

Spam Detection by Hashtag Relationships

CS 533 Proposal Presentation

Yalim Doğan, M.Enes Karaca, Soner Koç

29 March 2016

Presentation Outline

- Problem Description
- Motivation
- Methodology
- Current Progress
- Expected Results
- References



Problem Description

- The popularity of Twitter attracts spammers for advertising, propaganda, adult content etc.
- The tweets are considered as spam for following reasons:
 - Containing >3 hashtags (can be considered Hashtag Abuse) and some of them are unrelated.
 - Containing links generated by URL shorteners. Considered as a strong indication.
 - Includes words that are considered as spam (advertisement and adult content related)
 - Posted by different(or same) users in same time period (in the same minute at shortest) which are not by Retweeting.

Problem Description (Examples)



A Tweet Containing Spam Related Words Together with shortened URL



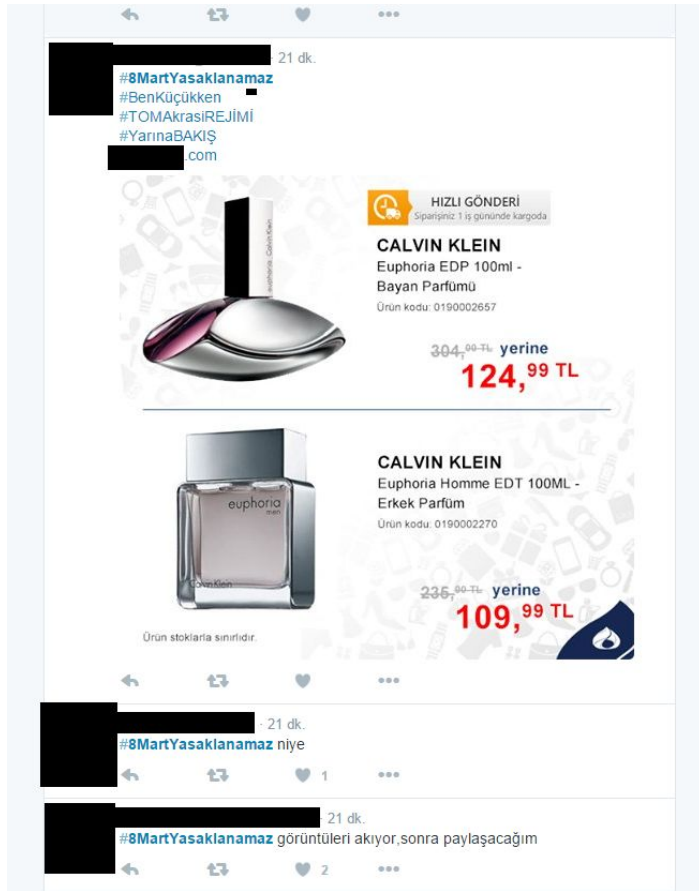
Advertising Spam.

- 46 dk.
Kıbrıs Vakıflar İdaresi'nden 8 Mart Dünya Kadınlar Günü mesajı
[com/haber/42602/ki...](https://www.kypr.gov.tr/haber/42602/ki...)
#8MartYasaklanamaz #BenKüçükken
Özeti görüntüle
- 46 dk.
Tatar: "Rum tarafının akli, gerçeklere dayalı bir anlaşmada değil"
[com/haber/42601/ta...](https://www.kypr.gov.tr/haber/42601/ta...)
#8MartYasaklanamaz #BenKüçükken
Özeti görüntüle
- 46 dk.
#BenKüçükken büyüyen çocuklardandım. Şimdi gençliğimde ihtiyarladım....

