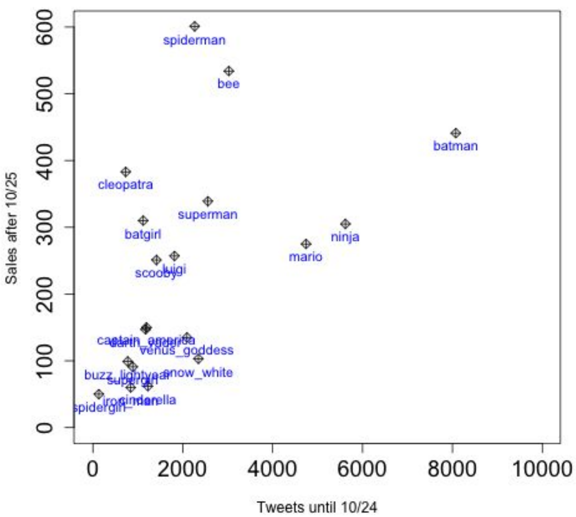
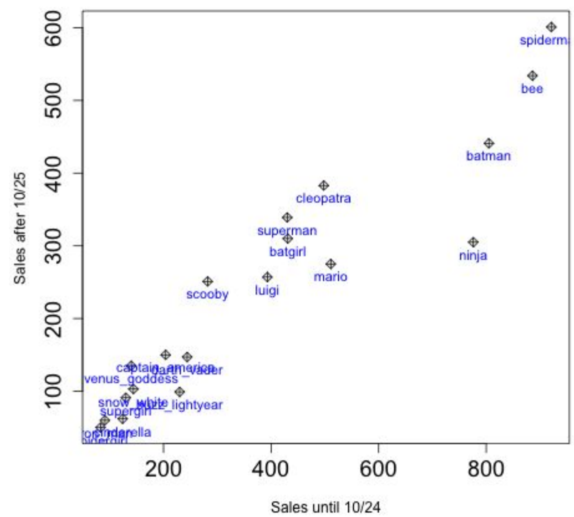
CS 341: Halloween Costume Predictions using Twitter

Anirudh Venkatesh, Onkar Dalal, Praveen Bommannavar

Introduction

For this project we had access to Walmart's sales data from the Northern Illinois region for October 7, 2011 - November 7, 2011 as well as 3 billion tweets from a similar time frame. We began this project with an aim to predict the sales of Halloween costumes for Walmart using Twitter as a signal. The analysis we conducted showed that one could indeed correlate the sales of Halloween costumes with activity on Twitter mentioning the costume name. The left plot shows a scatterplot of costume sales after October 25 versus tweets mentioning the costume until October 24.





The right plot shows the same response plotted against costume sales until October 24, which appears to be a much stronger predictor. Nonetheless, this analysis showed that Twitter contains a signal for the popularity of items, one which could be made stronger with appropriate tuning for better precision.

Problem Description

In light of this, we turned our attention to the problem of detecting which costumes were actually worn on Halloween. Knowing this would allow us to answer the question: **which costumes should Walmart have sold on Halloween?**

To explore this question, we focused on the time frame immediately before Halloween, which fell on a Monday this year; that is, we analyzed approximately 100 million English language tweets from October 28 - October 31, the days immediately preceding Halloween.

Related Work

Many recent attempts have been made to form predictions based on activity on Twitter. Asur and Huberman show that the success of newly released movies can be forecasted using sentiment analysis on tweets relevant to movie titles [1], while Bollen and Mao use moods contained in tweets to predict the stock market [2]. More recently, Globalpoint Research did a sentiment analysis of tweets related to the New Hampshire primary to accurately predict the outcome [4].

The methods which we shall employ in this work are closely related to those presented in [3], where Agichtein and Gravano describe a method which they term 'Snowball,' for generating a table of relations between keywords and attributes of interest. This method begins with seed keywords and attributes, and finds relevant contexts around which these appear. Using these contexts, additional items are added to the table.

Methodology:

Our methodology for mining the Twitter data for costumes which were worn on Halloween follows the approach taken in [3], with modifications to account for the fact that we are simply searching for relevant costumes rather than any attribute in particular. That is, while the Snowball method looks for structured relations of the form

Organization => Location
Microsoft => Redmond
Exxon => Irving
IBM => Armonk

we are simply interested in popular costumes. This introduces additional challenges, which we shall describe in detail. Also, we incorporated helper keywords to guide our search, an element not present in [3]. Specifically, whereas the Organization=> Location example looked through all documents mentioning some seed organizations with the mention of a seed location, we guided the search for halloween costume by first filtering tweets on reliable keyword surrogates.

The procedure we used is outlined below:

Step 1: Generate keywords that are relevant to our task

Although a number of alternatives exist for this step, we looked for words co-occurring with the terms {halloween, costume, dress up} by looking for large values of

`P[word | {halloween} or {costume} or {dress up}] - P[word].`

Large values correspond to a word being strongly associated with the terms "halloween", "costume" or "dress up". Other alternatives are dividing by `P[word]` rather than subtracting, or multiplying by inverse document frequency (IDF).

Step 2: Context generation

Part a)

In this part, tweets are inspected for the presence of costume seeds (such as batman or zombie) and surrounding so-called *contexts* are given points for each particular costume depending on the number of keywords (from Step 1) that appear in the tweet.

Here, a context is defined as a tuple `(l_context, r_context)` where `l_context` corresponds to a sequence of words appearing before a costume name, and `r_context` corresponds to a sequence of words appearing after a costume name.

When this part is complete, we can organize the point totals into a matrix where rows correspond to different contexts and columns correspond to different costumes. The point total for context i and costume j is denoted by `value_ij` in the table below.

	costume_1	costume_2	costume_3	…
(l_context, r_context)_1	value_11	values_12	value_13	…
(l_context, r_context)_2	value_21	values_22	value_23	…
(l_context, r_context)_3	value_31	values_32	value_33	…

Part b)

Using this matrix, we score each of these rows, making sure to account for diversity, that is, we give a context greater credibility if many costumes are expressed using that context. The score is also increased if the values are high.

The specific mechanism by which scores are computed is as follows:

- Each costume j can give a context i at most two points.
  - One point is given for the presence of context i (indicator variable) in any tweet about costume j.
  - The fraction `log(1 + value_ij)/log(1 + max_i(costume_j))` points are given to emphasize the value. The quantity `max_i(costume_j)` is the maximum point total of costume j over all contexts.

From the computed scores for each context, we produce a short-list of costumes that have at least a threshold score. This is to ensure that all contexts we choose have enough frequency with any given costume, but also have a large enough diversity (in other words, the context is expressed along with a sufficient number of different seed costumes). After accounting for the score threshold, we have a shortlist of roughly 150 contexts.

Step 3: Costume prediction

The final phase of our procedure involves using the generated keywords from Step 1 and the generated contexts from Step 2, along with their associated scores, to output a more extensive set of valid Halloween costumes than is given by the seed set of costumes. To do this, we examine tweets in the time period from October 28 - October 31 and generate candidate costume names based on proximity to all input contexts. For example, for the context (“dress up as” , “for Halloween”), the candidate costumes produced would be the sequence of words that occurs between “dress up as” and “for Halloween” over all the input tweets. In the case that a context is completely on the right side or left side (e.g. (“want to dress up as”, “”)), we generate candidate costumes of up to 3 words in length preceding or following the context, respectively.

For each generated costume j, we compute and output an aggregate score `total` as follows:

`total = 0`

- For every tweet in which costume j is flanked by any of the input contexts:
    - For every occurrence of an input keyword, add the associated score of that keyword (as computed in Step 1) to `total`
    - For every occurrence of an input context, add the associated score of that context (as computed in Step 2) to `total`

To prevent skewing of computed total score by retweeted tweets, we omit tweets that have “RT @” in them from consideration. While this reduces the error caused by the retweets, we were unable to eliminate the retweets where the tweet text was copied with an additional remarks. Using the computed aggregate score `total` for each costume as a measure of its relevance as a Halloween costume, we can then rank the generated costumes by aggregate score to produce a list of the costumes most likely worn during Halloween in 2011.

