The *Epistle to Cangrande* through the Lens of Computational Authorship Verification

Silvia Corbara

Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche
56124 Pisa, Italy
silvia.corbara@isti.cnr.it

Abstract. The *Epistle to Cangrande* is one of the most debated documents in the production of the Italian poet Dante Alighieri. For more than a hundred years scholars have been debating over its real paternity, whether it should be considered a work by Dante or a malicious forgery by an unnamed author. In this work, we try to address this philological problem through the methodologies of computational authorship verification and machine learning, by training a classifier on a dataset of medieval Latin prose texts and by using a set of authorship-related features. Although the project is still in a preliminary phase, the early results seem to confirm the hypothesis of a forgery.

Keywords: Machine Learning · Authorship Verification · Digital Humanities · Dante Alighieri · Medieval Latin.

1 The *Epistle to Cangrande*, and the Debate about its Author

The *Epistle to Cangrande*, from now on EpXIII, is the thirteenth of the letters from Dante's epistolary corpus survived until our times. Written in Latin, it is addressed to Can Francesco della Scala, known as Cangrande I, the ruler of the Italian cities of Verona and Vicenza at the beginning of the 14th century. Traditionally, it is divided into 90 paragraphs, where two thematic portions can be defined: the first one (Ep. 1–13) is the dedicatory section, with proper epistolary characteristics, while the second one (Ep. 14–90) contains an exegesis of Dante's *magnum opus*, including an *accessus* of the *Divina Commedia* and an *expositio textus* of the first few lines of the third cantica, the *Paradiso*. Among Dante's letters, it became rather renowned over the centuries, especially because it would be the only analysis we have received from Dante Alighieri about his own masterpiece.

However, in more recent times the authenticity of EpXIII has been questioned, starting from the study of Scolari in 1819 [22, pp. 12-21]. Generally speaking, the academic community can be divided among those who consider EpXIII authentic, and those who consider it a spurious work, either in its entirety or at least in its second portion. To support these claims, on the one hand, some scholars, like for example [3, 6, 20], point out numerous places in the composition where the logical sequence is ambiguous, rusty, or even incoherent within itself or with other writings by Dante. Moreover, many have noticed that there is a deep dissimilarity between the dedicatory and the exegetical section of the letter, in its themes, style and rhythm (see for example [11]). Even figuring out a time-frame when the letter could

have been written is not a trivial problem. On the other hand, others (like [2, pp. 280-1] [12, pp. 67-8]) note that there is a lexical coherence spreading through the entire EpXIII, and an inner cohesive logic. Additionally, [1] observes that a forger would have followed more closely Dante Alighieri's prose, and thus the dissimilar style should be seen paradoxically as a further proof of paternity. Many also note that the author of EpXIII offers some non-traditional and potentially controversial explanations for some exegetical and linguistic issues, which could be the proof of a prominent author, since a common forger would have probably stayed on more ordinary, "safer" grounds. For a more comprehensive outline over this authorship debate, see the analysis by [2, 21].

Given this debated and yet unsolved problem, and in order to gain a fresh perspective over it and thus offer the scholars yet another useful tool for investigation, this project was set to apply the methodologies and technologies of Computational Authorship Verification to the mystery of EpXIII.

2 Computational Authorship Verification: An Overview

Generally speaking, *Authorship Analysis* (AA) can be defined broadly as "any attempt to infer the characteristics of the creator of a piece of linguistic data" [13, p. 238], which includes the author's biographical information (age, gender, mother tongue, etc.) and writing style (use of pronouns, vocabulary richness, etc.), as well as his or her identity. The core of this practice, also known as "stylometry", relies on the idea to identify a certain author not by the artistic value of her/his work, or the meaning of the concepts proposed within it, but via a measure of her/his style. Here, "style" is intended as a summary statistic emerging from one or more numerical features that describe linguistic events present in the written texts, which are believed to remain more or less constant in one's production and conversely vary in noticeable fashion across different authors [13, p. 241].

These unique stylistic features are also known as "style markers". This definition embraces every kind of textual event, as long as it can be counted (hopefully: easily counted). It is the researcher's task to identify and extract the most discriminative features for the type of research s/he is dealing with. In particular, scholars started experimenting with this practice (well before the age of computers) by employing a single set of features comparable to the linguistic elements studied in classical philology, such as the frequencies of certain terms or of word and sentence lengths [18, 24]. However, in the late 20th century, starting from the work of Mosteller and Wallace on the *Federalist Papers*[19], the practice veered towards employing several sets of high-frequency features in parallel. Even though this approach captures textual components of apparently minimal significance, this practice has proven effective in a variety of tasks, since the phenomena involved tend to be out of the conscious control of the author, hence difficult to modify or imitate. The noted historian Carlo Ginzburg describes this change in the cultural paradigm in his essay *Clues*, calling it the *Evidential Paradigm* [9].