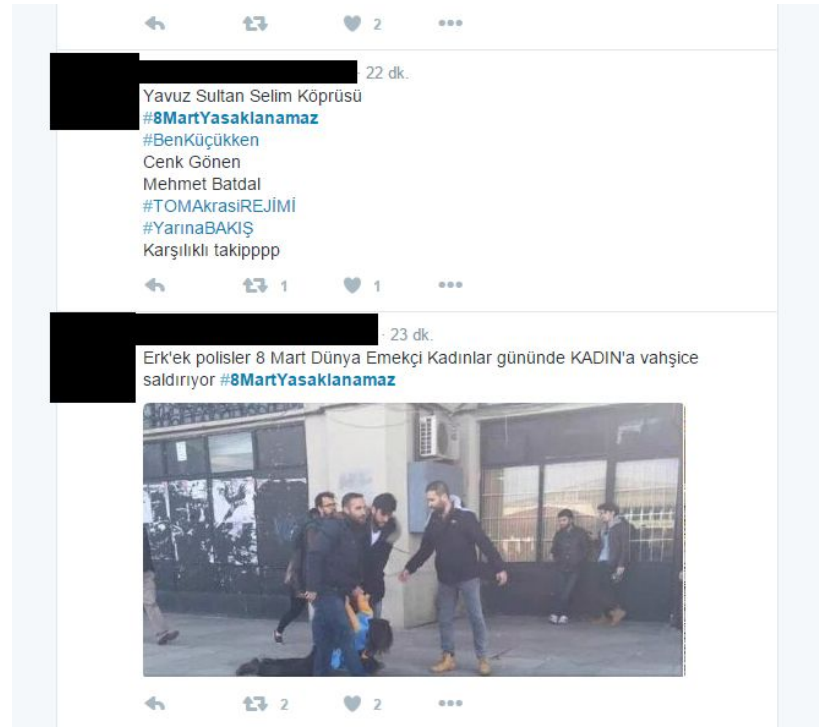
News Spam

- 53 dk.
#BenKüçükken hep Toni Montana olmak isterdim, öğretmenlerim desteklerdi
hala çözemedim neden? Olamadım ve destekledi.
1 2
- 53 dk.
#8MartYasaklanamaz
#BenKüçükken
#TOMAkraSiREJİMİ
#YarınaBAKİŞ
youtube.com/watch?v=6UlyhG...
- İLK ŞARKIM SİZCE NASIL OLMUŞ ? TAMAMEN
ŞÖZLERİ BEN YAZDIM... her türlü yoruma ve eleştiriye
açığım dinlediğiniz için teşekkürler...
youtube.com

Hashtag Abuse



Commercial through Hashtag Abuse



An Attempt for gaining more Followers. An Example of Hashtag Abuse.

Motivation

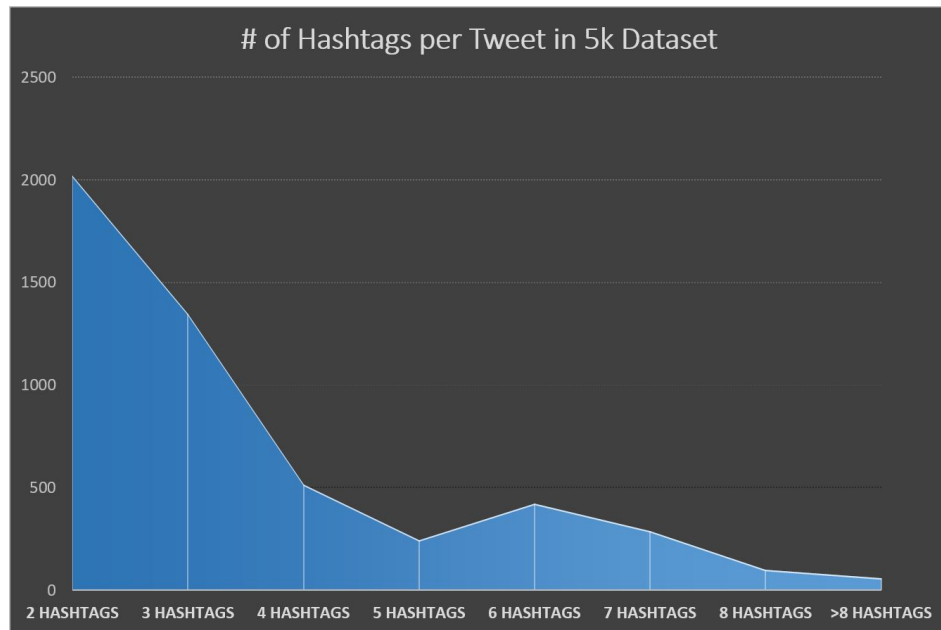
- There are several methods proposed in order to fight spamming which consider the common elements of spam tweets.
 - URL's posted in tweets [1]
 - Number of Hashtags (**BUT** doesn't consider the relation between them)
 - If the tweets are duplicate (Not RT)
 - Usage of spam words [2]
- We propose a system that checks the relationship between hashtags in tweets, in addition to the above parameters.
- In this way, we are aiming to improve the accuracy of spam detection on Twitter.

