Literary Detective Work on the Computer. Michael Oakes.

Walter Daelemans

Literary detective work on the computer. Michael P. Oakes. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2014. ISBN: 978-90-272-4999-9.

Michael Oakes is a reader in Computational Linguistics at the University of Wolverhampton and an expert in statistical and corpus-based methods for language research. This book integrates, in textbook style, recent and old research about the application of quantitative methods to the automatic analysis of authorship and author profiles on the basis of properties of the text, a discipline often called computational stylometry. In addition, other language-related "detective work", such as decipherment of old scripts and plagiarism detection are discussed as well.

The currently dominant approach in computational stylometry is based on automatic text categorization, the field of computer science that also brought us spam filtering. As an example, for author identification, this approach would entail (i) defining a representation of text in terms of (mainly) linguistic properties, (ii) training a model using statistical or machine learning methods on the basis of such representations of texts with known authorship, and (iii) applying the learned model to unknown texts in order to decide authorship. This approach is based on the belief that when the linguistic properties used to represent a text are well chosen, the quantitative methods in (ii) can learn how individual authors differ in style: author style would then be an idiosyncratic combination of preferences in the use of the linguistic properties modeled in the text representations. Also authorship methods based on some form of similarity or overlap between linguistic properties of different texts fit this scheme as an instance of nearest-neighbor classification. Apart from this supervised learning, unsupervised techniques are used as well, clustering the texts after step (i) automatically into groups with similar stylistic properties.

For readers new to the subject, a more extensive and systematic introduction from the point of view of this current practice, and the different steps it entails, would have been useful as a framework to interpret the first two chapters, in which old and new methods, scenarios and problems, mathematical background, supervised and unsupervised machine learning methods, as well as experimental design decisions related to the different stages described above, are mixed. For understanding the issues in step (i), a systematic introduction to the state of the art of Natural Language Processing (NLP), i.e. morphological analysis, parsing, and semantic analysis, would be essential as well. However, readers who already acquired a good mental framework of the field will find a wealth of issues and methods, clearly described, if not always in the places where you would expect them. The unsystematically interspersed R commands and programs in the text are probably confusing for beginners and too elementary for many practitioners, but don't hinder the flow of the text too much. Still, I think it is a missed chance that these were not collected more systematically in a separate chapter or appendix. Before starting with this book, it might therefore be useful to read some older but more systematic introductions such as Juola (2006), Koppel et al. (2009), or Stamatatos (2009).

A curious point of view implicit in Chapter 1 is that Oakes seems to suggest that 'optimal' linguistic feature selections and 'best' machine learning methods exist for the task of authorship identification and profiling in general. In reality, the best features and method will be different for different specific data sets and tasks. Evaluation is only addressed late in this chapter whereas it is central in Machine Learning approaches in other areas. Indeed, the absence of proper evaluation makes many of the older studies in authorship attribution mentioned in this book unreliable if not irrelevant. In

Chapter 2, plagiarism and spam filtering are addressed, admittedly examples of "detective work", but not really applications of computer stylometry. Exceptions are intrinsic plagiarism detection (based on the detection of style changes within the same text) and detection of the fact that a text is rewritten rather than original, but these topics are now buried deeply into the chapter.

The next two chapters address several cases of authorship attribution in practice, related to Shakespeare and religious texts (New Testament, Book of Mormon, and Qu'ran). These are fascinating applications that drive home the message that the quantitative approach in computational stylometry can be a game-changing tool for solving important questions in the humanities and may become the standard approach, not unlike what carbon dating has been for dating problems in archeology. Of course, for those problems we don't know the truth, so it is essential that the methods used to attribute "dubitanda" texts should be tested by some form of cross-validation on that part of the data which is (relatively) certain, for example the undisputed texts of Shakespeare. This does not seem to be the case for many of the older studies described in these chapters, which in my opinion makes their results dubious. The overview also shows that many (especially supervised) learning systems have not yet been tried on these interesting datasets; unsupervised clustering techniques seem to dominate these applications. Again, not all previously published work in this area would stand the test of current methodology. Chapter 5 addresses decipherment of languages. This work definitely fits the title of the book but not so much the field of stylometry introduced in preface and introductory chapters. The focus here is on the captivating cases of Rongorongo (possibly an Easter Island language script) and the Indus valley texts.

This book is a valuable repository of techniques, methods, tasks, cases, and background relevant to computational stylometry. I admire the way in which Oakes' interpretation of his own research and that of others is supports deeper understanding of the tasks tackled. This is very different from the current practice of computational stylometry in machine learning and NLP which seems to focus on accuracies and f-scores rather than on explaining the results (see Daelemans, 2013 for a critique of that approach). The book could have been structured more didactically, however. Everything is there, but unsystematically distributed over the monograph. I would expect beginners to be put off by the way the material is ordered and introduced. But as soon as a certain level of experience with the field is reached, the book will be a great companion and source of inspiration.

Daelemans, W. (2013). Explanation in computational stylometry. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 451-462). Springer Berlin Heidelberg.

Juola, P. (2006). Authorship attribution. Foundations and Trends in information Retrieval, 1(3), 233-334.

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. Journal of the American Society for information Science and Technology, 60(1), 9-26.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. Journal of the American Society for information Science and Technology, 60(3), 538-556.