The Challenge of Sentiment Quantification*

Fabrizio Sebastiani
Qatar Computing Research Institute
Hamad bin Khalifa University
PO Box 5825 Doha, Qatar
fsebastiani@qf.org.qa

Among the many challenges that sentiment analysis (SA) faces, I want to concentrate on one which has not received much attention within the SA community, but that is going to play a major role in future applications: *Sentiment Quantification* (SQ) (Esuli and Sebastiani, 2010). *Quantification* is defined as the task of estimating the prevalence (i.e., relative frequency) of the classes of interest in a set of unlabelled data via supervised learning (Forman, 2008); examples of SQ are (i) determining the prevalence of endorsements in a set of tweets about a political candidate, or (ii) determining the prevalence of rebuttals in a set of reviews of a given book. A naïve way to tackle quantification is by classifying each unlabelled item independently and computing the fraction of such items that have been attributed the class. However, a good classifier is not necessarily a good quantifier: assuming the binary case, even if $(FP+FN)$ is comparatively small, bad quantification accuracy results if FP and FN are significantly different (since perfect quantification coincides with the case $FP = FN$). This has led researchers to study quantification as a task on its own right, rather than as a byproduct of classification.

Within SA, quantification plays a major role, since in many applications we are interested in estimating sentiment not at the individual level, but at the aggregate level. For instance, when SA is applied to tweets, it is rarely (if at all) the case that we are interested in the sentiment conveyed by an individual tweet (Gao and Sebastiani, 2015): it is the sentiment of the crowd, and how it is distributed, that we are instead interested in, and monitoring this distribution over time is a holy grail within fields such as the social sciences, market research, and online reputation management.

Quantification is still an under-researched area, due to the fact that for years it has not been identified as a task on its own. Challenges that are still in need of satisfactory solutions are:

- Can we devise quantifiers that deliver high accuracy irrespectively of the level of *distribution drift* (i.e., the difference between class prevalence in the labelled and in the unlabelled sets)?
- Can we devise accurate quantifiers that do not use the classification of individual items as an intermediate step? And is this the best way to tackle quantification?
- Aside from the binary case, can we devise accurate quantifiers also for the multiclass case (i.e., when the mutually exclusive classes are > 2) and for the ordinal case (i.e., when there is a total order defined on the set of classes)?

References

Andrea Esuli and Fabrizio Sebastiani. 2010. Sentiment quantification. *IEEE Intelligent Systems*, 25(4):72–75.

George Forman. 2008. Quantifying counts and costs via classification. *Data Mining and Knowledge Discovery*, 17(2):164–206.

Wei Gao and Fabrizio Sebastiani. 2015. Tweet sentiment: From classification to quantification. In *Proceedings of the 7th International Conference on Advances in Social Network Analysis and Mining (ASONAM 2015)*, pages 97–104, Paris, FR.

*The author is on leave from Consiglio Nazionale delle Ricerche, Italy

Proceedings of NAACL-HLT 2016, page 173,
San Diego, California, June 12-17, 2016.