Twitter Opinion Topic Model: Extracting Product Opinions from Tweets by Leveraging Hashtags and Sentiment Lexicon

Kar Wai Lim, Wray Buntine
3 October 2014
Outline

• Motivation
• Background and Previous Work
• Twitter Opinion Topic Model
• Experiments
Online Reviews

• Abundant
Online Reviews

• But...
  – Fake reviews are everywhere too.
    • Consumers trust reviews more than ads.
    • 1 star increase in Yelp = 5-9% increase in revenue *
    • 1 bad review => 30 customers loss
  – Cheaper compared to advertising.
  – Estimated about 30% of online reviews are fake.

* from Luca (2011)
Alternatives

• Opinions from Social Media
  – usually meant for friends and family.
  – Hence are usually truthful opinions.
  – People more willing to post social update than write a proper review.
  – Less targeted by malicious companies due to lower reach.
Problems

• Social updates tend to be short with little details.
• Improper language makes it harder to analyse with existing NLP approach.
• Sarcasm:
Outline

• Motivation
• Background and Previous Work
• Twitter Opinion Topic Model
• Experiments
Background

• Aspect-based opinion mining

Example:

• Target: Sony, Microsoft, Nintendo...
• Aspect: Game consoles
  • PS4 – impressive...
  • XboxOne – cool...
  • Gameboy – retro...
Background

- LDA-based models considered state-of-the-art for aspect-based opinion mining (Moghaddam, 2012).
- LDA is the simplest Bayesian topic model.
- Topic Model
  - assigns a categorical label (topic) to each word in each document.
  - Allow us to analyse the words of each topic,
  - and also topic composition of each document.
LDA

- LDA models the words in each document.
  - In our case a document is a tweet.
LDA

• LDA assigns a topic label to each word.
  – Topic label is latent (unobserved).
LDA

• Words and topics are generated from probability distributions.
  – Theta : document-topic distributions

  θ_{m1}: 60% topic1, 40% topic2

  θ_{m2}: 30% topic1, 70% topic2
LDA

- Words and topics are generated from probability distributions.
  - Theta : document-topic distributions
  - Psi : topic-word distributions

Psi1 : 10% “awesome”, 3% “hard”, etc...

Psi2 : 20% “users”, 5% “consumer reports”, etc...
LDA

- Probability distributions are assigned Dirichlet priors (for LDA).
  - Can use other priors:
    - Hierarchical Dirichlet
    - Hierarchical Dirichlet Process
    - Pitman-Yor Process
- A flexible prior is important for learning.
ILDA

- Interdependent LDA (ILDA)
  - Extension of LDA for aspect-based opinion mining.
ILDA

- ILDA separates “target” and “opinion” words.
  - a : aspect/topic
  - t : target
  - r : sentiment/rating
  - o : opinion

---

**ILDA**

latent

observed
ILDA

- ILDA models the sentiments of each aspects.
  - What is the proportion of positive sentiment for aspect “mobile phone”?
ILDA

- Theta : document-topic distributions
- Psi : aspect-target distributions
- Phi : sentiment-opinion distributions
- Eta : aspect-sentiment distributions

Alphas are the priors
ILDA

• Problem with ILDA
  – Sentiment/Rating is arbitrary.
    • Need to manually inspect and give them positive/neutral/negative labels.
  – Does not consider target-opinion interaction directly.
    • eg: “short camera quality” is plausible in ILDA.
Outline

• Motivation
• Background and Previous Work
• Twitter Opinion Topic Model
• Experiments
Twitter Opinion Topic Model (TOTM)

- Designed to extract opinions from tweets.
- Use state-of-the-art Bayesian non-parametric modelling - Hierarchical Pitman-Yor process
Twitter Opinion Topic Model (TOTM)

- Model target-opinion interaction directly.
  - Tasty burger is more likely than friendly burger.
Twitter Opinion Topic Model (TOTM)