Methodology

- Hashtag Abuse Detection
 - Considering tweets that contain => 2 hashtags.
 - Determining the relation between hashtags by
 - Extracting keywords from each hashtag
 - Categorizing hashtags using keywords (Can have multiple, each weighted)
 - Checking if they fall into similar categories (Clustering)
- Link Status
 - Checking the presence of shortened URL's
- Spam and Non-Spam Words
 - Searching for words in tweets that are spam context
- Duplicate Tweets
 - Checking the latest 10 tweets of each user, to check for any duplicate tweet [2]

Current Progress

- We have gathered 5k tweets that contain more than 1 hashtag.
- The hashtags are written in CamelCase which makes it much more easy to extract the keywords in popular hashtags.
 - CamelCase: WeLikeReading, IBelieveIn etc. NOT currentlylistening
- We consider automatically declaring tweets that have more than 7 (subject to change) hashtags as spam.
- Hashtag categories can be as:
 - *Culture, Art, History, Sport, Science, Health, Philosophy, Religion* etc.
 - There will be subcategories



Hashtag Analysis of Tweets

(1start)UniteBlue4Bern(+)RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:14:51(+)+2(end)

(1start)N8ohv2fLBC6L3PQ(+)+RT @WNC4Bernie: Don't give in to fear! #DemocracyIn5Words

#GOTVforBernie #FeelTheBern #VoteForTheFuture #CaucusForBernie #VoteBernie [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:13:22(+)+3(end)

(1start)paulstewart35(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:13:20(+)+4(end)

(1start)michaelwoodlock(+)+RT @Trump_ego2016: We need you Donald Trump.

#DemocracyIn5Words #AlwaysTrump #TrumpTrain #MakeAmericaGreatAgain [https://t.co/aypa3PKWMJ\(+\)](https://t.co/aypa3PKWMJ(+))27.03.2016 22:11:35(+)+5(end)

(1start)Story27368218(+)+RT @ExtraSolarGroup: Stand with Bernie and VOTE! #DemocracyIn5Words #BirdieSanders #FeelTheBern [https://t.co/CJEXcyGhXi\(+\)](https://t.co/CJEXcyGhXi(+))27.03.2016 22:11:35(+)+6(end)

(1start)rdsathene(+)+RT @DrJillStein: People, Planet, Peace over profit. #DemocracyIn5Words #ImWithHer #DemocracyInEmojis #ItsInOurHands #ImWithJill(+)+27.03.2016 22:11:35(+)+7(end)

(1start)DaddyGuess(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:10:10(+)+8(end)

(1start)___Roar___(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:08:19(+)+9(end)

(1start)StoneSculptorJN(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 22:00:01(+)+10(end)

(1start)AttilioArmeni(+)+Hey #SuperDelegates "People's will is your mandate" #DemocracyIn5Words(+)+27.03.2016 21:59:15(+)+11(end)

(1start)Chewbacca Sound(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 21:57:38(+)+12(end)

(1start)i314159265(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 21:56:20(+)+13(end)

(1start)elm3c(+)+RT @SadlyCatless: Put a bird on it! #DemocracyIn5Words #BirdieSanders [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 21:55:53(+)+14(end)

(1start)teamgovforever(+)+RT @TheGovernorTWD_: #DemocracyIn5Words. This isn't a democracy anymore! (It's better when I say it. I mean it!) #TheWalkingDead [https://t.co/7fD9kwrk4J\(+\)](https://t.co/7fD9kwrk4J(+))27.03.2016 21:55:53(+)+15(end)

(1start)SadlyCatless(+)+@NerdyPodcast @xanderberkeley #FeelTheBerk could start a #2 worldwide trend, like my tag #DemocracyIn5Words did yesterday. #humblebrag(+)+27.03.2016 21:55:53(+)+16(end)

(1start)BagwellBuc(+)+RT @TEN_GOP: Time to build the WALL

#DemocracyIn5Words

#Trump #Trump2016 [https://t.co/z4HtrPhXHe\(+\)](https://t.co/z4HtrPhXHe(+))27.03.2016 21:30:18(+)+17(end)

(1start)C Lush87(+)+RT @CityCadillac: us It all about #American 🇺🇸 <https://t.co/73XpnFcZgt> us #DemocracyIn5Words [https://t.co/LHzUQxvETy\(+\)](https://t.co/LHzUQxvETy(+))27.03.2016 21:29:19(+)+18(end)

(1start)kamba722(+)+RT @ExtraSolarGroup: Stand with Bernie and VOTE! #DemocracyIn5Words #BirdieSanders #FeelTheBern [https://t.co/CJEXcyGhXi\(+\)](https://t.co/CJEXcyGhXi(+))27.03.2016 21:27:04(+)+19(end)

(1start)dona1d_lips(+)+RT @DonnyTrumpy: Building a stupid fucking wall. #DemocracyIn5Words #MakeAmericaFuckingStupid(+)+27.03.2016 21:26:22(+)+20(end)

(2start)dona1d_lips(+)+RT @DonnyTrumpy: Building a stupid fucking wall. #DemocracyIn5Words #MakeAmericaFuckingStupid(+)+27.03.2016 21:26:22(+)+21(end)

