headsets are surprisingly limited. Our only real choice was Oculus Rift. Another headset, the HTC Vive, is set for release soon. Most headsets are developed and sold as part of a complete VR system.

For the chair-assembly, a unique on-campus resource simplified our choice. The University of Oklahoma houses a high-powered physics fabrication lab. We worked with them to develop a custom railed-chair assembly (ergonomically designed for a 360° range of motion). This railed-chair allows the computer to reside under the chair and out of the way. For a robust virtual environment, the computer contains a GeForce GTX 980 graphics card. It delivers a 75-frames/second refresh-rate (the human eye generally resolves 25 frames/second), insuring an instantaneous visual experience when manipulating 3D objects or when turning one's head.

Finally, by integrating networking software into OVAL, a shared VR experience can occur across a range of clients. All changes made on a master workstation—including scale, rotation, lighting, and background imagery—are immediately transmitted to all co-participants, regardless of their physical location. In a classroom environment, for example, this means that students automatically see what the teacher sees. But this also allows OVAL to become a worldwide network. To facilitate such a network, all 3D models are uploaded via a public Dropbox, which immediately syncs with all OVAL clients. This means that all uploaded 3D asset are available to all OVAL clients. For a shared VR experience, each client only needs a short set of instructions concerning file names and how to manipulate them during a session.

Research and teaching

In our presentation, we will also discuss ongoing uses of OVAL at the University of Oklahoma and explore their implications. Despite its recent completion, OVAL has already had extensive use. Undergraduate biology students have analyzed the atomic structure of hemoglobin and oxyhemoglobin. Architecture faculty has analyzed student projects for unseen flaws pertaining to safety and accessibility of interior spaces. The Sam Noble Museum of Natural History has uploaded their recently discovered *Aquilops Americanus* skull into the OVAL system for curators and researchers. Art History faculty has begun analyzing sculpturally significant 3D scans for preserving what was once ephemeral art. A budding partnership with the Medical Imaging Facility has demonstrated how CT-to-OVAL workflows facilitate mammographic research. Finally, Bill Endres has begun to develop guided, immersive tours of the St Chad Gospels, an 8th-century illuminated manuscript.

Conclusion

The rapid production of 3D models makes having VR systems available for their viewing a pressing concern. 3D models of massive structures, such as the large Buddhas of Bamiyan, highlight the limitations of interacting through a computer screen. OVAL provides one cost-efficient solution. In our next phase, we plan to add collaborators and make OVAL available. We are also interested in hosting 3D assets in an archive-quality database. However, the most effective and efficient means of doing these has yet to be determined. We are looking forward to presenting at DH 2016 and conversing about possibilities for OVAL and the wide-ranging opportunities for research and teaching through VR.

Bibliography

Dylla, K., et al. (2010). Rome Reborn 2.0: A Case Study of Virtual City Reconstruction Using Procedural Modeling Techniques. Frischer, B., Crawford, J. and Koller, D. (eds), *Making History Interactive: 37*th *Proceedings of the CAA Conference*, Williamsburg, VA, March 2009. Oxford: Archaeopress, pp. 62-66.

Grün, A., Remondino, F. and Zhang, L. (2004). Photogrametric Reconsctruction of the Great Buddha of Bamiyan, Afghanistan. *The Photogrammetric Record,* **19**(107): 177-99.

Hermon, S. and Kalisperis, L. (2011). Between the Real and the Virtual: 3D Visualization in the Cultural Heritage Domain – Expectations and Prospects. *Virtual Archaeology Review,* **2**(4): 59-63.

Shapiro, L. (ed) (2014). *The Routledge Handbook of Embodied Cognition*. New York: Routledge.

Outliers or Key Profiles? Understanding Distance Measures for Authorship Attribution

Stefan Evert
stefan.evert@fau.de
Universität Erlangen-Nürnberg, Germany

Fotis Jannidis
fotis.jannidis@uni-wuerzburg.de
University of Würzburg, Germany

Thomas Proisl
thomas.proisl@fau.de
Universität Erlangen-Nürnberg, Germany

Thorsten Vitt
thorsten.vitt@uni-wuerzburg.de
University of Würzburg, Germany

Christof Schöch
christof.schoech@uni-wuerzburg.de
University of Würzburg, Germany

Steffen Pielström
pielstroem@biozentrum.uni-wuerzburg.de
University of Würzburg, Germany

Isabella Reger
isabella.reger@uni-wuerzburg.de
University of Würzburg, Germany

The state of the art

Burrows' Delta is one of the most successful algorithms in computational stylistics (Burrows 2002). A series of studies have proven its usefulness (e.g. Hoover 2004, Rybicki & Eder 2011). There are two essential steps in Burrows' Delta. The first is to standardize the relative frequencies of words in a document-term-matrix through a z-score transformation. In the second step, the distances between all texts are calculated. For each word, the difference between the z-score of the word in one and the other text are calculated. The absolute values of the differences are added for all words taken into account. The usual interpretation is that the smaller the sum, the more similar two texts are stylistically, and the more likely it is that they have been written by the same author.