The values of these stylistic features are collectively used as a simplified representations of the respective text, and employed for analyzing its authorship. This may be done via various methods; usually, these methodologies are classified as similarity-based or machine learning -based. In particular, in the latter class, a classifier is trained from a number of labelled training examples, using vectors of the chosen features (the style markers) as representations of the texts of interest; this enables the machine to leverage the values of the features in the training examples in order to classify new unlabelled documents into the proper class. These techniques come from the field of text classification and may be tuned to a variety of subtasks: the classes may be literary genres, topics, languages, and so on, depending on the

goal of the research. In fact, this is the principle behind many modern applications, from email-spam detecting to authorship identification for forensic cases.

In computational AA, the most popular methods still make use of similarity-based measures or "classical" machine learning algorithms, such as support vector machines (SVMs) and logistic regression (LR), even if deep learning algorithms have sometimes proved more accurate. This trend was also confirmed in the PAN 2018 Author Identification shared task [15], where the majority of the systems presented was based on SVMs. This is due to two different reasons: on the one hand, in some application domains there is a systematic scarcity of annotated data, which clashes with the fact that deep learning methods typically require very large training sets; on the other hand, deep learning methodologies notoriously lack on the explainability side, which could be daunting when the investigation concerns a case of genuine controversy, and it is indeed desirable that the factors supporting the conclusion drawn by the system can be properly exhibited [13, p. 307]. These issues are especially true in the humanities, where the documents available are usually rather limited in number (as in the case of medieval Latin), and the main objective of the computational studies is certainly not to replace the professional philologist, but to support her/his research with supplementary evidence and tools, which then need to be as explicit as possible. Some examples of AA in the humanities can be found in [4, 14, 23].

Within this eclectic landscape, the problem of EpXIII is an instance of *Authorship Verification* (AV), a subtask of AA that aims to design methods and techniques to determine whether a document of unknown or disputed paternity has been written by a given candidate author. It is thus different from *Authorship Attribution*, where the goal is to infer, for a document of unknown or disputed paternity, the most likely author among a finite set of candidate authors [16]. The verification problem itself is often formulated as a binary classification task, where the works from authors different from the candidate are used as negative training examples.

3 AV Methods Applied to the Dantean Case

We approach the problem of the authorship of EpXIII as a supervised binary classification task implemented via linear classifier. In Section 3.1, we give the details of the system employed for the task, while in Section 3.2 we show the training corpus we have assembled; finally, in Section 3.3 we list the results of the related experimentation.

3.1 Classifier and authorial features

After a few initial tests with LR and SVMs, we decided to employ the former, since the preliminary experiments on our data had indicated that the two have a similar level of accuracy, and, unlike SVMs, the output of LR admits a probabilistic interpretation, i.e., it can be interpreted as the ("posterior") probability that the document belongs to the class. In particular, after the training phase, the algorithm takes the independent variables (the feature values for the anonymous document) as an input, and computes a value within the logistic function - thus in the interval $(0, 1)$ - as an output. The output is the probability, given the variables, for a certain outcome (belonging to a class, in this case to the Dante class) to happen. See [5, pp. 205-6] for a more complete description of LR.

As already detailed in Section 2, computational methods for AV represent the textual document as a vector of variables representing some linguistic phenomena. To this aim, we have selected a combination of various feature types, since this approach usually assures better performances while dealing with multivariate

methods [10]. Each feature type has been shown to be effective to some extent in similar authorship-related tasks. The resulting set of features is composed as follows:

- Character n-grams ($n \in \{3, 4, 5\}$)
- Word n-grams ($n \in \{1, 2\}$)
- Function words (from a list of 74 Latin function words)
- Verbal endings (from a list of 245 regular Latin verbal endings)
- Word lengths (from 1 up to 23 characters)
- Sentence lengths (from 3 up to 70 words)

We subject the features resulting in a sparse distribution to feature selection (via Chi-square) and feature weighting (via TF-IDF).

