

[Inspiratron.org](https://inspiratron.org) - Natural language processing, machine learning and cybersecurity

Personalized relevance classifier of sentences

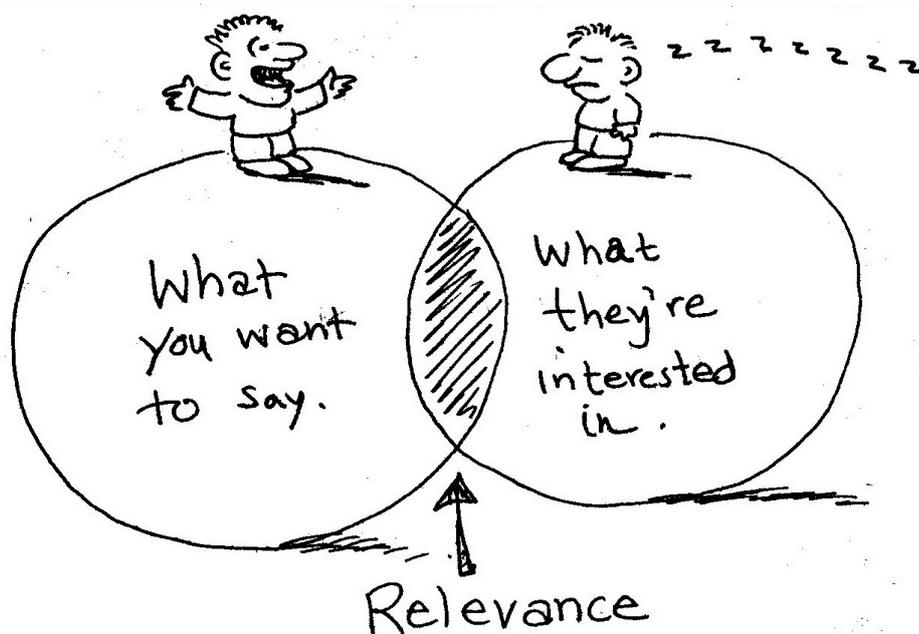
by Nikola Milošević - Saturday, July 13, 2013

<https://inspiratron.org/blog/2013/07/13/personalized-relevance-classifier-of-sentences/>

In this article I would like just to pitch idea about personalized classifier, and I would like to hear your opinion if this approach could be good and what can be problems with it. So what is the problem? I would like to build personalized relevance classifier.

Problem definition

Every user is tracking mentions of some term on internet or social media. Terms are usually brands they want to watch if they are some marketing guys or business owners, or some events, names etc. Since term can be ambiguous, user has opportunity to tell the program that some sentence is irrelevant for him. For example if user enter "Apple", first it will show all mentions of Apple company and fruit called apple. If user don't want to see fruits anymore, he marks couple of sentences that are referencing to fruits and application should learn how to distinguish relevant sentences against irrelevant ones. Also we can call this classifier "**Personalized relevance classifier**".



Idea behind solution

When we look at the problem we see that solution lays in creating classifier for each word that each user enters for following. Of course, when he enters word, there is no data, so no classifier and all the data is shown. When user starts to label sentences that are not relevant to him, behind that action is one machine learning task to build classifier from that data.

Since there is too much data to hold structures for these classifiers in memory we would need to hold these structures in database. When user labels sentences, in database will be created id for that classifier, and under that id sentences, words and word counts will be stored, creating structure for Naive Bayes or some other classifier. User should not need to classify all his current data, just some selected frame. But in that frame (for example one day, or 100 sentences), he needs to label all the sentences. Doing this classifier will be built and it will filter the output data. User will be able to ignore classifier (show all data) or to rebuild classifier.

Suggestions?

I would appreciate all the suggestions about building personalized classifier for users. If anyone has some experience doing this, please leave a comment or contact me. I hope I was concise describing ideas. Also if you see problems in this approach, let me know.

All rights reserved and copyrighted by inspiratron.org and Nikola Milosevic