XRCE Personal Language Analytics Engine for Multilingual Author Profiling

Notebook for PAN at CLEF 2015

Scott Nowson, Julien Perez, Caroline Brun, Shachar Mirkin, and Claude Roux

Xerox Research Centre Europe
{first name.last name}@xrce.xerox.com

Abstract This technical notebook describes the methodology used – and results achieved – for the PAN 2015 Author Profiling Challenge by the team from Xerox Research Centre Europe (XRCE). This year, personality traits are introduced alongside age and gender in a corpus of tweets in four languages – English, Spanish, Italian and Dutch. We describe a largely language agnostic methodology for classification which uses language specific linguistic processing to generate features. We also report on experiments in which we use machine translation to accommodate for languages in which there is less training data. Native language results are successful, but socio-demographic signals in language seem to be lost under MT conditions.

1 Introduction

Personal Language Analytics is a branch of text mining in which the object of analysis is the author of a document rather than the document itself. Language use in text (or indeed, speech) can reveal a great deal about a person: it can reveal one's gender, age or nationality, among other demographic traits. It can also provide clues as to one of the most important individual differences: personality. For example, when writing personal emails, out-going, social Extraverts are more likely to start by saying 'hi' while Introverts opt for 'hello' [7].

Personality traits (and indeed the other human attributes mentioned) are a valuable source of information for applications such as user modeling or social media engagement. Work in this area, particularly in the computational recognition of personality, is garnering increasing interest with a number of workshops being organized in recent years (e.g. [3], [21]). The addition of personality as a target trait in the PAN Author Profiling challenge in 2015 [18] serves as further evidence.

This paper presents the contribution of Xerox Research Centre Europe to the Author Profiling challenge 2015. We leverage our experience in multi-lingual processing by using language specific tools for each the four languages of the data set (see section 2 for more details). However, our methodology beyond this processing is broadly language agnostic: as much as possible we use a comparable feature set across each language; we also use the same parameters in our experiments.

One notable aspect of the dataset is the varying size of the corpora for the different languages. Therefore, in addition to exploring classification within each language in

isolation, we have also used statistical machine translation in order to generate larger datasets. MT has shown to be of use with NLP tasks such as sentiment analysis [2]; we explore its utility in Author Profiling, where the targets of classification are socio-demographic labels.

2 Data

The data for the Author Profiling task is drawn from Twitter. For each user, the data consists of a number of tweets (the average is approximately 100 per subject) and a series of gold standard labels: gender (Male or Female), age-class (one of 18-24, 25-34, 35-49, 50-xx) and personality. The labels are provided by the author, with scores on five personality traits being calculated via self-assessment responses to the short big5 test (BFI-10, [17]), normalized between -0.5 and +0.5. Table 1 shows the volume of data per language for the training set. As can clearly be seen, the Italian and Dutch data sets are considerably smaller than the Spanish and English.

Language	Authors	Tweets
English (EN)	152	14166
Spanish (ES)	100	9879
Italian (IT)	38	3687
Dutch (NL)	34	3350

Table 1. Data volume – by number of authors and tweets – across the four languages of the training dataset.

2.1 Data preprocessing

There are a number of differences between the data provided for the challenge and data typically collected directly from twitter.

- The data has been anonymised to the extent that all user mentions have been replaced with '`@username`'
- Unicode characters typically representing 'emojis' – a commonly occurring phenomena in Tweets – have not been encoded in the data. Thus, their use have been replaced by unknown character markers, e.g. '`?????`'

Other features of tweets, such as URLs and hastags, remain as per the original data.

2.2 Evaluation

The task of the Author Profiling Challenge is to predict an author's demographics from their tweets. Performance will be evaluated on the prediction of gender and personality traits in the four languages, along with age for the larger corpora, English and Spanish. For the official challenge, age and gender will be ranked by accuracy, the personality traits by Root Mean Squared Error (RMSE).

3 Methodology

In this section we describe the methods we have combined to form the core pipeline of our Personal Language Analytics engine: firstly we report on the linguistic analysis which forms our pre-processing and feature extraction steps; secondly, the techniques of the learning framework are outlined. We also introduce the machine translation (MT) process we employed (as introduced earlier) to explore the effect on classification of using translated data to boost smaller corpora.

3.1 Linguistic Processing

In order to feed the prediction models, we use a robust dependency syntactic parser [1] to extract a wide range of textual features, from standard n-grams to more sophisticated linguistic features.

