Technical Sentiment Analysis
Measuring Advantages and Drawbacks of New Products Using Social Media

Filippo Chiarello, Andrea Bonaccorsi, Gualtiero Fantoni, Giacomo Ossola, Andrea Cimino and Felice Dell’Orletta

University of Pisa: Engineering School and Italian Natural Language Processing Group
Advantages and Drawbacks of Inventions

Achieving users goals

Chiarello, Filippo, Gualtiero Fantoni, and Andrea Bonaccorsi. "Product description in terms of advantages and drawbacks: Exploiting patent information in novel ways." ICED 2017
Advantages and Drawbacks of Inventions

Achieving users goals

Positive Behavior

Negative Behavior

Chiarello, Filippo, Gualtiero Fantoni, and Andrea Bonaccorsi. "Product description in terms of advantages and drawbacks: Exploiting patent information in novel ways." ICED 2017
Advantages and Drawbacks of Inventions

Achieving users goals

Positive Behavior
- efficiency
- safety
- affordability
- robust
- not expensive
- precise
- protect
- effectively
- ergonomic
...

Negative Behavior
- cracked
- dangerous
- useless
- inaccurate
- out of control
- inappropriate
- expensive
- insecure
- unstable
...

Chiarello, Filippo, Gualtiero Fantoni, and Andrea Bonaccorsi. "Product description in terms of advantages and drawbacks: Exploiting patent information in novel ways., ICED 2017
Advantages and Drawbacks of Inventions

Achieving users goals

Positive Behavior

- efficiency
- safety
- affordability
- robust
- not expensive
- precise
- protect
- effectively
- ergonomic
- ...

10,175

Negative Behavior

- cracked
- dangerous
- useless
- inaccurate
- out of control
- inappropriate
- expensive
- insecure
- unstable
- ...

16,053

Chiarello, Filippo, Gualtiero Fantoni, and Andrea Bonaccorsi. "Product description in terms of advantages and drawbacks: Exploiting patent information in novel ways.", ICED 2017
Methodology
Methodology

Products → Data Collection
Methodology

Products → Data Collection → Data Filtering
Methodology

Data Collection → Data Filtering

Sentiment Analysis

Technical Sentiment

Products → Twitter → Data Collection
Methodology

Products → Data Collection → Data Filtering

Sentiment Analysis

Technical Sentiment

Results Comparison

CARMA 2018
Microsoft unveiled **Xbox One X**, a high-end version of Xbox One which has been released on 7th of November 2017.

First teased at E3 2016 under the codename "Project Scorpio", **Xbox One X** features upgraded hardware that is designed primarily to play games at **4K resolution**, and “super sampling” to provide graphical improvements on **high-definition displays**.
The **New Nintendo 2DS XL** is a handheld game console and is the sixth system in the *Nintendo 3DS family* of handheld consoles.

It is a streamlined version of the New Nintendo 3DS XL, retaining most of the enhancements it carries over the original 2DS and 3DS models, including its updated processor, additional controls, and Amiibo support, but with no autostereoscopic 3D display, and a thinner build.
We selected an event able to polarise Twitter users’ attention.

The Electronic Entertainment Expo, commonly referred to as E3, is a premier trade event for the video game industry.

In 2017, the event was held since 13th June to 15th June.
Data Collection

Search vs Streaming API

- Tweeter Search API allows to retrieve all tweets that match a small set of keywords
- Streaming API allows to stream all tweets that Twitter makes accessible through the API itself (roughly 1% of all tweets)

We decided to go with **Streaming API** because it is free
Data Collection

Search vs Streaming API

- Tweeter Search API allows to retrieve all tweets that match a small set of keywords
- Streaming API allows to stream all tweets that Twitter makes accessible through the API itself (roughly 1% of all tweets)

We decided to go with **Streaming API** because it is free

Search query

We built a search query composed from both:

- Hashtags
- Keywords
We collected tweets before, during, and after the tradeshow:

- From 11th June 2017 to 31st July 2017

**Search vs Streaming API**

- Tweeter Search API allows to retrieve all tweets that match a small set of keywords
- Streaming API allows to stream all tweets that Twitter makes accessible through the API itself (roughly 1% of all tweets)

We decided to go with **Streaming API** because it is free

**Search query**

We built a search query composed from both:

- **Hashtags**
- **Keywords**

**Observation time**

We collected tweets before, during, and after the tradeshow:

- From 11th June 2017 to 31st July 2017
Data Collection

Search vs Streaming API

- Tweeter Search API allows to retrieve all tweets that match a small set of keywords
- Streaming API allows to stream all tweets that Twitter makes accessible through the API itself (roughly 1% of all tweets)

We decided to go with **Streaming API** because it is free

Search query

We built a search query composed from both:

- Hashtags
- Keywords

Observation time

We collected tweets before, during, and after the tradeshow:

- From **11th June 2017** to **31st July 2017**

Data size

- We collected more than **28 million** tweets
- We evaluated both timespan and data size to decide when to stop data collection process
Data Collection

<table>
<thead>
<tr>
<th>Xbox One X</th>
<th>New Nintendo 2DS XL</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Xbox</td>
<td>#Nintendo</td>
</tr>
<tr>
<td>#XboxE3</td>
<td>#2DS</td>
</tr>
<tr>
<td>#Scorpio</td>
<td>#2DSXL</td>
</tr>
<tr>
<td>#ProjectScorpio</td>
<td>#NewNintendo2DS</td>
</tr>
<tr>
<td>#XboxScorpio</td>
<td>#NewNintendo2DSXL</td>
</tr>
<tr>
<td>#xboxnext</td>
<td></td>
</tr>
<tr>
<td>#xboxone</td>
<td></td>
</tr>
<tr>
<td>#XboxOneX</td>
<td></td>
</tr>
</tbody>
</table>

Note: Microsoft’s console were initially marketed as “Project Scorpio”. The company disclosed the official name during the tradeshow.
## Data Filtering

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
<th>N. of Output Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets written in different languages</td>
<td><strong>Language classifier</strong> to identify english tweets</td>
<td>7.165.216</td>
</tr>
<tr>
<td>Advertising</td>
<td><strong>Supported Vector Machine classifier</strong> to identify relevant tweets</td>
<td>66.796</td>
</tr>
</tbody>
</table>
SVM classifier for irrelevant tweets

We defined characteristics that make a tweet:

- **relevant**: posted by users or containing words or opinions related to our products of interests and their functionalities
- **irrelevant** tweets containing advertisings, links to e-commerce websites or messages related to other products or subjects

A researcher manually classified a subset made up of randomly extracted tweets, composed of 6,500 finding 105 relevant tweets and 6,395 irrelevant tweets.
SVM classifier for irrelevant tweets

<table>
<thead>
<tr>
<th>Text</th>
<th>Y/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does Xbox really think we don't know the One X is a One with a stronger processor and 499 is a decent price for a modest hardware upgrade?</td>
<td>Y</td>
</tr>
<tr>
<td>@NintendoAmerica Please explain how my 2DS can play Mario Maker, but can't download Super Mario World. Why do I need to buy a 2DSXL?</td>
<td>Y</td>
</tr>
<tr>
<td>Trappin out the condo 30 shot extendo pop the cartridge in like we play Nintendo</td>
<td>N</td>
</tr>
<tr>
<td>Emotionally connecting with your friends enables you to stay i... More for Xbox <a href="https://t.co/7gZyxKHU02">https://t.co/7gZyxKHU02</a></td>
<td>N</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

SVM model was then trained using this dataset, and computed a probability for each tweet to be relevant or irrelevant. A threshold of 0.7 has been chosen to label a tweet as relevant or irrelevant.
Sentiment Analysis & Technical Sentiment
Sentiment Analysis & Technical Sentiment

- We used a deep learning (RNN) classifier developed by Cimino (2016).

- We used a deep learning (RNN) classifier developed by Cimino (2016).
- We pre-processed the tweets by removing mentions (@ character), URLs, product hashtags, emoticons and single characters.

