

Tone Formality Concerning Hashtag Research

Written by Facebook AI researchers Jason Weston, Sumit Chopra, and Keith Adams, the publication “#TagSpace: Semantic Embeddings from Hashtags” describes a new system of suggesting appropriate hashtags for arbitrary textual posts and compares it to other well-known baselines. This system, dubbed #TagSpace, uses a computational model based on the biological nervous systems of animals, called a convolutional neural network, to train on posts consisting of over five billion words.^{1(par. 1)} With #TagSpace, Facebook researchers were able to rank hashtags based on how well they were related to a specific post and outperform several other systems including one meant to be used in industry for hashtag prediction.^{1(par. 27)} This piece was found on the Facebook page for its research publications which is located at www.facebook.com/publications under the “Artificial Intelligence” section. Hashtag prediction falls under both the computer science fields of Natural Language Processing and Artificial Intelligence because it deals with the analysis of a human language and executes on strings that have not been preprogrammed into the system. This piece exemplifies some of the difficulties of writing a formal paper on one of the many informal topics that populate the web culture aspect of computer science. Informal writing is a skill that often doesn’t come naturally to computer scientists, such as myself, who are used to utilizing technical terms and speaking bluntly; this publication illustrates a real-world example of a moment when formal writing is crossed with an informal topic and how the compatibility issues are dealt with.

The research document, in its entirety, consists of eight sections in total. This computer science research paper possesses a structure that is similar to those found in reports for the natural sciences. There is an abstract that summarizes the findings of the experiments conducted, a brief introduction to the topics concerned, and a vivid description of the experiments and their conclusions. Also, the voice of the paper is entirely passive and although terms such as “WSABIE” or “word2vec” are introduced, the authors assume their audience already has an adequate understanding of what these terms refer to. For a field that differs so much from the natural sciences in how it is applied and how relatively new it is, it’s interesting that

documenting and reporting research in the field of computer science adopts a similar structure to that of publications in chemistry or physics. This, of course, may only seem surprising because computer scientists usually receive very little exposure to research in their field. Computer science is often very product-oriented, focusing more on developing software applications and documenting their specifications than delving deeper into concepts such as finding better algorithms for suggesting hashtags for Facebook posts and running experiments to prove the superiority of these algorithms.

Aside from the structure of this publication, an interesting point of note is the disparity between the seriousness of how the piece is described and the content of the piece as it is generally viewed outside of the professional sphere. Hashtags are most often used by adolescents on social media websites such as Twitter, Instagram, or Facebook; because of this, the use of hashtags is commonly associated with immaturity or casual situations. This research paper maintains its credibility by using advanced diction and approaching the subject of hashtags from a very technical perspective, as evidenced from the first paragraph of the introduction:

Hashtags (single tokens often composed of natural language n-grams or abbreviations, prefixed with the character ‘#’) are ubiquitous on social networking services, particularly in short textual documents (a.k.a. posts). Authors use hashtags to diverse ends, many of which can be seen as labels for classical NLP tasks: disambiguation (chips #futurism vs. chips #junkfood); identification of named entities (#sf49ers); sentiment (#dislike); and topic annotation (#yoga). Hashtag prediction is the task of mapping text to its accompanying hashtags. In this work we propose a novel model for hashtag prediction, and show that this task is also a useful surrogate for learning good representations of text. ^{1(par. 2)}

Although the authors acknowledge the existence of terminology that is often used by those who use hashtags such as “posts” or “#sf49ers”, they distance the research piece itself from the culture associated with the content by putting such terminology in parenthesis. The only instance where this is not the case is when the authors describe hashtags; if they were to be consistent, the first sentence would read: “Single tokens often composed of natural language n-grams or abbreviations, prefixed with the character ‘#’ (hashtags)...”. This lack of inconsistency is likely

to exist only because the term “hashtag” is mentioned in the title so it would unusual to have the body of the paper distance itself from the term. Even disregarding the authors’ choice of diction, the whole topic of hashtags is still approached in a very analytical and professional perspective. Often described as “posters” or “facebookers” in popular culture, users of hashtags and Facebook posts are, instead, described as “authors”. In addition, hashtags are classified based off their ability to resolve ambiguity, refer to entities, express emotion, or emphasize topics, an act that is not often practiced due to the culture that surrounds those who hashtag.

