Maintaining sentiment polarity in translation of user-generated content

Pintu Lohar, Haithem Afli and Andy Way

ADAPT Centre, School of Computing, Dublin City University
Contents

- Objective & Motivation
- Sentiment analysis of user-generated content
- Data Preparation
  - Corpus development
  - Sentiment annotation and classification
- Experiments
  - Sentiment Translation Architecture
  - Results
  - Discussion
- Conclusions and future work
Objective

- Analyse sentiment preservation & MT quality in the context of user-generated content (UGC)
Objective

- Analyse sentiment preservation & MT quality in the context of user-generated content (UGC)

- Focus on whether sentiment classification helps improve sentiment preservation in MT of UGC
Motivation

• Translation quality *per se* is not the main concern
Motivation

• Translation quality *per se* is not the main concern

  ▪ *Sentiment preservation* is (arguably more) important

  e.g. companies want to know what their customers think of their products and services.

  It is **crucial** that user sentiment in one language is preserved in the target language (typically, English).
Motivation

Microsoft → Customer feedback in Japanese
Motivation

Japanese data → Translate → English data → Sentiment analysis → Sentiment classes → Customer feedback in Japanese
Track Record in UGC

Number of tweets: 835,725

Top Issues

Focused Issue

Candidate Mentions

Enda Kenny
18,682 Tweets
38% Positive / 42% Negative

Top Party Mentions

A World Leading SFI Research Centre
Track Record in UGC

13 languages and 24 language pairs

85,047,110 tweets in total
Sentiment analysis of UGC

- UGC includes blog posts, podcasts, online videos, tweets etc.

- UGC is usually multilingual and of varying quality (sometimes deliberately)

- Sentiment analysis of UGC has many applications
Crosslingual sentiment analysis (CLSA):

- The task of predicting the polarity of the opinion of a text in a language using a classifier trained on the corpus of another language (Balamurli et al. (2012))
Crosslingual sentiment analysis (CLSA):

- The task of predicting the polarity of the opinion of a text in a language using a classifier trained on the corpus of another language (Balamurli et al. (2012))

MT-based CLSA:

- MT is utilized to leverage its capability, existing SA resources available in English to classify sentiment in other languages (Mihalcea et al. (2012))
MT can alter the sentiment (Mohammad et al. (2016))

Google Translate from English to German on 25/05/2017

English: he is out of the world cup
German: Er ist aus des weltmeisterschaft

English: negative
German: neutral
• Can a sentiment classification approach help improve sentiment preservation in the target language?
• Can a sentiment classification approach help improve sentiment preservation in the target language?

• Is it useful to select a specific-sentimented MT model to translate the UGC with the same sentiment?
Corpus development:

- Twitter data set comprising 4,000 English tweets from the FIFA World Cup 2014 and their manual translations into German
Data preparation

Corpus development:

- Twitter data set comprising 4,000 English tweets from the FIFA World Cup 2014 and their manual translations into German

- Informal translations of English tweets into German

e.g. English tweet: *Goaaaaal*  German tweet: *Tooooor*
Sentiment annotation
Manually annotated sentiment scores between 0 and 1
Sentiment annotation and classification

- **Sentiment annotation**
  Manually annotated sentiment scores between 0 and 1

- **Sentiment classes**
  
  (i) **Negative**: sentiment score \( \leq 0.4 \)
  
  (ii) **Neutral**: sentiment score \( \approx 0.5 \)
  
  (iii) **Positive**: sentiment score \( \geq 0.6 \)

  e.g. Tweet: 
  
  *injured Neymar out of World Cup*  
  Sentiment score: 0.2
- Manual annotation of Twitter data is considered as the “gold-standard”
Sentiment annotation and classification

- Manual annotation of Twitter data is considered as the “gold-standard”
- 50 tweets per sentiment (negative, neutral and positive) are held out for tuning and testing purposes

<table>
<thead>
<tr>
<th>Data</th>
<th>Train</th>
<th>Development</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#neg</td>
<td>#neu</td>
<td>#pos</td>
</tr>
<tr>
<td>Twitter</td>
<td>3,700</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Data distribution of Twitter data for Training, development and test
- Flickr and News commentary (``News'') data are used as additional resources
- Automatic sentiment analysis tool (Afli et. al. (2017)) is applied to Flickr and News data
Flickr and News commentary (``News’’) data are used as additional resources.

Automatic sentiment analysis tool (Afli et. al. (2017)) is applied to Flickr and News data.

**Performance accuracy:**

- 2,994 tweets out of 4,000 correctly classified by this tool when compared to the ‘gold standard’ data.
- Accuracy = 74.85%
### Data distribution after sentiment classification

<table>
<thead>
<tr>
<th>Data</th>
<th>Sentiment classification</th>
<th>#neg</th>
<th>#neu</th>
<th>#pos</th>
<th>#total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>manual</td>
<td>919</td>
<td>1,308</td>
<td>1,473</td>
<td>3,700</td>
</tr>
<tr>
<td>Flickr</td>
<td>automatic</td>
<td>9,677</td>
<td>11,065</td>
<td>8,258</td>
<td>29,000</td>
</tr>
<tr>
<td>News</td>
<td>automatic</td>
<td>111,337</td>
<td>14,306</td>
<td>113,200</td>
<td>238,843</td>
</tr>
</tbody>
</table>
I. Translation without sentiment classification
Experiments

I. Translation without sentiment classification

II. Translation with sentiment classification
   i. Manual sentiment classification (only Twitter data)
   ii. Automatic sentiment classification (Flickr & News data)
Experiments

I. Translation without sentiment classification

II. Translation with sentiment classification
   i. Manual sentiment classification (only Twitter data)
   ii. Automatic sentiment classification (Flickr & News data)

III. Translation by wrong MT engines
   i. Negative tweets by positive model
   ii. Neutral tweets by negative model
   iii. Positive tweets by neutral model
Sentiment Translation Architecture

Sentiment Classification

Parallel corpus

No Sentiment Classification
Sentiment Translation Architecture

- Manual
- Sentiment Classification
- Parallel corpus
- Automatic
- No Sentiment Classification
Sentiment Translation Architecture

- Manual
  - Negative model
  - Neutral model
  - Positive model

- Sentiment Classification

- Parallel corpus

- Automatic

- No Sentiment Classification
Sentiment Translation Architecture

- **Sentiment Classification**
  - Manual
    - Negative model
    - Neutral model
    - Positive model
  - Automatic
    - Negative model
    - Neutral model
    - