TSHRDLU Modifications – Project Phase # 3

Jim Evans
Department of Linguistics
University of Texas at Austin
`j.s.evans@utexas.edu`

Jason Mielens
Department of Linguistics
University of Texas at Austin
`jmielens@utexas.edu`

1 Behaviors

We modified the behavior of the automated Twitter user "tshrdlu." First we give a brief description of the four behaviors, with more details in the following subsections. The first behavior is a listening behavior. Every thirty minutes the bot searches for tweets that contain both the word "vaccine" and a link. The bot uses a corpus-trained Naive Bayes classifier to identify the linked webpage as either "pseudoscience" or "science." The bot tweets about classification of the webpage and provides the link to it in the tweet. The bot also has four possible behaviors for responding to tweets directed at it (two of those are the original tshrdlu behaviors):

- Paraphrase: it responds by taking the original tweet and replacing longer words with randomly-selected synonyms of those words.

- Original tshrdlu response: searches Twitter using a query made up of a subset of the original tweet's words and tweets the most relevant result (original tshrdlu behavior),

- Synonym searching response: the same as the previous behavior, except the query is made up of synonyms of the words from the original tweet, rather than the exact words from the tweet.

- Follow: follows users that people ask it to follow (original tshrdlu behavior),

Which response the bot takes when tweeted at depends on the contents of the tweet, and is also probabilistic. As with tshrdlu, there is only one response to a tweet telling the bot to follow someone (assuming it is in the right format – "@jm_anlp follow <some username>"). Namely, the bot obeys the request to follow the indicated user and then tweets that it has done so. If the tweet to the bot is not a follow request, the the bot randomly chooses, with some probability, whether to use its "paraphrase" behavior or the "synonym searching" behavior. Currently the ratio is set as 90% synonym searching and 10% paraphrase.

The synonym-based behaviors make use of a freely available thesaurus from OpenOffice.org, which we turn into a map from input word to synonyms. This thesaurus is included in the bot repo under /src/main/resources/dict/.

1.1 Listening Behavior

The bot has one listening behavior (i.e. it monitors Twitter on its own). Periodically, the bot searches for tweets that contain both the word "vaccine" and a link (signified by the presence of "http"), and finds the most recent such tweet. The bot then classifies the linked page as "science" or "psuedoscience" and tweets about the results. The linked page is retrieved using an HTML renderer (w3m) so that we are hopefully classifying only the actual article, and not a lot of

The Naive Bayes classifier is trained on the corpus created by Beth and Evans (2012), in which documents (blog posts) are tagged as either "science" or "pseudoscience". The test accuracy is approximately 88 percent, and the training accuracy is about 92 percent. The classifier is included in the bot repo under /src/main/resources.

While listening, the bot keeps track of the latest tweet that it has seen. This enables it to only consider new tweets on the next search, and potentially not even say anything at all if it can't find any new relevent tweets. The listening behvaior is implemented as a separate thread that continues until the program is killed.

1.2 Behaviors for Mentions

Two novel mention behaviors (i.e. actions to take when people tweet at the bot) were created for the bot, while retaining tshrdlu's original "fol-

low" behavior, and "backing off" to tshrdlu's original "standard" behavior in some situations. As with tshrdlu, there is only one response to a tweet telling the bot to follow someone (if it is in the right format), and that is that the bot obeys, and then tweets that it has followed the requested person.

The default mention behavior is the synonym search, which requires that a candidate reply contain synonyms for at least two words in the original tweet. This seems to improve the relevance of the returned tweets, at the expense of being unable to find tweets in some instances. If no tweets can be found that statisfy this requirement, the bot backs-off to the original tshrdlu behavior of searching for any words in the original tweet.

2 Output analysis

Here are some examples of interactions with the bot, and associated commentary.

- @jm_anlp You're too quiet today.

- @evans_anlp This rain is so relaxing it has me sleepy again and I just woke up a couple of hours ago its going to be a long day!

In the first example, the bot takes "quiet" and "today" and gets a query of "relaxing AND day," which returned the response tweet as the most recent relevant tweet. This is an excellent example of how the bot finds relevant tweets that would not be found if it only made queries from the exact words in the tweet that it is responding to. Now consider the next example interaction:

- @ jm_anlp Ready for round two?

- @ evans_anlp Getting Ready For Believe Tour, Banging Out The Album And Dancing Round My Room, Can't Wait OMB!!!!!!

In this instance, because the word-to-synonym function can potentially map words to themselves, the bot did search only for words from the tweet: "ready AND round." In this case, it looks just like the original tshrdlu behavior, but it is not (even though our bot does back off to that behavior in some cases, as mentioned in the section on 1.2). The following is an example of the paraphrase behavior:

- @jm_anlp robot of cruelty !!!

- @evans_anlp So in other words: mechanism of malice ! ! !

The mapping of longer words to synonyms is self-explanatory. Finally, here is an example of a spontaneous tweet about a vaccine-related article:

- I read this article, not sure what to think.. http://www.theblaze.com/stories/2013/0

In this case, the classification of "pseudo-science" is probably not right. However, it isn't a particularly scientific- or skeptic-sounding article either. The linked article is mostly reporting on a politician's position about funding for vaccinations. It is for this general reason that the bot hedges the classification results (i.e. "Not sure what to think" vs. "Sounds pretty legit").

3 Future Work

We are content with the varied behaviors of our current bot at this phase of the project, but we have several more ideas to further enhance the bot for the final project. One thing we plan to do is to create a behavior in which the bot looks at the user who has tweeted to it, finds more tweets that the user has posted, and creates an n-gram model of that user's language in order to generate a response tweet.

We also plan to enable the both to save all past interactions, so that it can remember what kinds of things particular users like to talk to the bot about. It could use this information to customize its responses to users who tweet to the bot more than once.

We also would like to expand the classification to health topics other than vaccination (i.e. to use the rest of the Beth and Evans corpus, which covers topics like acupuncture).

References

[Beth and Evans2012] Bradley Beth, and James Evans. 2012. A Theme-based Classifier for Skeptic and Pseudoscience Blog Posts using Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993-1022.

[McCallum2002] Andrew McCallum. 2002. MALLET: a Machine Learning for Language Toolkit. http://www.cs.umass.edu/mccallum/mallet.