Processing Steps Processing here includes tokenization, morpho-syntactic analysis, POS tagging – which is performed via a combination of hand-written rules and HMM – Named Entity Detection, chunking and finally, extraction of dependency relations such as subject, object and modifiers between lexical nodes.

This is the stage of processing in which, as mentioned previously, we use-language specific tools. Several grammars have been developed for this parser, among which are the grammars for the PAN languages, i.e. English, Spanish, Italian and Dutch. These grammars are in different stages of development, the English one being more advanced than the others. Consequently, the set of features extracted is different from one language to another (see Table 2).

This parser has also been customized to parse social media data, and detects hashtags, mentions, and (ASCII) emoticons, along with labelling the latter with their polarity. For English, we have integrated a normalization dictionary (from [10]) in the preprocessing steps of the analysis. The English grammar also includes a polarity lexicon and a sentiment analysis layer to detect opinionated relations.

Feature extraction We apply the parser on the different sets of PAN input data, and select a broad set of linguistically interesting features. In order to be closer to our aim of language independence, we do not draw on the deepest level of morpho-syntactic analysis which our toolset provides. For example, Spanish adjectives can have gender inflections, while English adjectives typically do not at the same level. We recognise that in doing this we may not be using the features of a given language as much as possible. However – tool performance aside – this is in-line with our broader aims.

The features extracted are of two types: word-level or class-based features. Word-level features are associating information to the surface and lemma forms of the words directly, while class-based features are more abstract and more generalised: they encode the presence of a given POS, semantic type, hashtags, etc, without tying the feature to the surface form.

- word-level features: unigram, bigram and trigram of surface and lemmatized form of the words; part-of-speech of surface and normalized word; words with negation,

words with at least three repeated letters; bigram of repeated character (cc), trigram of repeated character (ccc), quadrigram of repeated characters (cccc);
- class-based features: named entities (places, persons, organisation, dates and time expressions); unigram, bigram and trigram of POS tags, positive emoticons, negative emoticons, other emoticons; hashtags, mentions and http links; use of feminine or masculine firstnames and pronouns; capitalized words.

Table 2 summarizes the set of features we retained for each language.

Feature type	**en**	**es**	**it**	**nl**	**Examples (for English)**
Surface unigrams, bigrams & trigrams	✓	✓	✓	✓	uni(going), bi(going to), tri(going to talk)
Lemmatized unigrams, bigrams & trigrams	✓	✓	✓	✓	unilem(go), bilem(go to), trilem(go to talk)
word with negation	✓	-	-	-	NEG(nice) in *it is not nice*
Lemmatized word & POS	✓	✓	✓	✓	NOUN(dog), VERB(be)
Bigram, trigram, quadrigrams of repeated characters	✓	✓	✓	✓	bichar(ii), trichar(iii) quadrichar(iiii)
lemmatized unigrams, bigrams & trigrams of POS	✓	✓	✓	✓	unipos(VERB), bipos(VERB PREP), tripos(VERB PREP VERB)
Hashtags	✓	✓	✓	✓	HASHTAG
Http links	✓	✓	✓	✓	HTTPLINK
mentions	✓	✓	✓	✓	MENTION
Fully capitalized words	✓	✓	✓	✓	ALLCAP
Named entities (Pers, Loc, Org, Date, City, Country)	✓	✓	✓	-	Pers, City, Country Org, Date
Time expressions	✓	-	-	-	TIMEX (e.g. “three days after”)
Positive or negative word	✓	-	-	-	POSW (e.g. “awesome”), NEGW (e.g. “disaster”)
Positive or negative emoticon	✓	✓	✓	✓	POSEMOT (e.g. “:>”), NEGEMOT (e.g. “:<”)
Other emoticons	✓	✓	✓	✓	EMOT
Feminine or Masculine firstname	✓	✓	-	-	FN-FEM, e.g. “Mary”, FN-MASC, e.g. “Peter”
Word with at least three repeated letters	✓	✓	✓	✓	REPEATLET, e.g. for *niiiiiiiiice*

Table 2. Features used per language. where en, es, it and nl denote English, Spanish, Italian and Dutch respectively.

3.2 Learning Framework

Our learning framework is composed of 3 elements. First, the exhaustive tag-set produced by the linguistic preprocessing step is pruned using frequency thresholding determined by cross-validation. This reduces the occurrences of heavily under-used features. In the second step, the resulting index of features is compressed using truncated singular value decomposition. Finally, ensemble models are produced for each personality and demographic trait.

