.446	0.554
Hed	0.707	0.293	0.834	0.166
Sel	0.87	0.13	0.834	0.166
Uni	0.769	0.231	0.792	0.208
Ben	0.793	0.207	0.815	0.185
Tra	0.236	0.764	0.229	0.771
Con	0.52	0.48	0.624	0.376
Sec	0.656	0.344	0.631	0.369

Table 1: Distribution of classes in the two countries in our corpus

We observe that for most values, the class distribution is unbalanced. This has an impact on the strategy we used to calculate the baseline, which will be discussed in 5.1.1.

LIWC Psycholinguistic Lexicon

LIWC (Linguistics Inquiry and Word Count) (Pennebaker et al., 2001) is a multilingual lexicon that organize words in different categories according to their psychological characteristics, such as positive and negative emotions, family, social relations, curiosity, well-being, and different pronouns. For example, the French word "parfum" (perfume) can be found in the following categories: "affect", "emopos" (positive emotion) and "perception", and the word "sucré" (sweet or sweetened) can be found in these categories: "verb", "verb past" (past tense verb), “perception”, “biological”, and “food”.

LIWC has been used in several studies on detecting personality traits from text (Pennebaker et al., 2001; Mairesse et al., 2007; Majumder et al., 2017) [3]. It transforms a document (in our case, a review) into a vector, the dimensions of the vector correspond to the different linguistic categories in LIWC.

Language Algorithms and Models

3.1 Classification Algorithms

The different classification algorithms used in the experiments are: decision tree, SVM and deep neural networks. This allows us to observe how the classical algorithms, from simple to more sophisticated ones, perform on human value detection.

Decision tree is a tree structure where each branch represents a possible decision, and the leaf (or node) following that branch represents the outcome of that decision. SVM consists of creating a hyperplane which optimally separates the objects (in our case the reviews transformed into vectors) projected to a high-dimensional space. The deep neural networks used in our experiments are convolutional networks, bi-directional LSTMs, and pre-trained bidirectional text representation models, followed by fully connected layers. The architecture of convolutional networks is supposed to be able to capture short-distance linguistic features, while LSTM is supposed to be able to manage the memory of information that goes across a longer distance.

3.2 Language Models

We used vector representation at the word level and the document level respectively. The word embedding models used are Word2Vec (Mikolov et al., 2013) and fastText (Bojanoski et al., 2017). Both of these models provide a vector representation of a word. This representation is calculated according to the context in which the word is found in the training corpus. While Word2Vec has a fixed vocabulary, fastText can handle out-of-vocabulary tokens because it takes character-level information into account. Besides, unlike Word2Vec which only supports English, fastText is available in many languages.

The document embedding models applied to the US corpus are BERT (Devlin et al., 2018) and RoBERTa (Liu et al., 2019). As for the French corpus, CamemBERT (Martin et al., 2019) and FlauBERT (Le et al., 2019) were used. All of these models are based on Transformer architecture which is based on the self-attention mechanism (Vaswani et al., 2017), and are designed to pre-train bidirectional representations of texts. RoBERTa is a "replication" of the original BERT with some modifications in the training configurations. FlauBERT and CamemBERT are the French versions of BERT and RoBERTa.

[3] It has many versions as well. In our experiments, the 2007 version (Pennebaker et al., 2007) is applied to the French corpus, and the 2022 version (Boyd et al., 2022) is applied to the American corpus.

Formalization of the Task

Given a respondent of the survey i, for each of this respondent's review $R_{i,j}$, for each one of the 10 values v, we apply model M_v to this review and get its output $O_{i,j,v} = M_v(R_{i,j})$. We then examine if this output is the same as this respondent's auto-evaluation for this particular value E_v.

Experiments

This part presents the experiments and their results. The hyperparameters of all the algorithms we used have been tuned, and the results obtained with the optimal hyperparameters are shown below. For BERT models, all the layers have been tuned. The best F1 scores are shown in bold characters. The optimal hyperparameters can be found in the appendix. For each of the algorithms and methods used, we present the results obtained with the US corpus first, followed by results obtained with the French corpus.

