Sarcasm Detection: Beyond Machine Learning Algorithms

Vineet Kumar
Capgemini
{vineet.k}@capgemini.com

ABSTRACT

Noise in online networks especially knowledge networks such as *Quora*, *Yahoo! Q&A*, *reddit* can be attributed to jokes, redundancy, insults, sarcasm. As the size of the content on these websites grows in a manner not possible to be monitored manually, there is a need to automatically detect the undesired text to increase the signal (useful content) to noise ratio. Popular machine learning algorithms does a good job at predicting noise in online social networks, such as sarcasm in Twitter, Davidov et. al. (2010) [1]. However, sarcasm detection remains a challenge for learning algorithms because of the complexity involved in learning the patterns. Some statement in topic *Physics* on *Quora* might be sarcastic without containing any word features normally found in general sarcastic sentences. For instance: "*I know that this defies the law of gravity, but, you see, I never studied law*" is a joke without containing any typical word features like oh!, !, yeah! etc., found in sarcastic comments. However to tag it as a sarcasm, reader must know the context i.e. law of gravity and law subject. Even humans will have difficult time classifying whether a piece of text is sarcastic or not without knowing the context. Since humans need context to determine sarcasm, so must be the learning algorithms as experimented by Wallace et. al. (2014) [2]. This generates a need to develop intelligent machine learning algorithms which can take complex context and subject matter knowledge to detect specialized sarcasm.

BODY

Learning Algorithms just like humans need context to detect sarcasm.

REFERENCES

[1] D. Davidov, O. Tsur, and A. Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceeding of the 23rd international conference on Computational Linguistics (COLING)*, July 2010.

[2] B. C. Wallace, D. K. Choe, L. Kertz, and E. Charniak. Humans require context to infer ironic intent (so computers probably do, too). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 512–516, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

Volume 3 of Tiny Transactions on Computer Science