Truncated Singular Value Decomposition Singular Value Decomposition (SVD) [8] is a widely used technique for predictive data analysis in sparse dataset situations. It decomposes a given input matrix into a product of three matrices such that $X = USV^T$. Thereby, U and V are unitary matrices which essentially rotate the dataset. S is a diagonal matrix (producing a scaling) with the ordered singular values as entries.

In the truncated version, the purpose of the method is to compute an approximation of X instead of the exact decomposition such as, for instance, in Principal Components Analysis (PCA) [11,9]. Indeed, by producing a low-rank approximation, the method copes with the noise present in the data by extracting the principal dimensions describing the data and projecting the data at the same time. Furthermore, the problem of data sparsity and high-dimensionality in the context of text analysis is addressed because the resulting representation of the points of the compressed dataset are dense and of low-dimension. Truncated SVD technically requires the setting of the smaller valued k diagonal entries in S to 0.

The resulting reconstruction US_kkV^T has a rank k. Neglecting all but the first k components is justified since the data noise perturbs the small eigenvalues, whereas the first k components supposedly capture the underlying structure of the data. Selecting the cutoff value k defines the so-called model-order selection problem of truncated SVD. In our framework, the selection has been determined through cross-validation.

Ensemble decision models Ensemble methods [19,6] are learning algorithms that construct a set of classifiers with new data being classified by an integrating over the resulting set of predictions. The original ensemble method is Bayesian averaging but more recent algorithms include error correcting output coding bagging and boosting. The efficiency of such an approach for non-convex learning model has been often demonstrated by the capability to cope with variance and biases due to the challenging nature of the considered data. For each personality and demographic trait, an ensemble of 10 classifiers is trained and used for inference.

Sub-data classification Our framework enables these ensemble classifiers to operate at different levels for any given data point – in the case of this challenge a data point is considered to be a single author. In the first instance, a 'user-level' decision consists in inferring a given trait from the compressed representation of an aggregatied view of the features of the entire dataset, i.e. the full set of tweets.

A second level – in this setting 'tweet-level' – is to submit each sub-data point (i.e. each tweet) for a decision from the inference model. These sub-decisions are then combined to produce an expected decision at the higher level.

3.3 Machine translation models

We created machine translation models from English to each one of Spanish, Italian and Dutch, in order to increase the size of the training data of these languages. The details of these models are described below.

Parallel corpora We wished to use the same setting for all language-pairs. To that end, we chose parallel corpora that are available for all language pairs, namely: the European Parliament proceedings [13][1] and the TED[2] talks parallel corpus, WIT3 [4].[3] WIT3, consisting of spoken-language transcripts, represents a corpus which is closer in nature to the tweet data used in the challenge. Europarl was chosen mostly for its size. Our combined training data consists of approximately 2 million bi-sentences for each languages-pair, with 50 million tokens for each language. The Europarl corpus accounts for more than 90% of this data. The two corpora were concatenated to create the training data for the MT models.

Translation System Moses [14], a popular, open-source phrase-based MT system[4], was used to train translation models and translate the tweets data.

Preprocessing We used the standard Moses tools to preprocess the data, including tokenization, lowercasing and removal of bi-sentences where at least one of the sentences is empty or longer than 80 tokens.

Recasing and Language models We used SRILM [20] version 1.7.1 to train 5-gram language models on the target side of the parallel corpus, with modified Kneser-Ney discounting [5]. A recasing model was trained from the same corpus, with a 3-gram KenLM [12] language model.

Tuning We tuned the translation models using MERT [16]. For tuning data, we used the development set of the above mentioned campaign consisting of 887 bi-sentences for each language-pair.

Translation and post-processing Each of the tweets of the PAN training set was preprocessed in the same fashion as was the training data. It was then translated with the trained model of the corresponding language-pair, and finally underwent quick post-processing, namely recasing and detokenization.

4 Experiments

In this section we outline our own internal evaluations of our system. First we report experiments into the parameters of our core pipeline. Following this, our experiments in using machine translation to improve performance of the smaller language subsets.

[1] Version 7, from: `http://www.statmt.org/europarl/`
[2] `http://www.ted.com`
[3] Data from IWSLT 2014 evaluation campaign: `https://wit3.fbk.eu/mt.php?release=2014-01`.
[4] Version 3.0, downloaded 16 Feb 2015 from `http://www.statmt.org/moses/`.