3.2 Assembling the corpus

As already explained in Section 1, EpXIII consists of two sections that are very distinct for purpose and style; moreover, many scholars theorize a different hand for each of them. Hence, it seemed imperative to follow the traditional division and split the AV problem into two sub-problems, training a specific classifier for the paragraphs 1-13 (from now on: EpXIII(I)) and another one for the remaining paragraphs 14-90 (from now on: EpXIII(II)).

In order to offer to the classifier an adequate representation of the two authorial classes (Dante and NotDante), we created two corpora of medieval Latin texts by collecting documents which can be considered, linguistically and stylistically speaking, close to the related portion of EpXIII. Understandably, the positive class (Dante) is represented by the works unquestionably ascribed to Dante: the other 12 letters for EpXIII(I), *De Vulgari Eloquentia* and *Monarchia* for EpXIII(II). Conversely, for the negative class (NotDante), two sets of texts of coeval Latin authors have been assembled: a number of epistles from various authors (EpXIII(I)) and a collection of different textual works, mostly literary comments and treatises (EpXIII(II)). All the documents are dated between the 13th and the 15th century. Additionally, we subject the training documents to a segmentation policy in order to further increase the number of training samples. Using a large number of training segments for each author when only few long texts are available is a common practice in the field [17, p. 514]. The final result of this procedure is shown in Table 1.

Table 1. The number of full documents compared to the number of training documents

	# full docs	**# training docs** (full + segments)
EpXIII(I)	294	1310
EpXIII(II)	30	12312

3.3 The Classification Hypotheses

In order to determine the degree of reliability of the classifier, and hence establish the trustworthiness of the classification hypotheses it generates, we subject the algorithm to a "leave-one-out" validation test; more specifically, we create an AV classifier for each author that has more than one positive sample in the corpus, setting the author as the positive class. In particular, as to recreate the condition of the actual classification of EpXIII(I) and EpXIII(II) and to avoid overlapping between test and training samples, we only test the full original documents and,

when the document t is used as test, all the segments derived from t are excluded from the training. The results of the validation test are shown in Table 2; we use Macro-F_1 and Micro-F_1 as evaluation measures. As it can be seen, the classifiers that have been trained, and especially the one for EpXIII(I), obtain a good level of accuracy, in line with other state-of-the-art methods.

Table 2. The results of the validation test

	Macro-F_1	**Micro-F_1**
EpXIII(I)	0.88	0.98
EpXIII(II)	0.68	0.77

Finally, the classification hypotheses are shown in Table 3, along with the F_1 values for the class Dante. As it can be seen, the classifiers consider both portions of EpXIII written by someone other than Dante, with a slightly lower probability, and thus higher confidence in the negative attribution, for EpXIII(I). Considering the specific accuracy level obtained in the validation, while the value for EpXIII(I) is satisfying, the same cannot be said about the value for EpXIII(II); the classifier is penalized for not attributing *Monarchia* to Dante, while it correctly classifies 26 out of the 28 negative samples.

Table 3. The classification hypotheses

	Hypothesis	Pr**(Dante)**	F_1**(Dante)**
EpXIII(I)	NotDante	0.24	0.95
EpXIII(II)	NotDante	0.39	0.40

Henceforth, the hypotheses computed seem to align with the theory that the entire EpXIII was the production of a malicious forger. Nevertheless, the evaluation presented here should not be considered conclusive: in future developments we plan to define some improvements to the system, such as an expansion of the training corpus and of the feature set exploited, as well as experimenting with some different visualization techniques.

4 Conclusion

In this project, we explore a case of AV applied to Latin prose, in order to infer the real author of the *Epistle to Cangrande*, a letter of dubious Dantean paternity, via computational and quantitative methods.

The classification hypotheses we reach seem to confirm the forgery thesis. However, the current stage of the research is still preliminary. It is important to underline that our goal is not to reach some definitive answer for the philological dilemma; especially for this kind of problems, where the truth is debated even among specialists, the hypothesis given by a classification algorithm can not, and should not aim to, be final. Nevertheless, we are eager to explore other methodological possibilities that could determine an improvement in the system accuracy, and generally create a more accessible and solid tool that can aid the study of the researcher, not replace it.

Acknowledgments. The research described in this paper is part of the work carried out by the author for her MSc thesis in Digital Humanities at the University of

Pisa (supervisors: Alejandro Moreo, Fabrizio Sebastiani, Mirko Tavoni), discussed in February 2019 and also reported in [7, 8].