Despite the fact that Burrows' Delta is as simple as it is useful, there is still a lack of a good explanation why the algorithm works so well. Argamon (2002) has shown that the second step in Burrows' Delta is equivalent to taking the Manhattan distance between two points in a multi-dimensional space. He suggests, among other things, using the Euclidean distance instead. An empirical test of his proposals has shown, however, that none of them lead to an improvement in performance (Jannidis et al. 2015).

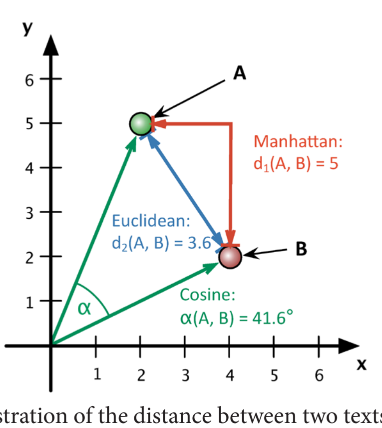

Figure 1: Illustration of the distance between two texts made up of just two words

Smith and Aldrige (2011) have suggested to use the cosine of the angle between the document vectors for the second step, as is customary in Information Retrieval (Baeza-Yates & Ribeiro-Neto 1999:27). The Cosine variant of Delta (Delta *Cos*) outperforms Burrows' Delta (Delta *Bur*) in many different settings and has the advantage of not showing the drop in performance typical ofother Delta variants when large numbers of MFW are used (Jannidis et al. 2015). The question now is why Delta *Cos* is so much better than Delta *Bur* and other variants, that is, in what way Delta *Cos* captures the authorship signal more clearly than other variants of Delta.

Of decisive importance for our further analyses was the insight that using the Cosine Distance is equivalent to a vector normalization in the sense that (in contrast to Manhattan and Euclidean Distance) the length of the vector does not play a role for the calculation of the distance (see figure 1). Previous experiments have shown that an explicit, additional vector normalization also substantially improves performance of the other Delta measures (Evert et al. 2015).

Hypotheses

Having discovered that impact of the normalization effect, we have developed two empirically testable hypotheses:

- (H1) Performance differences are caused by single extreme values, so-called outliers. These are particularly large positive or negative *z-scores* specific to single texts rather than all texts of a single author. As the Euclidean distance should be more sensitive to single extreme values than the Manhattan distance, this hypothesis would explain the comparatively bad performance of Argamon's "Quadratic Delta" Delta *Q*. The positive effect of vector normalization originates from the reduction of outlier amplitudes ("outlier hypothesis").
- (H2) The author specific "style profile" manifests itself more in the qualitative combination of word preferences, i.e. in the pattern of over- and under utilization of vocabulary, rather than in the actual amplitude of *z-scores*. A text distance measure is particularly successful in authorship attribution if emphasizing structural differences of author style profiles without being too much influenced by actual amplitudes ("key-profile hypothesis"). This hypothesis explains directly why vector normalization results in such impressive improvements: it standardizes the amplitudes of author profiles in different texts.

New insights

Corpora

For the experiments in this paper, we use three similarly composed corpora in German, English and French. Each corpus contains 25 different authors with 3 novels each, thus 75 texts in total. The corpora have been described in Jannidis et al. (2015). Due to space issues, the following section will only present our observations on the German corpus. The results for the corpora in both other languages show only small deviations and also support our findings.

Experiments

To further investigate the role of outliers and thus the plausibility of H1, we complement Delta *Bur* and Delta *Q* with additional variants based on the general Minkowski distance (for $p \geq 1$):

$$\Delta_p = \left(\sum_{i=1}^{m} |z_i(D_1) - z_i(D_2)|^{p} \right)^{1/p}$$

We generally name these distance measures L *p*-Delta. The specific case $p = 1$ equals the Manhattan distance (L *1*-Delta = Delta *Bur*), $p = 2$ the Euclidean distance (L *2*-Delta = Delta *Q*). The higher the value for *p*, the larger the influence of single outliers on L *p*-Delta.

Fig. 2 compares four different L *p* distance measures (for p=1, √2, 2, 4) with Delta *Cos*. The method of comparison is the same as in Evert et al. (2015): 75 text are automatically clustered in 25 groups according to Delta distances; clustering quality is estimated with the adjusted rand index (ARI). An ARI of 100% signifies perfect author recognition whereas a value of 0% shows that the clustering is entirely random. The performance of L *p* Delta obviously decreases with increasing *p*. Additionally, the robustness of the measures also decreases with an increasing number of MWF used. As already reported in Jannidis et al. (2015) and Evert et al. (2015), $\text{Delta}_{\text{Bur}}$ (L_1) consistently outperforms Argamon's Delta *Q* (L *2*). Especially if many features, i.e. a large number of MFW is considered, high p values result in low performance. Delta *Cos* is more robust than other variants and achieves almost perfect attribution success (ARI > 90%) over a wide range of the MFW.