(2start)MLCzone(+)+RT @BRios82: Bernie Sanders supports Hillary Clinton. #DemocracyIn5Words #TrustHillary #ImWithHer [https://t.co/mKJu3Tmhy8\(+\)](https://t.co/mKJu3Tmhy8(+))27.03.2016 21:20:03(+)+22(end)

(1start)TonyEquality(+)+RT @BRios82: Bernie Sanders supports Hillary Clinton. #DemocracyIn5Words #TrustHillary #ImWithHer [https://t.co/mKJu3Tmhy8\(+\)](https://t.co/mKJu3Tmhy8(+))27.03.2016 21:19:05(+)+23(end)

(2start)redhead67(+)+RT @BRios82: Bernie Sanders supports Hillary Clinton. #DemocracyIn5Words #TrustHillary #ImWithHer [https://t.co/mKJu3Tmhy8\(+\)](https://t.co/mKJu3Tmhy8(+))27.03.2016 21:07:41(+)+24(end)

(2start)shibinn(+)+RT @ProjectBernie16: We believe in our future. #BirdieSanders #WAcaucus #DemocracyIn5Words [https://t.co/jsIPTQ0VHU\(+\)](https://t.co/jsIPTQ0VHU(+))27.03.2016 21:07:10(+)+25(end)

(2start)FercarolAnne(+)+RT @Lucy_Duncan: "Nobody's free until everybody's free." #DemocracyIn5Words #FannieLouHamer(+)+27.03.2016 21:06:05(+)+26(end)

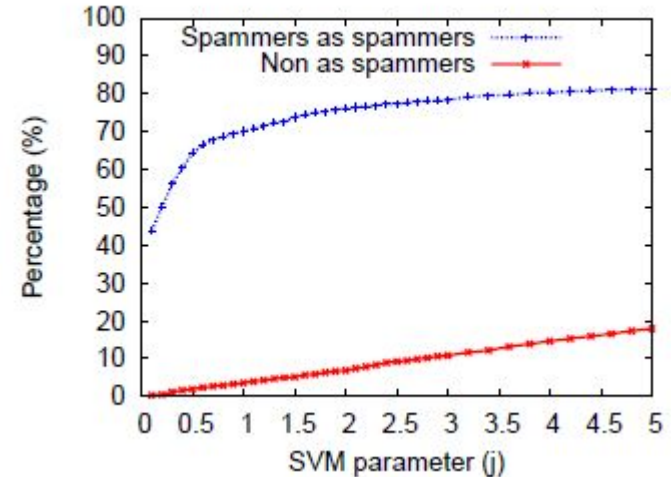
(2start)Sarahbear80(+)+RT @ProjectBernie16: We believe in our future. #BirdieSanders #WAcaucus #DemocracyIn5Words [https://t.co/jsIPTQ0VHU\(+\)](https://t.co/jsIPTQ0VHU(+))27.03.2016 21:05:57(+)+27(end)

(2start)rholtzman(+)+RT @BRios82: Bernie Sanders supports Hillary Clinton. #DemocracyIn5Words #TrustHillary #ImWithHer [https://t.co/mKJu3Tmhy8\(+\)](https://t.co/mKJu3Tmhy8(+))27.03.2016 21:04:41(+)+28(end)

(2start)rholtzman(+)+RT @BRios82: Bernie Sanders supports Hillary Clinton. #DemocracyIn5Words #TrustHillary #ImWithHer [https://t.co/mKJu3Tmhy8\(+\)](https://t.co/mKJu3Tmhy8(+))27.03.2016 21:04:41(+)+29(end)

Expected Results

- We expect to have high amount of precision; at least close to performance of some spam detection systems proposed for Twitter.
- We will use Precision, Recall and F-Measure to determine the performance of our system.
- Aim to have least amount of false positives as possible.
- We might consider getting duplicate tweets together beforehand, to improve the speed of the process.
- The process will be repeated frequently to detect spam tweets in an instant.



		Predicted	
		Spammer	Non-spammers
True	Spammer	70.1%	29.9%
	Non-spammer	3.6%	96.4%

Some Results from [1]

References

- [1] F. Benevenuto, G. Magno, T. Rodrigues, V. Almeida, “Detecting Spammers on Twitter”, Electronic messaging, Anti-Abuse and Spam Conference (CEAS), 2010.
- [2] A Collection of 14 Million Tweets for HashtagOriented Spam Research. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15).
- [3] Wang, A.H.: Detecting spam bots in online social networking sites: A machine learning approach. In: Foresti, S., Jajodia, S. (eds.) Data And Applications Security and Privacy XXIV.LNCS, vol. 6166, pp. 335-342. Springer, Heidelberg (2010)

#Thank #You #For #Listening