References

1. Ascoli, A.R.: Access to authority: Dante in the Epistle to Cangrande. In: Baranski, Z.G. (ed.) Seminario Dantesco Internazionale / International Dante Seminar 1, pp. 309–52. Le Lettere, Firenze (1997)
2. Azzetta, L.: Nuova edizione commentata delle opere di Dante, vol. 5, chap. Epistola XIII, pp. 271–487. Salerno Editrice, Rome, Italy (2016)
3. Baranski, Z.G.: Comedía. Notes on Dante, the Epistle to Cangrande and medieval comedy. Lectura Dantis (8), 26–55 (1991)
4. Binongo, J.N.G.: Who wrote the 15th Book of Oz? An application of multivariate analysis to authorship attribution. Chance **16**(2), 9–17 (2003)
5. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
6. Casadei, A.: Sempre contro l'autenticità dell'Epistola a Cangrande. Studi danteschi **LXXXI**, 215–46 (2016)
7. Corbara, S.: The Epistle to Cangrande through the Lens of Computational Authorship Verification. Master's thesis, Department of Philology, Literature, and Linguistics, University of Pisa, Pisa, IT (2019)
8. Corbara, S., Moreo, A., Sebastiani, F., Tavoni, M.: L'epistola a Cangrande al vaglio della computational authorship verification: Risultati preliminari (con una postilla sulla cosiddetta "XIV Epistola di Dante Alighieri"). In: Casadei, A. (ed.) Atti del Seminario "Nuove Inchieste sull'Epistola a Cangrande". Pisa University Press, Pisa, IT (2019), forthcoming
9. Ginzburg, C.: Clues, Myths and the Historical Method, chap. Clues: Roots of an Evidential Paradigm, pp. 96–125. Johns Hopkins University Press, Baltimore (1989)
10. Grieve, J.: Quantitative authorship attribution: An evaluation of techniques. Literary and Linguistic Computing (22), 251–70 (2007)
11. Hall, R.G., Sowell, M.U.: Cursus in the Can Grande Epistle: A forger shows his hand? Lectura Dantis (5), 89–104 (1989)
12. Jenaro-MacLennan, L.: The Trecento commentaries on the Divina Commedia and the Epistle to Cangrande. Clarendon Press, Oxford (1974)
13. Juola, P.: Authorship attribution. Foundations and Trends in Information Retrieval **1**(3), 233–334 (2006). https://doi.org/10.1561/1500000005
14. Kabala, J.: Computational authorship attribution in medieval Latin corpora: The case of the Monk of Lido (ca. 1101–08) and Gallus Anonymous (ca. 1113–17). Language Resources and Evaluation pp. 1–32 (2018)
15. Kestemont, M., Tschuggnall, M., Stamatatos, E., Daelemans, W., Specht, G., Stein, B., Potthast, M.: Overview of the Author Identification Task at PAN-2018: Cross-domain authorship attribution and style change detection. Working Notes Papers of the CLEF 2018 Evaluation Labs pp. 1–25 (2018)
16. Koppel, M., Schler, J., Argamon, S.: Computational methods in authorship attribution. Journal of the American Society for Information Science and Technology **60**(1), 9–26 (2009)
17. Luyckx, K., Daelemans, W.: Authorship attribution and verification with many authors and limited data. In: Proceedings of the 22nd International Conference on Computational Linguistics. pp. 513–20. Manchester, UK (2008)
18. Mendenhall, T.C.: The Characteristic Curves of Composition. Science (9), 237–249 (1887)
19. Mosteller, F., Wallace, D.: Inference and Disputed Authorship: The Federalist. CSLI Publications, Stanford, 2007 edn. (1964)
20. Nardi, B.: Il punto sull'Epistola a Cangrande. In: "Lecturae" e altri studi danteschi, pp. 205–25. Le Lettere, Firenze (1990)
21. Sasso, G.: Sull'Epistola a Cangrande. La Cultura (3), 359–446 (2013)
22. Scolari, F.: Note ad alcuni luoghi delli primi cinque canti della Divina Commedia. Tip. Picotti, Venezia (1819)

23. Tuccinardi, E.: An application of a profile-based method for authorship verification: Investigating the authenticity of Pliny the Younger's letter to Trajan concerning the Christians. Digital Scholarship in the Humanities **32**(2), 435–47 (2016)
24. Yule, G.U.: On sentence length as a statistical characteristic of style in prose with application to two cases of disputed authorship. Biometrika (30), 363–90 (1938)