Experiments and the Results

1.1.3 Baseline

As seen before (Table 1), the class distribution is unbalanced for most of the values in our corpus. For that reason, we use a simple dummy classifier with the stratified strategy[4] as baseline. This baseline method generates random predictions with respect to the class distribution of the training corpus (it favors the majority class of the training corpus, but not systematically). This makes the baseline more difficult compared to the uniform strategy.

	Baseline (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.758	0.131	0.138	0.134
Ach	0.582	0.337	0.34	0.338
Sti	0.492	0.532	0.516	0.524
Hed	0.591	0.73	0.704	0.717
Sel	0.765	0.863	0.863	0.863
Uni	0.67	0.797	0.776	0.786
Ben	0.649	0.774	0.777	0.775
Tra	0.642	0.202	0.223	0.212
Con	0.469	0.468	0.491	0.479
Sec	0.564	0.653	0.681	0.667

Table 2: US corpus baseline

	Baseline (France)			
Value	Accuracy	Precision	Recall	F1
Pow	0.79	0.03	0.041	0.035
Ach	0.642	0.237	0.254	0.245
Sti	0.485	0.416	0.42	0.418
Hed	0.714	0.834	0.824	0.829
Sel	0.706	0.815	0.83	0.822
Uni	0.654	0.798	0.769	0.783
Ben	0.708	0.827	0.816	0.821
Tra	0.62	0.202	0.194	0.198
Con	0.551	0.661	0.634	0.647
Sec	0.539	0.628	0.641	0.635

Table 3: French corpus baseline

1.1.4 LIWC Features + Decision Tree / SVM

We applied the decision tree and SVM to LIWC vectors for the classification task.

Experiment 1:

	LIWC + Decision Tree (US)			
Value	Accurac	Precision	Recal	F1
Pow	0.745	0.102	0.112	0.10
Ach	0.582	0.342	0.354	0.34
Sti	0.552	0.552	0.928	0.69
Hed	0.734	0.734	1	0.84
Sel	0.859	0.859	1	**0.92**
Uni	0.784	0.784	1	0.87
Ben	0.78	0.78	1	0.87
Tra	0.64	0.212	0.245	0.22
Con	0.522	0.52	0.526	0.52
Sec	0.64	0.64	1	0.78

Table 4: LIWC features + decision tree, US corpus

Experiment 2:

	LIWC + SVM (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.832	0.171	0.06	0.089
Ach	0.608	0.354	0.299	0.324
Sti	0.543	0.542	1	**0.703**
Hed	0.737	0.736	1	**0.848**
Sel	0.859	0.859	1	**0.924**
Uni	0.784	0.784	1	0.879
Ben	0.783	0.782	1	**0.878**
Tra	0.723	0.28	0.179	0.219
Con	0.499	0.499	1	0.666
Sec	0.642	0.641	1	0.781

Table 5: LIWC features + SVM, US corpus

Experiment 3:

	LIWC + Decision Tree (France)			
Value	Accurac	Precision	Recall	F1
Pow	0.806	0.077	0.092	0.08
Ach	0.644	0.224	0.246	0.23
Sti	0.511	0.455	0.445	0.45
Hed	0.856	0.856	1	**0.92**
Sel	0.823	0.823	1	**0.90**
Uni	0.794	0.794	1	0.88
Ben	0.832	0.832	1	**0.90**
Tra	0.629	0.253	0.256	0.25
Con	0.654	0.654	1	**0.79**
Sec	0.625	0.625	1	**0.76**

Table 6: LIWC features + decision tree, French corpus

Experiment 4:

	LIWC + SVM (France)			
Value	Accurac	Precisio	Recal	F1

[4] https://github.com/scikit-learn/scikit-learn/blob/7f9bad99d/sklearn/dummy.py#L33