4.1 Experimenting with Learning Framework Parameters

Training data is first passed through the linguistic processing pipeline as described in section 3.1. Subsequently, the data encoded as features, along with the labels are passed to the learning framework. We experimented with a number of parameters which included:

- numeric representation of the features: binary, normalised, or absolute frequency
- feature thresholding (only including features with a frequency greater than a set value).
- dimensionality of the compressed feature space (see section 3.2 for more details)
- the level of classification decisions: per user, or per tweet.

For each combination of settings, we employed the following conditions:

- We use leave-one-out cross-validation on the training data
- Due to random seeding in the bagging used in the cross-validation of the SVD calculations (see section 3.2) we run each setting five times, and average the result.
- Since age is a scale, we use regression as our classification model. To do this, we convert the classes into an ordered scale: 0, 1, 2, 3. Performance on age is reported as mean-squared error, similar to the personality traits.

Results In the interests of space, we do not report all runs here. Generally, we found that thresholding the feature space at $n >= 5$ provided the best results, balancing model performance and computational runtime. Similarly, while increasing the dimensionality generally improved performance, too great an increase significantly impacts runtime. We report only those experiments on the optimum value across all settings of 500. Results are reported in table 3.

The most distinct result is gender: across all languages it is the model trained at a per-tweet level using binary representation of feature frequency that performs the best. Conversely, age – though limited to two languages – shows best performance with normalised frequency at a per-user level.

Results for personality traits are less clear. Overall, the same conditions as for age – per-user, normalised frequency – perform best. In many of the cases in which they do not, the difference in performance is insignificant.

4.2 Experimenting with SMT

Personality labelled data sets can often be smaller than the ones typically used for text classification tasks. This is largely due to the personal nature of the information and the complexity of collecting such labels. This issue of size is particularly clear in the Dutch and Italian datasets (see table 2).

One alternative approach to collecting personality labels is the use of perception ratings – wherein the personality labels are judgements made by third parties. In this work, we explored another approach to answering the data sparsity question, namely machine translation.

Language	Frequency	Level	Gender	Age	EXT	STA	AGR	CON	OPN
EN	user	binary	0.737	0.575	0.168	0.218	0.165	0.153	0.143
EN	user	norm	0.492	**0.464**	0.171	0.223	0.173	0.144	0.146
EN	tweet	binary	**0.805**	0.500	**0.153**	**0.197**	**0.154**	**0.144**	**0.132**
EN	tweet	norm	0.709	0.552	0.156	0.203	0.155	0.144	0.137
ES	user	binary	0.828	0.549	0.161	0.195	0.161	0.172	0.167
ES	user	norm	0.410	**0.327**	**0.153**	0.191	0.163	**0.156**	0.164
ES	tweet	binary	**0.917**	0.525	0.154	**0.188**	**0.155**	0.168	**0.160**
ES	tweet	norm	0.814	0.550	0.156	0.192	0.155	0.168	0.161
IT	user	binary	0.800		0.143	0.170	0.156	0.117	0.153
IT	user	norm	0.395		**0.095**	0.172	0.148	0.106	**0.137**
IT	tweet	binary	**0.893**		0.137	**0.168**	**0.142**	**0.098**	0.141
IT	tweet	norm	0.689		0.158	0.168	0.192	0.136	0.191
NL	user	binary	0.676		0.122	0.182	0.144	0.113	0.109
NL	user	norm	0.318		**0.088**	**0.117**	**0.118**	**0.086**	**0.098**
NL	tweet	binary	**0.852**		0.108	0.164	0.138	0.104	0.104
NL	tweet	norm	0.704		0.109	0.169	0.139	0.104	0.109

Table 3. Results on all language corpora under different feature 'frequency' and classification decision 'level' conditions. Gender is measured in accuracy, age with mean squared error (MSE) and personality traits with root MSE. **Bold** is used to highlight the best result per language-trait pair.

Our main approach was to use the largest corpus – the English – to supplement the remaining smaller datasets. The intention was to see if increasing the size of the dataset, by leveraging non-native labelled data, would improve results. The experiments where conducted thus:

- The English dataset was translated (using the models described in section 3.3) into each of Spanish, Italian and Dutch.
- Each enlarged data set was processed using the linguistic pipeline configured for that language.
- Using the same settings as described for the native language experiments, similar trait classification models were trained using the combined dataset. The results reported here are on the same "leave-one-out" approach, though only the original native non-English dat was used to compute the reported results.