- Makes use of emoticons to learn sentiment.
  - Positive opinions tend to come with positive emoticon 😊
Twitter Opinion Topic Model (TOTM)

• Hierarchical priors for opinion words.
  – Model both target-specific and general sentiment-opinion distributions.
Twitter Opinion Topic Model (TOTM)

- Can use existing sentiment lexicon as prior.
  - We use SentiStrength and MPQA lexicons.
Sentiment Prior Formulation

• TOTM uses a tuneable parameter $b$ that control the strength of the sentiment lexicon:

$$\phi_{rv}^* \propto (1 + b)^{X_{rv}}$$

• $X = $ the sentiment score for sentiment $r$
  – Higher value means stronger sentiment.

• Easy to differentiate => simple to learn $b$. 
Sentiment Prior Formulation

• How to formulate X?

\[ X_{rv} = \begin{cases} 
S_v & \text{if } r = 1 \text{ (positive)} \\
-|S_v| & \text{if } r = 0 \text{ (neutral)} \\
-S_v & \text{if } r = -1 \text{ (negative)} 
\end{cases} \]

• S = sentiment score from lexicon
  – Assumed positive sentiment => positive S
  – Negative sentiment => negative S
Example

• For the word “happy”:
  – SentiStrength score, $S = +2$
  – So $X = +2$ for positive sentiment
    $X = -2$ for neutral sentiment
    $X = -2$ for negative sentiment
• Hence it is a priori more likely for “happy” to be given a positive sentiment.
Training TOTM

• Collapsed Gibbs Sampling for Hierarchical PYP
  – Probability distributions are integrated out.
  – Store information as counts, like LDA.
  – Algorithm consists of decrementing and incrementing the counts.
  – More details in the paper.
Learning Hyperparameters

• For PYP hyperparameters
  – Use auxiliary variable sampler (Teh, 2006).

• For tuneable sentiment strength parameter $b$
  – Use gradient ascent:
    • new $b = \text{old } b + \text{gradient } \times \text{learning rate}$
    • Gradient $t'(b) = \frac{1}{(1 + b)} \sum_r \sum_v c_{rv} (X_{rv} - \mathbb{E}_{\phi_r}[X_r]) + \rho'(b)$
  – Quite intuitive:
    • Increase $b$ if the sentiment score is greater than expected.
Outline

• Motivation
• Background and Previous Work
• Twitter Opinion Topic Model
• Experiments
Data

• Use 3 corpus:
  – From Twitter 7 dataset (Yang & Leskovec, 2011)
    • Query 9 millions tweets on Electronic Products.
    • Non-English tweets are removed.
    • Tweets containing URL are removed.
  – Sentiment 140 tweets (Go et al., 2009)
    • 1.6 millions tweets annotated using emoticons.
  – SemEval tweets (Nakov et al, 2013)
    • 6322 tweets annotated by humans (Mechanical Turk).
Data Preprocessing for TOTM

- Convert raw tweets to target-opinion pairs.

![Diagram showing the data preprocessing steps for TOTM]

- Tweets → Twitter NLP → Part of Speech Tagging → Normalization → Stanford Dependency Parser → Target-opinion Extraction, Emotion Indicator Extraction → Hashtag Aggregation, Remove Infrequent Tags → Decapitalize, Remove Stop Words, Common Words and Infrequent Words → Processed Tweets
Data Preprocessing for TOTM

• Part-of-Speech Tagging
  – with TwitterNLP (Owoputi, 2013).
  – State-of-the-art for tweets.
  – Also tokenise the tweets.

@user yep, the quality of the new iPhone is really good :) ! i like it .