Pow	0.865	0.17	0.105	**0.13**
Ach	0.68	0.24	0.206	0.22
Sti	0.549	0.499	0.493	0.49
Hed	0.856	0.856	1	**0.92**
Sel	0.823	0.823	1	**0.90**
Uni	0.794	0.794	1	0.88
Ben	0.832	0.832	1	**0.90**
Tra	0.691	0.331	0.246	**0.28**
Con	0.653	0.654	0.996	0.79
Sec	0.625	0.625	1	**0.76**

Table 7: LIWC features + SVM, French corpus

1.1.5 Word Embedding + Deep Neural Networks (CNN / Bi-LSTM) + Fully Connected Layer

We applied pre-trained word embeddings followed by a deep neural network (CNN or Bi-LSTM). When Word2Vec is applied to the US corpus, the out-of-vocabulary words are randomly[5] vectorized. The fixed document length is of 56 words for the US reviews, and of 52 words for the French reviews. Longer reviews are trimmed, while shorter reviews are padded with special padding tokens.

In the CNN experiments, we used three different kernel sizes (1, 2, 3 or 2, 3, 4). The number of kernels varies between 50 and 100.

Experiment 5:

	Word2Vec + CNN (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.842	0.302	0.123	**0.153**
Ach	0.609	0.39	0.45	0.411
Sti	0.543	0.543	0.987	0.698
Hed	0.736	0.735	1.0	0.844
Sel	0.861	0.862	0.997	0.923
Uni	0.796	0.806	0.976	0.881
Ben	0.784	0.786	0.992	0.875
Tra	0.743	0.403	0.226	0.265
Con	0.501	0.499	0.972	0.655
Sec	0.648	0.649	0.982	0.779

Table 8: Word2Vec + CNN, US corpus

Experiment 6:

	fastText + CNN (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.801	0.19	0.115	0.13
Ach	0.608	0.397	0.495	**0.432**
Sti	0.545	0.545	0.972	0.695
Hed	0.736	0.736	0.999	0.844
Sel	0.861	0.862	0.997	0.923
Uni	0.8	0.807	0.982	**0.884**
Ben	0.786	0.786	0.997	0.876
Tra	0.722	0.331	0.303	**0.291**
Con	0.515	0.506	0.959	0.659
Sec	0.645	0.644	0.998	0.781

Table 9: fastText + CNN, US corpus

Experiment 7:

	Word2Vec + Bi-LSTM (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.863	0.037	0.012	0.019
Ach	0.677	0.204	0.032	0.055
Sti	0.538	0.54	0.987	0.695
Hed	0.736	0.735	1.0	0.844
Sel	0.858	0.857	1.0	0.921
Uni	0.786	0.786	1.0	0.878
Ben	0.778	0.778	1.0	0.873
Tra	0.782	0.012	0.005	0.007
Con	0.525	0.514	0.972	**0.667**
Sec	0.641	0.641	1.0	0.779

Table 10: Word2Vec + Bi-LSTM, US corpus

Experiment 8:

	fastText + Bi-LSTM (US)			
Value	Accuracy	Precision	Recall	F1
Pow	0.855	0.019	0.012	0.015
Ach	0.679	0.262	0.058	0.093
Sti	0.535	0.539	0.982	0.693
Hed	0.734	0.734	1.0	0.844
Sel	0.859	0.858	1.0	0.922
Uni	0.786	0.786	1.0	0.878
Ben	0.779	0.779	1.0	0.873
Tra	0.768	0.074	0.007	0.014
Con	0.516	0.51	0.978	0.664
Sec	0.641	0.641	1.0	0.779

Table 11: fastText + Bi-LSTM, US corpus

Experiment 9:

	fastText + CNN (France)			
Value	Accurac	Precisio	Recal	F1
Pow	0.893	0.22	0.095	0.12
Ach	0.714	0.367	0.3	0.32
Sti	0.549	0.488	0.639	**0.54**
Hed	0.847	0.846	1.0	0.91
Sel	0.818	0.818	1.0	0.89
Uni	0.819	0.818	1.0	**0.89**
Ben	0.828	0.829	0.997	0.90
Tra	0.703	0.309	0.184	0.21
Con	0.655	0.654	0.992	0.78
Sec	0.632	0.63	0.996	**0.76**

Table 12: fastText + CNN, French corpus

Experiment 10:

	fastText + Bi-LSTM (France)			
Value	Accurac	Precisio	Recal	F1
Pow	0.901	0.08	0.021	0.03
Ach	0.764	0.187	0.038	0.06
Sti	0.544	0.359	0.079	0.12
Hed	0.844	0.844	1.0	0.91
Sel	0.818	0.819	0.997	0.89
Uni	0.814	0.814	1.0	0.89
Ben	0.825	0.825	1.0	0.90
Tra	0.747	0.207	0.046	0.07
Con	0.656	0.654	0.998	0.78
Sec	0.625	0.624	1.0	0.76

Table 13: fastText + Bi-LSTM, French corpus

1.1.6 BERT Family

The tables below are the results obtained with BERT and RoBERTa applied to the US corpus,

[5] The components are random values between -0.25 and 0.25 and follow the normal distribution.

and FlauBERT and CamemBERT applied to the French corpus.

Experiment 11:

	BERT (US)			
Value	Accurac	Precisio	Recall	F1
Pow	0.89	0.111	0.032	0.04
Ach	0.68	0.493	0.275	0.34
Sti	0.508	0.507	1.0	0.66
Hed	0.717	0.717	1.0	0.83
Sel	0.861	0.861	1.0	0.92
Uni	0.753	0.753	1.0	0.85
Ben	0.81	0.81	1.0	0.89
Tra	0.756	0.533	0.137	0.20
Con	0.56	0.56	1.0	0.71
Sec	0.649	0.649	1.0	**0.78**

Table 14: Experiment with BERT, US corpus

Experiment 12:

	RoBERTa (US)			
Value	Accurac	Precisio	Recall	F1
Pow	0.889	0	0	0
Ach	0.676	0.278	0.038	0.06
Sti	0.558	0.536	0.922	0.67
Hed	0.717	0.717	1.0	0.83
Sel	0.861	0.861	1.0	0.92
Uni	0.751	0.751	1.0	0.85
Ben	0.81	0.81	1.0	0.89
Tra	0.742	0	0	0
Con	0.56	0.56	1.0	0.71
Sec	0.649	0.649	1.0	**0.78**

Table 15: Experiment with RoBERTa, US corpus

Experiment 13:

	FlauBERT (France)			
Valu	Accurac	Precisio	Recall	F1
Pow	0.898	0	0	0
Ach	0.759	0	0	0
Sti	0.537	0	0	0
Hed	0.832	0.832	1.0	0.90
Sel	0.84	0.84	1.0	0.91
Uni	0.797	0.797	1.0	0.88
Ben	0.828	0.828	1.0	0.90
Tra	0.777	0.08	0.01	0.01
Con	0.62	0.62	1.0	0.76
Sec	0.601	0.601	1.0	0.74

Table 16: Experiment with FlauBERT, French corpus

Experiment 14:

	CamemBERT (France)			
Valu	Accurac	Precision	Recall	F1
Pow	0.87	0.193	0.089	0.11
Ach	0.707	0.394	0.349	**0.35**
Sti	0.559	0.517	0.434	0.46
Hed	0.832	0.832	1.0	0.90
Sel	0.84	0.84	1.0	0.91
Uni	0.797	0.797	1.0	0.88
Ben	0.828	0.828	1.0	0.90
Tra	0.757	0.299	0.105	0.14
Con	0.628	0.625	1.0	0.76
Sec	0.601	0.601	1.0	0.74

Table 17: Experiment with CamemBERT, French corpus

Discussion

We can observe that decision trees and SVM give good F1 scores for the human values with an unbalanced distribution in our corpus (hedonism, autonomy, universalism, benevolence, and security). As the positive class (class 1) is the majority class for these values, the accuracy and the precision scores are the same, the recall is 1, we can infer that the classifier just votes systematically for the majority class when being applied to the test corpus. This observation may be explained by the fact that LIWC is not suitable for our specific domain. For example, the validity of the word "parfum" (perfume) being categorized under "positive emotion" can be questionable, as it is highly likely that a disliked scent will elicit a negative emotion. Another example: while the word "shampoing" (shampoo) has a high frequency in our French corpus, it is not in this dictionary. [6] As a consequence, this piece of information is completely lost, while it can be useful for the model to do prediction.

With CNN model, we obtained better results compared to our baseline in terms of F1 score when it comes to the values of achievement, stimulation, tradition, and conformity of the US corpus, without having the classifier systematically predicting the majority class. As for the French corpus, we observe that the CNN gives better results compared to our baseline when it comes to achievement and stimulation. In our experiments with CNN, we used kernel sizes 1, 2, 3, and 4. This could suggest that certain linguistic features, such as bigrams and trigrams, may be useful indications for human value detection from text.

The sequential model (Bi-LSTM) that we tested favors the majority class too, especially when it comes to values that have an unbalanced class distribution (power, hedonism, self-direction, universalism, benevolence). If we make a comparison with the results obtained with CNN, does this mean that a longer memory

[6] In this article, we can only show examples of the French version of LIWC, because the content of the English version is not accessible for us.

does not do any help to human value detection from texts?

We can also observe that the models of the BERT family systematically favor the majority class for most values, this is the case for both US and French corpora. It would be interesting to do further studies on the effectiveness of complex language models like BERT in psycholinguistic topics, especially when we have a training corpus where the classes have an unbalanced distribution.

Conclusion

We tested decision tree, SVM, convolutional neural network (CNN), sequential neural network (Bi-LSTM), as well as BERT models to detect human values in the corpus we created. We observed that the decision tree, SVM and BERT models tend to always predict the majority class in our task. The CNN model has a performance that clearly exceeds our baseline when it comes to certain values.

To improve the performance of this task, we have a few ideas for future work:

- It would be interesting to study the relevance of using data augmentation methods in our task. It would also be interesting to adopt a cost - sensitive learning strategy during the training stage.
- We can create a psycholinguistic dictionary dedicated to field of sensory studies or adapt LIWC to this field.
- Instead of a fully connected layer at the end of a CNN, we can test other classifiers . We can also test the parallel CNN model. Israeli et al. (2022) reported good performance of this model.
- In the experiments presented in this article, we trained a model for each of the values independently. We can think of training a model for all the values at the same time, and then investigate if such a model takes into account the correlation that may exist between the different values.

As this is the first project about human value detection from consumer reviews about sensory products to our knowledge, we mainly applied and presented the results of the classical methods. We will apply more recent models and add domain specific knowledge as a next step.

Besides the experiments we have done, the bilingual corpus we created which contains more than 16 000 consumer reviews associated to the human value profile and the demographic information of the authors that can be used for different marketing purposes is also a first contribution of this kind. Taking into consideration the demographic information is also planned for next step.

Acknowledgement

This work is carried out within the framework of a CIFRE agreement, managed by the French National Association for Technology Research (ANRT), and established between the ERTIM laboratory of Inalco and the HCI department of IFF.

References

Piotr Bojanowski, Grave Edouard, Joulin Armand, and Mikolov Tomas. 2017. Enriching word vectors with subword information. *Transactions of the association for computational linguistics* 5: 135-146.

Ryan L. Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker. 2022. The development and psychometric properties of LIWC-22. The University of Texas at Austin.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Abraham Israeli, Alexander Kremiansky, and Oren Tsur. 2022. This Must Be the Place: Predicting Engagement of Online Communities in a Large-scale Distributed Campaign. In *Proceedings of the ACM Web Conference 2022* (pp. 1673-1684).

Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Eric Cambria. 2020. Personality trait detection using bagged svm over bert word embedding sets. arXiv preprint arXiv:2010.01309.

Huang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoit Crabbé, Laurent Besacier, and Didier Schwab. 2019. Flaubert: Unsupervised language model pre-training for french . arXiv preprint arXiv:1912.05372.