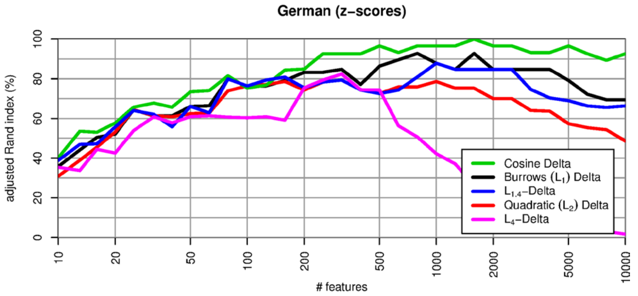

Figure 2: Clustering quality of different Delta measures as a function of the number of the MFW considered

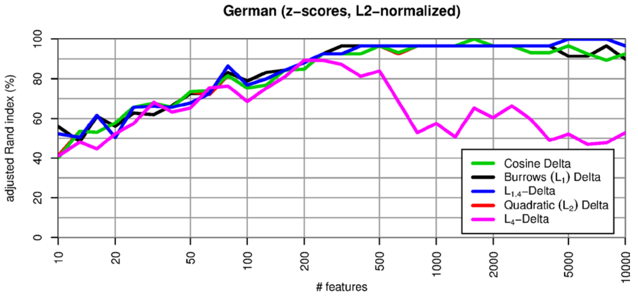

Figure 3: Cluster quality of various Delta measures with length-normalized vectors

Normalizing the feature vectors to length 1 improves the quality of all Delta measures significantly (fig. 3). In this case, Argamon's Delta *Q* is identical to Delta *Cos*: the red line is completely covered by the green one. The other Delta measures (Delta *Bur*, L *1.4*-Delta) now reach about the same quality as Delta *Cos*. Only L *4* Delta, which is especially prone to outliers, falls short considerably. These results seem to support H1.

A different approach to limit the influence of outliers is to truncate extreme *z-scores*. To do so, we set all $|z| > 2$ to +2 or −2, depending on the original *z-scores*'s sign. Fig. 4 shows the effects of various normalizations on the distribution of the feature values. Vector length normalization (lower left) produces only slight changes and practically does not reduce the number of outliers at all. Pruning large *z-score* values only affects words with above-average frequencies (upper right).

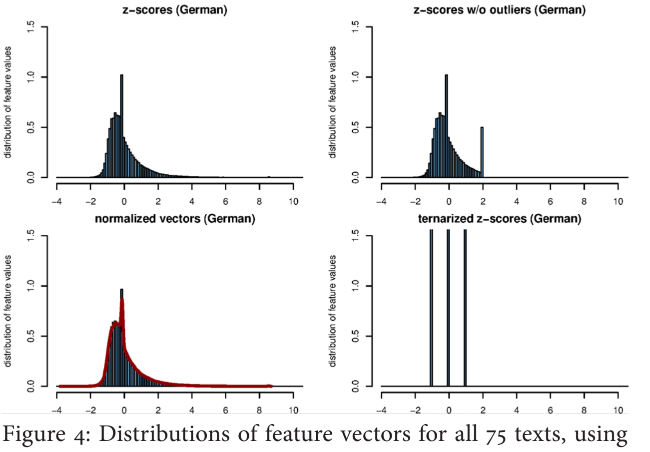

Figure 4: Distributions of feature vectors for all 75 texts, using vectors of 5000 most frequent words. The table shows the distribution of the original *z-scores* (upper left), the distribution after length-normalizing the vectors (lower left), the distribution after clamping outliers with $|z| > 2$ (upper right) and a ternary quantization to the values −1, 0 and +1 (lower right). The red curve in the lower left graph shows the *z-scores* before normalization; the direct comparison shows the normalization has only minimal effect and almost does not reduce outliers. The thresholds for the ternary quantization, $z < -0.43$ (−1), $-0.43 \leq z \leq 0.43$ (0) and $z > 0.43$ (+1), have been selected such that in an ideal normal distribution, a third of all feature values would fall into each of the classes −1, 0, and +1.