Results The results of these tests can be seen in table 4 along with the results from the previous native experiments for comparison. In the interests of space, we have selected the best result for each language-trait pair. Although Spanish is closest in size to the English corpus, it is also included for completeness.

Overall, the results suggest that translation does not help in the classification of socio-demographic traits. In fact, in many cases – particularly gender – it is considerably detrimental to performance. Despite previous finding that SMT to assist NLP tasks provides at least 'comparable' results, the effect here is worse than expected.

Language	Gender	Age	T0	T1	T2	T3	T4
ES	**0.917**	**0.327**	**0.153**	**0.188**	**0.155**	0.156	0.160
ES + EN2ES	0.798	0.472	0.175	0.206	0.157	**0.146**	**0.140**
IT	**0.893**		**0.095**	0.168	0.142	**0.098**	0.137
IT + EN2IT	0.706		0.115	0.167	**0.127**	0.117	**0.128**
NL	**0.852**		**0.088**	**0.117**	**0.118**	**0.086**	**0.098**
NL + EN2NL	0.740		0.103	0.180	0.143	0.122	0.110

Table 4. Results on the original non-English datasets, compared with a model trained on additional data translated from English. Gender is measured in accuracy, age with mean squared error (MSE) and personality traits with root MSE. **Bold** is used to highlight the best result per language-trait pair.

One issue with working in automatic personality classification is understanding how the manifestation of traits varies between data sources [15]. This likely extends to variations due to language as well. However, we do not enter into further discussion of this topic here, or other aspects which could effect performance such as translation quality. We intend to pursue this in future work.

5 PAN Challenge

In delivering models for the official PAN review, we chose models based on our desire to be as language agnostic as we could be. With this in mind, we chose a single, optimised combination of parameters across all traits and languages. The only variation on this is with gender, for which the settings had a significant – and consistent – impact. As per the settings discussed in section 3, the final parameters for the models uploaded to the evaluation platform are listed in table 5. Additionally, we retain the model dimensionality value of 500.

Trait	Encoding	n	Decision level	Decision model
Gender	binary	5	tweet-level	SVM
Age	normalised	5	user-level	linear regression
Personality Trait	normalised	5	user-level	linear regression

Table 5. Summary of the inference models

Note that though age is a regression in our setting, for the challenge it is converted to a class, rounding the value.

5.1 Challenge Results

The global results can be found in the overview paper for the PAN 2015 Author Profiling challenge [18]. Here we report the results of our system on the evaluation data in table 1. As discussed previously age and gender are measured by accuracy, the personality traits by Root Mean Squared Error (RMSE).

Language	Gender	Age	EXT	STA	AGR	CON	OPN
EN	0.775	0.165	0.167	0.206	0.165	0.148	0.142
ES	0.773	0.136	0.158	0.202	0.136	0.146	0.157
IT	0.806	0.124	0.091	0.215	0.124	0.160	0.169
NL	0.781	0.109	0.135	0.132	0.109	0.062	0.070

Table 6. Results on all language corpora under different feature 'frequency' and classification decision 'level' conditions. Gender and age are reported in accuracy; the remaining traits with root mean squared error (RMSE).

The difference between these results and those of our own tests naturally vary. Gender performance on evaluation is overall lower, but personality traits sees both improvements and worsening of performance. There is no clear pattern in this across language or trait, so there are no general conclusions which can be drawn. Large decreases in expected performance of any trait-language pair (for example NL Extraversion, testing: 0.088, evaluation: 0.135) suggests an overfitting of features under training. Despite attempting to minimise this outcome, with corpora of the sizes of Italian and Dutch, this is to be expected.

We cannot directly compare age, because we used a different metric (MSE compared with accuracy). However, when we compare our performance to others, we see that for English age prediction, we ranked among the lowest in the challenge. It is expected that this is largely due to our choice of regression modelling. As an ordered trait, even performing class-based learning as a regression makes sense. It is clear, however, that our naive approach of rounding our predicted value to a class label does not perform well.

6 Conclusions and Future Work

In this paper we have presented details of XRCE's Personal Language Analytics engine for multi-lingual author profiling. The system we have described leverages our capabilities in natural language processing and machine learning. We have chosen a largely language agnostic approach to this task, which has shown good performance on the four datasets.