- We used a deep learning (RNN) classifier developed by Cimino (2016).
- We pre-processed the tweets by removing mentions (@ character), URLs, product hashtags, emoticons and single characters.
- As a result, for each tweet we obtained a probability of belonging to a mood class.
- We used a deep learning (RNN) classifier developed by Cimino (2016).
- We pre-processed the tweets by removing mentions (@ character), URLs, product hashtags, emoticons and single characters.
- As a result, for each tweet we obtained a probability of belonging to a mood class.
- After a manual analysis, we used a class prediction probability threshold of 0.6 to filter out low confidence prediction, i.e. tweets that cannot be classified as positive or negative with a high confidence are classified as neutral instead.

- We used a deep learning (RNN) classifier developed by Cimino (2016).
- We pre-processed the tweets by removing mentions (@ character), URLs, product hashtags, emoticons and single characters.
- As a result, for each tweet we obtained a probability of belonging to a mood class.
- After a manual analysis, we used a class prediction probability threshold of 0.6 to filter out low confidence prediction, i.e. tweets that cannot be classified as positive or negative with a high confidence are classified as neutral instead.

- We used the lexicon developed in Chiarello (2017)
- Tweets are classified using a simple word count

Sentiment Analysis VS Technical Sentiment

- **Sentiment Analysis**
  - New Nintendo 2DS XL: 92.4%
  - Xbox One X: 87.4%

- **Adv/Drw Lexicon**
  - Negative: 0.9%
  - Positive: 8.8%
  - Neutral: 6.6%

**Polarity**
- Red: Negative
- Green: Positive
- Blue: Neutral

**Comparison**
- Graph shows the percentage of tweets for each console across different sentiment analyses.
Nintendo VS Xbox

- **New Nintendo 2DS XL**
  - Negative: 0.9
  - Positive: 6.6
  - Neutral: 45.4

- **Xbox One X**
  - Negative: 3.7
  - Positive: 8.8
  - Neutral: 59.4

**Analysis Tool**
- Adv/Dnw Lexicon
- Sentiment Analysis

**Polarity**
- Red: Negative
- Green: Positive
- Blue: Neutral
Hypothesis

The low signals in the outputs are related to:

- Different technology complexity of the products
- Different Marketing focus
Hypothesis

The low signals in the outputs are related to:
- Different technology complexity of the products
- Different Marketing focus

We can use technical reviews as a proxy
Google Trends

Queries entered into Google trends engine to search for **technical reviews**:

- **Xbox One X review**
- **New N2DS XL review**
Google Search

Queries entered into Google search engine to search for **technical reviews**:

<table>
<thead>
<tr>
<th>Search queries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Xbox One X</strong></td>
</tr>
<tr>
<td><strong>New Nintendo 2DS XL</strong></td>
</tr>
</tbody>
</table>

We obtained **1.117** articles about **Xbox** and only **52** about **Nintendo**
Marketing Campaigns

Xbox
@Xbox

Fully specced for 4K gaming. #XboxOneX
bit.ly/2t2zlGs

- 6 Teraflop GPU
- 12GB GDDR5 graphic memory
- 326GB/s memory bandwidth
- 8GB flash memory
- 1TB HDD internal storage

10:35 - 14 giu 2017
816 Retweet 3.112 Mi piace

Nintendo of America
@NintendoAmerica

Play the existing and upcoming library of 3DS games in style with the New Nintendo #2DSXL, available now! bit.ly/2veqANZ

11:00 - 28 lug 2017
212 Retweet 1.202 Mi piace
Conclusions

- Advantages and drawbacks analysis is more effective and efficient in giving technical-functional judgements about a product.
Conclusions

- Advantages and drawbacks analysis is more effective and efficient in giving technical-functional judgements about a product.

- It can give complementary results with respect to sentiment analysis when a product has a certain technological complexity and fuels a more technical social media discourse.
THANKS FOR YOUR ATTENTION!

Technical Sentiment Analysis: Measuring Advantages and Drawbacks of New Products Using Social Media

filippochiarello.90@gmail.com

@f_chiare