Results:

We now give the inputs and outputs associated with each step of the procedure outlined above. In our analysis, we began with the seed costume set

`C = {banana, batman, catwoman, devil, lady gaga, nicki minaj, nurse, pirate, spiderman, superman,vampire, zombie}`

<u>Keywords:</u>

We obtained the following list of keywords occurring with `{halloween, costume, dress up}`. The bold words have been manually selected as relevant keywords for which we shall give extra points for tweets containing them.

['**halloween**', '**costume**', 'a', 'for', 'as', 'happy', 'costumes', 'up', 'my', '**party**', 'dressed', '**dress**', '**tonight**', 'be', 'to', 'this', 'dressing', 'and', 'going', 'the', 'in', 'wear', 'or', 'year', 'on', 'happyhalloween', 'is', 'candy', 'have', 'slut', 'what', '**trick**', 'everyone', 'wearing', 'kids', 'weekend', 'gonna', 'of', 'out', 'your', 'girls', "it's", 'tomorrow', 'best', '**night**', 'scary', 'ur', 'today', 'its', 'all', 'are', 'i', 'slutty', 'was', 'just', 'last', 'at', 'some', 'think', 'school', 'cute', "i'm", 'little', 'fun', 'see', 'go', 'their', 'not', 'treating', 'so', 'should', '**treat**', 'wait', 'sexy', 'ready', 'look', 'macmiller', 'dont', "don't", 'monday', 'contest', 'pumpkin', 'our', 'excited', 'seen', 'lord_voldemort7', 'being', 'parties', 'need', 'idea', 'if', 'dear', 'these', 'people', 'pics', 'work', 'ideas', 'celebrate', 'any', 'cat', 'day', 'ever', 'mycostume', 'kid', 'snow', 'lil', 'got', 'saw', 'zombie', 'minute', 'sluts', 'old', 'awesome', 'an', 'girl', 'face', 'get', 'christmas', 'with', 'outfit', 'pic', 'part', 'house', 'most', 'put', 'enough', '**mask**', 'years', 'witch', 'door', 'even', 'lady', 'yet']

<u>Contexts:</u>

A sample of the highest-scoring contexts with associated scores are given:

to dress up as|	14.559763	be a|for halloween	12.851941
a|costume for halloween	11.541077	dress up as a|	10.941845
to be|for halloween	10.555063	to be a|for	10.192246

<table>
<tr><td>going to be a|</td><td>10.137222</td><td>wanna be|for halloween</td><td>10.133323</td></tr>
<tr><td>gonna be|for halloween</td><td>10.026384</td><td>to be a|for halloween</td><td>9.980893</td></tr>
<tr><td>dress up as|for</td><td>9.882041</td><td>being a|for halloween</td><td>9.650323</td></tr>
<tr><td>i wanna be a|</td><td>9.621495</td><td>up as a|for</td><td>9.527579</td></tr>
<tr><td>as a|for halloween</td><td>9.451828</td><td>should be|for halloween</td><td>9.416168</td></tr>
<tr><td>gonna be a|for</td><td>9.408806</td><td>gonna be a|for halloween</td><td>9.301002</td></tr>
<tr><td>going to be a|for</td><td>9.271627</td><td>i'm gonna be a|</td><td>9.239245</td></tr>
<tr><td>dress up as a|for</td><td>9.227627</td><td>i'm going to be a|</td><td>9.107306</td></tr>
<tr><td>going to be a|for halloween</td><td>9.10595</td><td></td><td></td></tr>
</table>

<u>Costumes:</u>

In the next step, we use these contexts to generate a list of costumes which occur in the specific context order from the tweets belonging the four days prior to Halloween. The results generated produced costume names with **precision 78.3%** and the same process on tweets from fourteen days before Halloween generated costumes with **precision 75.7%**. These precision values are the percentages of actual costumes among the candidate costumes generated, values were evaluated by manual inspection of results. The numbers can be explained by the adjectives picked and the longer phrases picked from the contexts which have only left or right phrase. Also, the difference between the precisions can be explained by the increasing relevance of the tweets (with the seed costumes and the contexts) to Halloween as the day approaches.