Simone Leonardi, Diego Monti, Giuseppe Rizzo, and Maurizio Morisio. 2020. Multilingual transformer-based personality traits estimation. Disclosure, 11(4), 179.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. .2019. Roberta: A robustly optimized bert pre-training approach. arXiv preprint arXiv:1907.11692.

François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, 30, 457-500.

Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. 2017. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2), 74-79.

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de La Clergerie, Djamé Seddah, and Benoît Sagot. 2019. CameBERT : a tasty French language model. arXiv preprint arXiv:1911.03894.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

James W. Pennebaker, Martha E. Francis, and Roger L. Booth. 2001. Linguistic inquiry and word count: LIWC 2001. Mahway : Lawrence Erlbaum Associates, 71(2001), 2001.

James W. Pennebaker, Roger L. Booth, and Martha E. Francis. 2007. LIWC2007: Linguistic inquiry and word count. Austin, TX: liwc . clean .

Soujanya Poria, Alexandar Gelbukh, Basant Agarwal, Erik Cambria, and Newton Howard. 2013. Common sense knowledge based personality recognition from text. In *Mexican International Conference on Artificial Intelligence* (pp. 484-496). Springer, Berlin, Heidelberg.

Shalom H. Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in experimental social psychology*, 25(1), 1-65.

Shalom H. Schwartz. 1996. Value Priorities and Behavior: Applying a theory of integrated value systems. In C. Seligman, JM olson & MP Zanna (Eds.), *The psychology of values: The Ontario symposium*, volume 8(pp. 1-24). Manwah , NJ: Erlbaum.

Shalom H. Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, 32(5), 519-542.

Shalom H. Schwartz. 2003. A proposal for measuring value orientations across nations. *Questionnaire Package of the European Social Survey*. 259-290.

Shalom H. Schwartz. 2006. The basic values of the person: theory, measurements and applications. *French journal of sociology*. 47 (4). 929-968.

Ricardo Lazo Vásquez, and José Ochoa-Luna. (2021, October). Transformer-based Approaches for Personality Detection using the MBTI Model. In *2021 XLVII Latin American Computing Conference (CLEI)* (pp. 1-7). IEEE.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin.. Attention is all you need. *Advances in neural information processing systems*, 30.

Stephane Warrenburg. 2002. Measurement of emotion in olfactory research.

Appendix A – An Extraction of the Corpus

Respondent ID	1	1	2	2
Review	I tried this product for the first time 4 months ago and I was so impressed with how it felt on my skin […]	The scented oil refills for electrical plug diffusers not only keep your space smelling nice for 90 days, […]	It is a classic masculine smell nothing fancy never changed not too overpowering […]	Basic deodorant complements the cologne perfectly Priced competitively and can be obtained at your local drugstore […]
Pow	0	0	0	0
Ach	0	0	0	0
Sti	1	1	0	0
Hed	1	1	0	0
Sel	1	1	1	1
Uni	1	1	1	1
Ben	1	1	0	0
Tra	0	0	0	0
Con	1	1	1	1
Sec	1	1	1	1

Table 1: Examples

We can observe that the reviews written by the same respondent (indicated by the ID) always have the same labels for each of the values, because these labels are calculated based on the same author's answers to the questionnaire.

Appendix B – The Hyperparameters Used in the Experiments

Experiment 1: LIWC features + decision tree, US corpus

Value	Parameters
Pow	'max_depth' : 40
Ach	'max_depth' : 57
Sti	'max_depth' : 2
Hed	'max_depth' : 2
Sel	'max_depth' : 2
Uni	'max_depth' : 2
Ben	'max_depth' : 2
Tra	'max_depth' : 35
Con	'max_depth' : 2
Sec	'max_depth' : 2