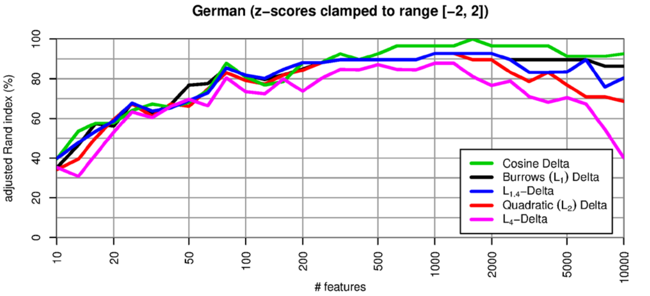

Figure 5: Cluster quality after clamping outliers, i.e. feature values with $|z| > 2$ have been replaced with the fixed values −2 or +2, depending on *z-score*'s sign

As Fig. 5 shows, this manipulation improves the performance of all L p Deltas considerably. However, its positive effect is noticeably smaller than that of vector normalization.

With these differing effects of the normalizations on outlier distributions and Delta results, H1 cannot be upheld. H2 is supported by the good results of vector length normalization. However, on its own, it cannot explain why clamping outliers leads to a considerable improvement as well. To examine this hypothesis further, we created pure "key profile" vectors that only discriminate between word frequencies that are above average (+1), unremarkable (0), and below average (−1; cf. Fig. 4, lower right).

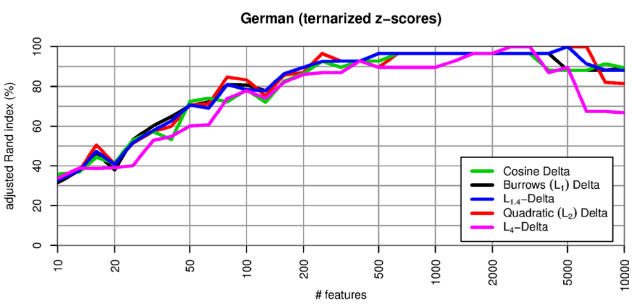

Figure 6: Cluster quality with ternary quantization of the vectors in frequencies that are above average (+1, $z > 0.43$), unremarkable (0, $-0.43 \leq z \leq 0.43$), and below average ($z < -0.43$)

Fig. 6 shows that these key profile vectors perform remarkably well, almost on par with vector normalization. Even the especially outlier-prone L *4* Delta reaches a quite robust clustering quality of more than 90%. We interpret this observation as giving considerable support to hypothesis H2.

Discussion and perspectives

H1, the outlier hypothesis, has been disproven as the vector normalisation hardly reduces the number of extreme values and the quality of all L p measures is still considerably improved. On the other hand, H2, the key profile hypothesis, has been confirmed. The ternary quantification of the vectors shows clearly that it is not the extent of deviation resp. the size of the amplitude, but the profile of deviation across the MFW which is important. Remarkably, the measures behave differently if more than 2000 MFW are used. Almost all variant show a decline for a very large number of features, but they differ in when this decline starts. We suppose that the vocabulary in those parts is less specific for an author than for topics and content. Clarifying such questions will require further experiments.

Bibliography

Argamon, S. (2008). Interpreting Burrows's Delta: Geometric and Probabilistic Foundations. *Literary and Linguistic Computing*, **23**(2): 131–47 doi:10.1093/llc/fqn003. http://llc.oxfordjournals.org/content/23/2/131.abstract.

Baeza-Yates, R. and Ribeiro Neto, B. (1999). *Baeza-Yates, Ricardo; Ribeiro Neto, Berthier (1999): Modern Information Retrieval. Harlow.* Harlow.

Burrows, J. (2002). 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, **17**(3): 267–87 doi:10.1093/llc/17.3.267. http://llc.oxfordjournals.org/content/17/3/267.abstract.

Eder, M. and Rybicki, J. (2011). Deeper Delta across genres and languages: do we really need the most frequent words?. *Literary and Linguistic Computing*, **26**(3): 315–21 doi:10.1093/llc/fqr031. http://llc.oxfordjournals.org/content/early/2011/07/14/llc.fqr031.abstract .

Evert, S., Proisl, T., Pielström, S., Schöch, C. and Vitt, T. (2015). Towards a better understanding of Burrows's Delta in literary authorship attribution. *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver CO.

Hoover, D. L. (2004). Testing Burrows's Delta. *Literary and Linguistic Computing*, **19**(4): 453–75 doi:10.1093/llc/19.4.453. http://llc.oxfordjournals.org/content/19/4/453.abstract.

Jannidis, F., Pielström, S., Schöch, C. and Vitt, T. (2015). Improving Burrows' Delta – An empirical evaluation of text distance measures. *Digital Humanities 2015 Conference Abstracts*. Sydney: ADHO http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empi/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empirical_.html.

Smith, P. W. H. and Aldridge, W. (2011). Improving Authorship Attribution: Optimizing Burrows' Delta Method. *Journal of Quantitative Linguistics*, **18**(1): 63–88 doi:10.1080/09296174.2011.533591. http://www.tandfonline.com/doi/abs/10.1080/09296174.2011.533591.