The following is the list of 50 highest-scoring costumes generated from using all tweets from the four day window preceding Halloween (October 28 - October 31) as input. The found costumes are displayed along with associated score:

cat	3255.7	zombie	1333.5	nerd	1273.65
witch	1065.75	vampire	946.4	pirate	821.45
pumpkin	789.6	ghost	698.95	hipster	676.55
clown	579.95	bunny	578.9	devil	561.05
princess	557.2	stripper	493.15	cop	488.95
carrot	460.25	spoon	445.2	banana	437.5
iphone 5	426.65	hippie	424.55	nurse	414.75
kitty	407.4	nun	405.6	ninja	401.8
angel	376.6	fairy	370.65	eskimo	362.95
sexy pan am stewardess	323.4	cowgirl	315.35	indian	305.9
playboy bunny	303.8	cow	302.4		
football player	288.75	normal person	284.9	mexican	281.4
smurf	269.5	nicki minaj	267.05	hobo	254.45
prostitute	253.4	cowboy	234.85	cheerleader	233.8
leprechaun	225.75	angry bird	208.25	bumble bee	206.15
tiger	204.75	lion	203.0	muggle	200.9
school girl	199.15	sailor	198.1	penguin	193.2

To examine whether we can predict consumption of the popular (high-scoring) costumes that we display in the previous table from tweets earlier in the month, we run Step 3 of our algorithm over all texts in the period of 14 days before Halloween, to produce the following results for top 50 high-scoring costumes:

cat	1467.55	zombie	930.3	pirate	805.7
vampire	787.15	cop	684.2	princess	540.75
pumpkin	502.25	nerd	493.5	bunny	485.45
witch	471.45	playboy bunny	427.7	nun	425.95
banana	424.2	angry bird	419.65	clown	415.1
ghost	402.85	ninja	388.15	stripper	380.1

indian	370.3	nurse	368.2	slutty xbox	325.5
ups delivery man	324.1	smurf	322.3	fairy	311.85
sailor	307.3	kitty	287.35	angel	284.2
ninja turtle	267.4	hipster	266.3	devil	263.2
unicorn	261.8	hooters girl	255.5	power ranger	252.7
nicki minaj	251.3	hippie	241.85	lion	241.15
crayon	232.4	cowgirl	229.95	cow	209.65
chocolate m&m	209.3	brick	200.55	carrot	198.8
cheerleader	198.45	snooki	197.05	amy winehouse	194.95
patriot	194.6	lady gaga	190.05	mermaid	187.95
cowboy	181.65	football player	173.95		

Comparing the two sets, 35 out of 50 costumes (Jaccard similarity of 0.54) predicted from the fourteen-days tweet data were also obtained in the four-days tweet data. This large overlap gives validity to the predictions obtained over the fourteen-day time period. It shows that contexts generated in Step 2 are clearly linked with Halloween costumes since potential noise from the entire 14-day period does not skew the costume generation results too far from those of the 4 day window immediately prior to Halloween. Thus, we can see that the procedure we use allows for utilizing tweet data from an earlier time period (in this case fourteen days) to actually predict which Halloween costumes will be popular to a reasonable degree.

Validation:

We compare the costume predictions with the results of the survey conducted by the National Retail Federation http://www.nrf.com/modules.php?name=News&op=viewlive&sp_id=1200 about popular Halloween costumes across the United States in 2011 (see Appendix for details). For the purposes of validation, we believe this source is a valid “ground truth” as to which costumes people actually wore for Halloween 2011.