As noted previously, this document refers to an experiment run with #TagSpace to compare it to other programs that offer similar hashtag prediction services. In order to express some of the data, the authors chose to include tables depicting some of the predictions made by #TagSpace to exemplify its abilities. The table reads as follows: ¹(Table 2)

Crazy commute this am, was lucky to even get in to work.	#nyc, #snow, #puremichigan, #snowday, #snowstorm, #tubestrike, #blizzard, #commute, #snowpocalypse, #chiberia
This can't go on anymore, we need marriage equality now!	#samelove, #equalrights, #equality, #equalityforall, #loveislove, #lgbt, #marriageequality, #noh8, #gayrights, #gaymarriage
Kevin spacey what a super hottie :)	#houseofcards, #hoc, #houseofcardsseason2, #season2, #kevinspacey, #frankunderwood, #netflix, #suits, #swoon, #hubbahubba
Went shopping today and found a really good place to get fresh mango.	#mango, #shopping, #heaven, #100happydays, #yummy, #lunch, #retailtherapy, #yum, #cravings, #wholefoods
Went running today -- my feet hurt so much!	#running, #ouch, #pain, #nopainnogain, #nike #marathontraining, #sore, #outofshape, #nikeplus, #runnerproblems
Wow, what a goal that was, just too fast, Mesut Ozil is the best!	#arsenal, #coyg, #ozil, #afc, #arsenalfc #lfc, #ynwa, #mesut, #gunners, #ucl
Working really hard on the paper all last night.	#thestruggle, #smh, #lol, #collegelife, #homework #sad, #wtf, #confused, #stressed, #work
The restaurant was too expensive and the service was slow.	#ripoff, #firstworldproblems, #smh, #fail, #justsaying #restaurant, #badservice, #food, #middleclassproblems, #neveragain
The restaurant had great food and was reasonably priced.	#dinner, #restaurant, #yum, #food, #delicious #stuffed, #goodtimes, #foodporn, #yummy, #winning
He has the longest whiskers, omg so sweet!	#cat, #kitty, #meow, #cats, #catsofinstagram #crazycatlady, #cute, #kitten, #catlady, #adorable

Table 2: #TAGSPACE (256 dim) predictions for some example posts.

The inclusion of this data from the experiment’s findings contrasts with the professional tone the paper was trying to set in the introduction. Aside from the simplicity of the posts being analysed, a close look at some of the predicted hashtags produces some interesting findings one wouldn’t typically see in a formal research document, most notable of these are: #crazycatlady, #foodporn,

and #hubbahubba. Given the amount of effort put into producing a serious tone in the opening paragraph, the authors likely had a reason for allowing a data table with such unprofessional undertones to exist inside their paper. This reason seems to be, again, to improve the credibility of the piece. Whereas the professional tone the authors adopt serves to have their paper taken seriously, the casual terms that populate the values of Table 2 help with the believability of the data. Obviously, the average sentence analyzed using #TagSpace will be quite similar to the ones used in the experiment, and the suggested hashtags reflect the culture of those that most often use hashtags. The stark contrast this table has with the rest of the piece emphasize that the authors hold the accuracy and adequacy of data above their need to sound professional which is very respectable in a research paper.

The computer science research document “#TagSpace: Semantic Embeddings from Hashtags”, says a lot about the field of computer science and how it handles research. Like in any other field, this research paper strives to be passive and professional in the way it presents itself. However, computer science often deals with subjects whose primary user bases are composed of non-professionals so the formality of a piece may need to suffer in order to accommodate for believability. In computer science, it’s apparent that writers need to be able to adapt their styles to be able to deal with both formal and informal subjects. While a piece may be structured in a way that calls for a solemn tone, and even achieves such a tone most of the time, the subject of the piece ultimately dictates whether or not such a piece will remain formal throughout.

Reference

1. Weston, Jason, Sumit Chopra, and Keith Adams. #TagSpace: Semantic Embeddings from Hashtags. Facebook, 2014. [cited 2014 Sept 11]. Available from: <https://www.facebook.com/publications/892028280826743/>