Do Cashtags Matter?: An Analysis of the Cashtag Attribute

Vincent Chee

Aaron Cass, Advisor

Introduction

- **Do cashtag Tweets matter?**
  - They are more relevant to the subject matter
  - But they may not represent consumer sentiment
- **A good source of public sentiment data is social media, namely, Twitter**
- **The idea is positive/negative public sentiment will lead to an increase/decrease in stock price**
- **Using tweet data, we can get sentiment data along with stock market data to build a classifier**
- **There are tweets with and without cashtags:**
  - Cashtags are stock ticker symbols, e.g. $INTC

Figure 1. Example of what I'm trying to model.

Research Focus

- **I would like to investigate whether including a cashtag attribute in a model would improve prediction**
  - Cashtag? attribute indicates whether the tweet contains a stock ticker symbol or not
  - If I include cashtag? how will it impact the model?
- **To answer this question:**
  - I will build two models using IBk
    - A model with the cashtag? attribute and a model without it
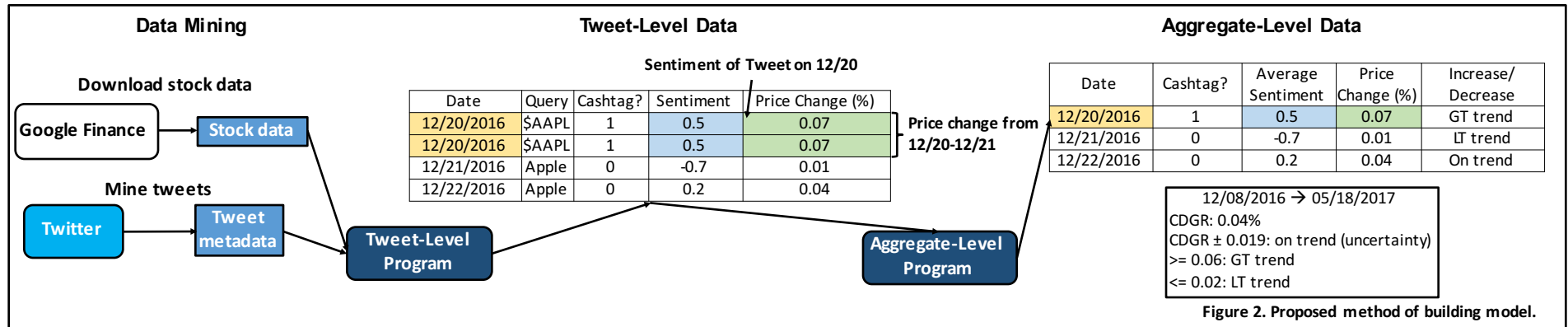 - Test for significance of the model with the cashtag attribute

---

Data Mining

Download stock data

Google Finance → Stock data

Mine tweets

Twitter → Tweet metadata → Tweet-Level Program

Tweet-Level Data

Sentiment of Tweet on 12/20

Date	Query	Cashtag?	Sentiment	Price Change (%)
12/20/2016	$AAPL	1	0.5	0.07
12/20/2016	$AAPL	1	0.5	0.07
12/21/2016	Apple	0	-0.7	0.01
12/22/2016	Apple	0	0.2	0.04

Price change from 12/20-12/21

→ Aggregate-Level Program

Aggregate-Level Data

Date	Cashtag?	Average Sentiment	Price Change (%)	Increase/ Decrease
12/20/2016	1	0.5	0.07	GT trend
12/21/2016	0	-0.7	0.01	LT trend
12/22/2016	0	0.2	0.04	On trend

12/08/2016 → 05/18/2017
CDGR: 0.04%
CDGR ± 0.019: on trend (uncertainty)
>= 0.06: GT trend
<= 0.02: LT trend

Figure 2. Proposed method of building model.

Results

- **The models were built using ~2 million tweets dated between Dec 8, 2016 – May 18, 2017 with the daily DJIA opening price data over the same period**
- **Model w/o cashtag attribute slightly outperformed model w/ cashtag attribute**
- **The corrected paired t-test indicates that there is a slight difference between the models, but it wasn't statistically significant**

Statistic	w/o Cashtag (1)	w/ Cashtag (2)
Total Correctly Classified	66.9 (v/ /*)	65.2 (0/1/0)
GT Trend Correctly Classified	78.4	76.6
LT Trend Correctly Classified	55.5	53.6

Key:
v – statistically significantly better vs model (1)
f – uncertain about statistcal significance
* – statistically signicantly worse than model (1)

Figure 3. Results from Weka.

Future Work

- **Removing or marking Twitterbot Tweets**
- **Explore other variables in existing financial models to add to the model**
- **Including the number of followers of the Tweet user in the model (more followers may be more influential)**
- **Include tweets pulled from the Twitter Firehose (streaming)**