We expect to continue this work, further refining our models. In particular we intend to explore the contribution of the individual categories of linguistic features to classification across languages and traits. This, we expect, will also lead to a further understanding of the nature of the relationship between language and personality traits in Twitter.

Related to this, we have also discussed the use of machine translation as a potential means to accommodate for the difficulty of acquiring labelled data of this nature. In the limited context explored here, this has not shown to be helpful. This suggests that though sentiment signals can often be maintained under translation (c.f [2]) the same cannot be said for socio-demographic signals. We intend to look at tuning translation models to be sensitive to these signals, as a step toward personalised translation systems.

References

1. Ait-Mokhtar, S., Chanod, J.P., Roux, C.: A multi-input dependency parser. In: IWPT (2001)
2. Balahur, A., Turchi, M.: Multilingual sentiment analysis using machine translation? In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis. pp. 52–60. WASSA '12, Association for Computational Linguistics, Stroudsburg, PA, USA (2012), http://dl.acm.org/citation.cfm?id=2392963.2392976
3. Celli, F., Lepri, B., Biel, J.I., Gatica-Perez, D., Riccardi, G., Pianesi, F.: The workshop on computational personality recognition 2014. In: Proceedings of the ACM International Conference on Multimedia. pp. 1245–1246. ACM (2014)
4. Cettolo, M., Girardi, C., Federico, M.: WIT^3: Web inventory of transcribed and translated talks. In: Proceedings of EAMT (2012)
5. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. In: Proceedings of the 34th annual meeting on Association for Computational Linguistics (ACL 1996). pp. 310–318 (1996)
6. Dietterich, T.G.: Ensemble methods in machine learning. In: Proceedings of the First International Workshop on Multiple Classifier Systems. pp. 1–15. Springer-Verlag (2000)
7. Gill, A.J., Oberlander, J.: taking care of the linguistic features of extraversion
8. Golub, G.H., Reinsch, C.: Singular value decomposition and least squares solutions. Journal of Numerical Mathematics 14, 403–420 (1970)
9. Halko, N., Martinsson, P.G., Tropp, J.A.: Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review 53(2), 217–288 (2011)
10. Han, B., Cook, P., Baldwin, T.: Automatically constructing a normalisation dictionary for microblogs. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012). pp. 421–432. Jeju Island, Korea (2012)
11. Hansen, P.C.: The truncated svd as a method for regularization. Tech. rep. (1986)
12. Heafield, K.: KenLM: faster and smaller language model queries. In: Proceedings of the EMNLP 2011 Sixth Workshop on Statistical Machine Translation. pp. 187–197. Edinburgh, Scotland, United Kingdom (July 2011), http://kheafield.com/professional/avenue/kenlm.pdf
13. Koehn, P.: Europarl: A parallel corpus for statistical machine translation. In: Proceedings of MT Summit (2005)
14. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E.: Moses: Open source toolkit for statistical machine translation. In: Proc. of ACL Demo and Poster Sessions (2007)
15. Nowson, S., Gill, A.J.: Look! Who's Talking? Projection of Extraversion Across Different Social Contexts. In: Proceedings of WCPR14, Workshop on Computational Personality Recognition at ACMM (22nd ACM International Conference on Multimedia) (2014)
16. Och, F.J.: Minimum error rate training in statistical machine translation. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1 (ACL 2003). pp. 160–167. ACL '03 (2003)
17. Rammstedt, B., John, O.P.: Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. Journal of Research in Personality 41(1), 203–212 (Feb 2007), http://dx.doi.org/10.1016/j.jrp.2006.02.001

18. Rangel, F., Celli, F., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd author profiling task at pan 2015. In: Cappellato, L., Ferro, N., Gareth, J., San Juan, E. (eds.) CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings, CEUR-WS.org (Sep 2015), http://www.clef-initiative.eu/publication/working-notes
19. Schapire, R.: The strength of weak learnability. Journal of Machine Learning Research 5 (1990)
20. Stolcke, A.: SRILM - an extensible language modeling toolkit. In: Proceedings Int. Conf. on Spoken Language Processing (INTERSPEECH 2002). pp. 257–286 (2002)
21. Tkalčič, M., Carolis, B.D., de Gemmis, M., Odić, A., Košir, A.: Preface: Empire 2014. In: Proceedingsof the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014). CEUR-WS.org (July)