Table 2

Experiment 2: LIWC features + SVM, US corpus

Value	Parameters
Pow	'C': 100, 'gamma': scale, 'kernel': 'rbf'
Ach	'C': 100, 'gamma': scale, 'kernel': 'rbf'
Sti	'C': 0.5, 'gamma': 0.1, 'kernel': 'rbf'
Hed	'C': 5, 'gamma': 0.1, 'kernel': 'rbf'
Sel	'C': 1, 'gamma': scale, 'kernel': 'rbf'
Uni	'C': 1, 'gamma': 0.01, 'kernel': 'rbf'
Ben	'C': 5, 'gamma': 0.1, 'kernel': 'rbf'
Tra	'C': 100, 'gamma': scale, 'kernel': 'rbf'
Con	'C': 1, 'gamma': 1, 'kernel': 'rbf'
Sec	'C': 5, 'gamma': 0.1, 'kernel': 'rbf'

Table 3

Experiment 3: LIWC features + decision tree, French corpus

Value	Parameters
Pow	'max_depth': 53
Ach	'max_depth': 35
Sti	'max_depth': 40
Hed	'max_depth': 2
Sel	'max_depth': 2
Uni	'max_depth': 2
Ben	'max_depth': 2
Tra	'max_depth': 35
Con	'max_depth': 2
Sec	'max_depth': 2

Table 4

Experiment 4: LIWC features + SVM, French corpus

Value	Parameters
Pow	'max_depth': 53
Ach	'max_depth': 35
Sti	'max_depth': 40
Hed	'max_depth': 2
Sel	'max_depth': 2
Uni	'max_depth': 2
Ben	'max_depth': 2
Tra	'max_depth': 35
Con	'max_depth': 2
Sec	'max_depth': 2

Table 5

Experiment 5: Word2Vec + CNN, US corpus

The activation function after the convolution is ReLU. The optimizer used is AdamW. A dropout of 0.5 is used when training.

Value	Parameters
Pow	100 kernels of sizes 1, 2, 3 ; lr=0.01
Ach	100 kernels of sizes 2, 3, 4 ; lr=0.001
Sti	100 kernels of sizes 2, 3, 4 ; lr=0.001
Hed	100 kernels of sizes 1, 2, 3 ; lr=0.01
Sel	50 kernels of sizes 2, 3, 4 ; lr=0.001
Uni	50 kernels of sizes 2, 3, 4 ; lr=0.001
Ben	100 kernels of sizes 2, 3, 4 ; lr=0.001

Tra	50 kernels of sizes 2, 3, 4 ; lr=0.001
Con	100 kernels of sizes 2, 3, 4 ; lr=0.001
Sec	100 kernels of sizes 1, 2, 3 ; lr=0.01

Table 6

Experiment 6: fastText + CNN, US corpus

The activation function after the convolution is ReLU. The optimizer used is AdamW. A dropout of 0.5 is used when training.

Value	Parameters
Pow	100 kernels of sizes 2, 3, 4 ; lr=0.001
Ach	50 kernels of sizes 2, 3, 4 ; lr=0.001
Sti	100 kernels of sizes 1, 2, 3 ; lr=0.01
Hed	100 kernels of sizes 2, 3, 4 ; lr=0.01
Sel	50 kernels of sizes 2, 3, 4 ; lr=0.001
Uni	50 kernels of sizes 2, 3, 4 ; lr=0.001
Ben	100 kernels of sizes 2, 3, 4 ; lr=0.001
Tra	50 kernels of sizes 2, 3, 4 ; lr=0.001
Con	100 kernels of sizes 2, 3, 4 ; lr=0.001
Sec	50 kernels of sizes 2, 3, 4 ; lr=0.001

Table 7

Experiment 7: Word2Vec + Bi-LSTM, US corpus

The optimizer used is AdamW. A dropout of 0.5 is used when training.

Value	Parameters[7]
Pow	64 : 2 : 0.01
Ach	64 : 2 : 0.01
Sti	64 : 2 : 0.01
Hed	128 : 1 : 0.001
Sel	128 : 1 : 0.01
Uni	128 : 1 : 0.01
Ben	128 : 1 : 0.01
Tra	128 : 1 : 0.01
Con	64 : 2 : 0.01
Sec	128 : 1 : 0.01

Table 8

Experiment 8: fastText + Bi-LSTM, US corpus

The optimizer used is AdamW. A dropout of 0.5 is used when training.