@ ! , D N P D A ^ V R A E , O V O ,

Noun Adjective Proper Noun
Data Preprocessing for TOTM

• Normalisation
  – Use a conversion dictionary from Han et al. (2012).
  – But do not convert proper noun (iPhone → phone).
  – Examples:
    • amazing → amazing
    • nite → night
    • 2morrow → tomorrow
Data Preprocessing for TOTM

• Target-opinion extraction
  – Use Stanford Dependency parser (De Marneffe et al., 2006).
  – Convert relations to target-opinion pairs.
  – Rules:
    
    \[
    \text{amod}(N, A) \rightarrow < N, A > \\
    \text{acomp}(V, A) + \text{nsubj}(V, N) \rightarrow < N, A > \\
    \text{cop}(A, V) + \text{nsubj}(A, N) \rightarrow < N, A > \\
    < h_1, m > + \text{conj\_and}(h_1, h_2) \rightarrow < h_2, m > \\
    < h, m_1 > + \text{conj\_and}(m_1, m_2) \rightarrow < h, m_2 > \\
    < h, m > + \text{neg}(m, \text{not}) \rightarrow < h, \text{not} + m > \\
    < h, m > + \text{nn}(h, N) \rightarrow < N + h, m > \\
    < h, m > + \text{nn}(N, h) \rightarrow < h + N, m > 
    \]
Data Preprocessing for TOTM

• Extract positive or negative emoticons
  – Use both eastern and western smileys:

<table>
<thead>
<tr>
<th></th>
<th>Eastern</th>
<th>Western</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>^_^ (^[ ]^)</td>
<td>:) =-)</td>
</tr>
<tr>
<td>Negative</td>
<td>&lt;( <code>^</code>)&gt; T_T</td>
<td>:@ :'(</td>
</tr>
</tbody>
</table>

– Use strong sentiment words
  • Such as “happy”, “sad”, etc.
Data Preprocessing for TOTM

• Aggregate tweets based on hashtags
  – Word co-occurrence to be used by topic model.
  – Give a different way to view the results.

• Remove stop words, common words and rare words.
  – These words are of less interest.
  – eg: “he”, “she”, misspellings, etc...
Experiments

• Compare TOTM with 2 baselines:
  – ILDA as mentioned previously.
  – LDA-DP
    • Vanilla LDA but apply ad hoc modification to the prior following He (2012).
    • Set $\phi_{rv}^*$ to 0.9 if sentiment for word $v$ is the same as $r$, else set to 0.05.
Experiments

• Quantitative Evaluations
  – Perplexity
  – Sentiment classification
  – Sentiment prior evaluation

• Qualitative Evaluations
  – Inspecting word distributions
  – Comparing opinions
  – Opinions extraction
Perplexity

- Commonly used to evaluate topic models.
- Negatively related to the log likelihood of observed words.
  - So lower perplexity is better.

\[
\text{perplexity}(W) = \exp \left( - \frac{\sum_{d=1}^{D} \log P(\vec{w}_d)}{\sum_{d=1}^{D} N_d} \right)
\]

Log likelihood for words in the test set

Normaliser (Number of words)
Perplexity

• Results

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>Opinion</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA-DP</td>
<td>N/A</td>
<td>510.15 ± 0.08</td>
<td>N/A</td>
</tr>
<tr>
<td>ILDA</td>
<td>594.81 ± 13.61</td>
<td>519.84 ± 0.43</td>
<td>556.03 ± 6.22</td>
</tr>
<tr>
<td>TOTM</td>
<td>592.91 ± 13.86</td>
<td>137.42 ± 0.28</td>
<td>285.42 ± 3.23</td>
</tr>
</tbody>
</table>

• Significant improvement on opinion words
  – since TOTM model target-opinion interaction directly, i.e. better prediction for opinion words.
Sentiment Classification

• Evaluate on annotated tweets.
• Predict sentiment by selecting the polarity that has higher likelihood given the sentiment-word distributions.

\[
polarity(d) = \arg \max_{r=\{-1, 1\}} \prod_i \phi_{r, o_{di}}
\]
Sentiment Classification