By comparing the top costume results that our algorithm predicted with the top 20 most popular Halloween costumes among adults from this survey, we see that 15 out of the top 20 costumes were produced by our algorithm as one of the top 50 costumes worn during Halloween in 2011. Thus, we can see that not only did our algorithm produce mostly feasible costume results (as evidenced by 78% precision), but the costumes it generated as being popularly worn in the 4-day time window before Halloween were correct to a reasonable degree, based on the results of this survey.

Conclusions:

In this project, we ran one iteration of the 3 Step method outlined above in order to generate a larger set of Halloween costumes. By running more iterations, using the output set of costumes from the previous iteration as the seed set of the next iteration, we can in-turn generate more contexts, and thus, more candidate Halloween costumes. However, with the precision values that we received above after one iteration (roughly 78%) , running further iterations of this procedure would require a quality scoring function else we run into the risk of introducing a large number false positives into the costume predictions. Since we prune the contexts generated to account for diversity, the set of ‘all contexts’ is not well-defined and hence, there is no easy way to compute the recall. However, a possible proxy for the set of ‘all contexts’ would be to generate an extended list of contexts based on the newly generated costumes (again accounting for the diversity).

The results of this project could have great utility for Walmart. From our analysis, and by looking at the set of costumes sold by Wamart (from the Walmart data set), we find that several of the highest-scoring costumes that we observe to be the most likely to be worn during Halloween 2011 were not carried in Walmart stores at that time. The following are examples of such costumes:

nerd	1273.65	hipster	676.55	carrot	460.25
spoon	445.2	iphone 5	426.65	nun	405.6
eskimo	362.95	nicki minaj	267.05	angry bird	208.25

These costumes can be used to answer the question raised in the problem statement about **which costumes Walmart should have sold**. As these costumes are observed in our analysis to be likely candidates for popular Halloween costumes in 2011, it probably would have been profitable for Walmart to include several of these in its 2011 Halloween costume offerings.

In addition, the implication we notice that tweets from an earlier time window can be used to predict costumes that will be worn during the Halloween time window to a reasonable degree could be extended and applied to a variety of other similar problems. Such applications of using social data for prediction could potentially give Walmart valuable information about the types of products that are currently trending that their stores currently do not carry.

References:

[1] *Predicting the Future with Social Media,* Sitaram Asur and Bernardo Huberman, in Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

[2] *Twitter mood predicts the stock market, Journal of Computational Science*, 2(1), March 2011, Pages 1-8

[3] *Snowball: Extracting Relations from Large Plain-Text Collections*, Eugene Agichtein and Luis Gravano, in Proceedings of the 5th ACM International Conference on Digital Libraries (DL), 2000

[4] http://mashable.com/2012/01/10/twitter-primary/

Acknowledgement:

We are extremely grateful to our mentor, Prof. Anand Rajaraman, for his constant guidance in this project and for taking the time out of his busy schedule to meet with us each week. We also would like to thank Prof. Jeff Ullman, Prof. Andreas Weigend, and Prof. Jure Leskovec for their support.

Appendix:

http://www.nrf.com/modules.php?name=News&op=viewlive&sp_id=1200

Ranks for Adults:

1 Witch	13.4%	7,305,503
2 Pirate	3.9%	2,107,817
3 Vampire	3.7%	2,012,007

4 Zombie	3.0%	1,628,768
5 Batman Character	2.2%	1,221,576
6 Cat	2.2%	1,197,623
7 Wench/Tart/Vixen	2.1%	1,125,766
8 Ghost	1.5%	838,336
9 Nurse	1.4%	766,479
10 Scary Costume	1.2%	670,669

Ranks for Kids:

1 Princess	11.0%	5,134,868
2 Witch	4.4%	2,058,908
3 Spider-Man	3.1%	1,438,755
4 Pirate	2.8%	1,314,725
5 Pumpkin	2.8%	1,289,918
6 Fairy	2.6%	1,215,500
7 Action/Super Hero	2.5%	1,190,694
T8 Batman	2.4%	1,141,082
T8 Vampire	2.4%	1,141,082
T9 Disney Princess	2.2%	1,041,857