Value	Parameters
Pow	64 : 2 : 0.01
Ach	64 : 2 : 0.01
Sti	64 : 2 : 0.01
Hed	64 : 2 : 0.01
Sel	64 : 2 : 0.01
Uni	64 : 2 : 0.01
Ben	64 : 2 : 0.01
Tra	64 : 2 : 0.01
Con	64 : 2 : 0.01
Sec	64 : 2 : 0.01

Table 9

Experiment 9: fastText + CNN, French corpus

The activation function after the convolution is ReLU. The optimizer used is AdamW. A dropout of 0.5 is used when training.

Value	Parameters
Pow	100 kernels of sizes 1, 2, 3 ; lr=0.01
Ach	100 kernels of sizes 1, 2, 3 ; lr=0.01
Sti	50 kernels of sizes 2, 3, 4 ; lr=0.001
Hed	100 kernels of sizes 1, 2, 3 ; lr=0.01
Sel	100 kernels of sizes 2, 3, 4 ; lr=0.01
Uni	100 kernels of sizes 1, 2, 3 ; lr=0.01
Ben	100 kernels of sizes 2, 3, 4 ; lr=0.01
Tra	100 kernels of sizes 2, 3, 4 ; lr=0.001
Con	100 kernels of sizes 2, 3, 4 ; lr=0.01
Sec	100 kernels of sizes 1, 2, 3 ; lr=0.01

Table 10

Experiment 10: fastText + Bi-LSTM, French corpus

Value	Parameters
Pow	64 : 2 : 0.01
Ach	64 : 2 : 0.01
Sti	64 : 2 : 0.01
Hed	128 : 1 : 0.001
Sel	128 : 1 : 0.01
Uni	128 : 1 : 0.01
Ben	128 : 1 : 0.01
Tra	128 : 1 : 0.01
Con	64 : 2 : 0.01
Sec	128 : 1 : 0.01

Table 11

Experiment 11: BERT

Value	Parameters[8]

[7] The hyperparameters separated by semicolons are: the number of hidden units per layer in the LSTM network; the number of layers in the LSTM network; the learning rate. The other LSTM experiments have the same structure

[8] AdamW optimizer is used. The learning phase is limited to 2 epochs for time reasons and to avoid over-learning. We have found that results are generally not improved beyond 2 epochs. The other experiments done with

Pow	lr=1e-04
Ach	lr=5e-05
Sti	lr=5e-05
Hed	lr=5e-05
Sel	lr=5e-05
Uni	lr=5e-05
Ben	lr=1e-04
Tra	lr=5e-05
Con	lr=5e-04
Sec	lr=5e-05

Table 12

Experiment 12: RoBERTa

Value	Parameters
Pow	lr=5e-05
Ach	lr=5e-05
Sti	lr=1e-05
Hed	lr=5e-05
Sel	lr=5e-05
Uni	lr=5e-05
Ben	lr=5e-05
Tra	lr=5e-05
Con	lr=1e-04
Sec	lr=5e-05

Table 13

Experiment 13: FlauBERT

Value	Parameters
Pow	lr=5e-05
Ach	lr=5e-05
Sti	lr=5e-05
Hed	lr=5e-05
Sel	lr=5e-05
Uni	lr=5e-05
Ben	lr=5e-05
Tra	lr=5e-05
Con	lr=5e-05
Sec	lr=5e-05

Table 14

Experiment 14: CamemBERT

Value	Parameters
Pow	lr=5e-05
Ach	lr=5e-05
Sti	lr=5e-05
Hed	lr=5e-05
Sel	lr=5e-05
Uni	lr=5e-05
Ben	lr=5e-05
Tra	lr=5e-05
Con	lr=5e-05
Sec	lr=5e-05

Table 15

models of the BERT family have the same configuration.