• Results

<table>
<thead>
<tr>
<th></th>
<th>Sent140 Tweets</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Precision</td>
<td>Recall</td>
<td>F1-score</td>
<td></td>
</tr>
<tr>
<td>LDA-DP</td>
<td>57.3</td>
<td>56.1</td>
<td>90.1</td>
<td>69.2</td>
<td></td>
</tr>
<tr>
<td>ILDA</td>
<td>54.1</td>
<td>56.9</td>
<td>55.3</td>
<td>55.9</td>
<td></td>
</tr>
<tr>
<td>TOTM</td>
<td>65.0</td>
<td>61.7</td>
<td>90.2</td>
<td>73.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SemEval Tweets</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Precision</td>
<td>Recall</td>
<td>F1-score</td>
<td></td>
</tr>
<tr>
<td>LDA-DP</td>
<td>52.1</td>
<td>65.0</td>
<td>58.3</td>
<td>61.4</td>
<td></td>
</tr>
<tr>
<td>ILDA</td>
<td>46.8</td>
<td>60.7</td>
<td>53.6</td>
<td>56.3</td>
<td></td>
</tr>
<tr>
<td>TOTM</td>
<td><strong>73.3</strong></td>
<td><strong>84.0</strong></td>
<td><strong>74.9</strong></td>
<td><strong>79.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

• TOTM performs best in sentiment classification.
Evaluating Sentiment Prior

• Use SentiWordNet to evaluate the learned sentiment-opinion distributions.

• SentiWordNet gives positive affinity and negative affinity for each word, eg:
  – “Active” -> positive 0.5, negative 0.125
  – “Supreme” -> positive 0.75, negative 0

• So can calculate both positivity and negativity of an opinion word distribution.
Evaluating Sentiment Prior

- Evaluation metric
  - Sentiment score – expected sentiment under an opinion word distribution.

\[
Score(\phi_r, Z) = E_{\phi_r}[Z] = \sum_{v=1}^{V_o} Z_v \phi_{rv}
\]

- \(Z\) = positive or negative affinity from SentiWordNet
- Opinion word distribution
Evaluating Sentiment Prior

- **Results**

<table>
<thead>
<tr>
<th></th>
<th>Electronic Product Tweets</th>
<th>Sent140 Tweets</th>
<th>SemEval Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negativity</td>
<td>Positivity</td>
<td>Negativity</td>
</tr>
<tr>
<td>No lexicon</td>
<td>17.82 ± 1.26</td>
<td>17.39 ± 0.45</td>
<td>22.63 ± 0.96</td>
</tr>
<tr>
<td>MPQA</td>
<td>23.91 ± 0.49</td>
<td>31.96 ± 0.09</td>
<td>24.10 ± 0.49</td>
</tr>
<tr>
<td>SentiStrength</td>
<td>23.19 ± 0.08</td>
<td>35.69 ± 0.33</td>
<td>24.29 ± 1.07</td>
</tr>
</tbody>
</table>

- No lexicon = use only emoticons
- SentiStrength is slightly better than MPQA lexicon.
- Sentiment lexicon gives significant improvement.
Experiments

• Quantitative Evaluations
  -- Perplexity
  -- Sentiment classification
  -- Sentiment prior evaluation

• Qualitative Evaluations
  – Inspecting word distributions
  – Comparing opinions
  – Opinions extraction
Inspecting Word Distributions

• We can inspect aspect-target word distributions to see if the target words are correctly clustered.
  – Some examples:

<table>
<thead>
<tr>
<th>Aspects (a)</th>
<th>Target Words (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>camera, pictures, video camera, shots</td>
</tr>
<tr>
<td>Apple iPod</td>
<td>ipod, ipod touch, songs, song, music</td>
</tr>
<tr>
<td>Android phone</td>
<td>android, apps, app, phones, keyboard</td>
</tr>
<tr>
<td>Macbook</td>
<td>macbook, macbook pro, macbook air</td>
</tr>
<tr>
<td>Nintendo games</td>
<td>nintendo, games, game, gameboy</td>
</tr>
</tbody>
</table>

– Target words are closely related.
Inspecting Word Distributions

- Similarly, we can inspect the opinion word distributions.
  - TOTM allows in depth analysis by looking at opinion word distributions for a particular target.

<table>
<thead>
<tr>
<th>Target (t)</th>
<th>+/−</th>
<th>Opinions (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>phone</td>
<td>−</td>
<td>dead, damn, stupid, bad, crazy</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>mobile smart, good, great, f***ing</td>
</tr>
<tr>
<td>battery life</td>
<td>−</td>
<td>terrible, poor, bad, horrible, non-existence</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>good, long, great, 7hr ultralong</td>
</tr>
<tr>
<td>game</td>
<td>−</td>
<td>addictive, stupid, free, full, addicting</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>great, good, awesome, favorite, cat-and-mouse</td>
</tr>
<tr>
<td>sausage</td>
<td>−</td>
<td>silly, argentinian, cold, huge, stupid</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>hot, grilled, good, sweet, awesome</td>
</tr>
</tbody>
</table>

* Words in **bold** are more specific and can only describe certain targets.
Comparing Opinions

• Aggregating tweets using hashtags allows additional analysis.
  – We inspect hashtags that correspond to electronic companies such as #sony, #canon, #samsung...
Comparing Opinions

- A snapshot

<table>
<thead>
<tr>
<th>Brands</th>
<th>Sentiment</th>
<th>Aspects / Targets’ Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Camera</td>
</tr>
<tr>
<td>Canon</td>
<td>-</td>
<td>camera → expensive small bad</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lens → prime cheap broken</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>camera → great compact amazing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pictures → great nice creative</td>
</tr>
<tr>
<td>Sony</td>
<td>-</td>
<td>camera → big crappy defective</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lens → vertical cheap wide</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>photos → great lovely amazing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>camera → good great nice</td>
</tr>
<tr>
<td>Samsung</td>
<td>-</td>
<td>camera → digital free crazy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>shots → quick wide</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>camera → gorgeous great cool</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pics → nice great perfect</td>
</tr>
</tbody>
</table>
Opinions Extraction

- Finally, TOTM allows us to query tweets that correspond to certain opinions.
  - Example: query opinions on iPhone

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT @user: the iPhone is so awesome!!! Emailing, texting, surfing the</td>
<td>@user awww thx! I can’t send an email right now bc my iPhone is stupid</td>
</tr>
<tr>
<td>same time! — Can do all that while talkin on the phone?...</td>
<td>with sending emails. Lol but I can tweet or dm u?</td>
</tr>
<tr>
<td>Ahhh! Tweeting on my gorgeous iPhone! I missed you! hehe am on my</td>
<td>It would appear that the iPhone, due to construction, is weak at holding</td>
</tr>
<tr>
<td>way home, put the kettle on will you pls : )</td>
<td>signal. Combine that with a bullshit 3G network in Denver.</td>
</tr>
<tr>
<td>Thanks @user for the link to iPhone vs Blackberry debate. I got the</td>
<td>@user @user Ah, well there you go.</td>
</tr>
<tr>
<td>iPhone &amp; it’s just magic! So intuitive!</td>
<td>The iPhone is dead, long live Android! : )</td>
</tr>
<tr>
<td>Finally my fave lover @user has Twitter &amp; will be using it all the</td>
<td>@user Finally eh? :D I think iphone is so ugly x.x</td>
</tr>
<tr>
<td>time with her cool new iPhone :)</td>
<td></td>
</tr>
</tbody>
</table>
Major Contributions

• Introduce TOTM for aspect-based opinion mining on tweets.
  – Makes use of emoticons and hashtags on tweets.

• Novel way of incorporating sentiment prior information into topic model.
  – Simple to implement and allow automatic learning of hyperparameters.

Thanks!
References

References


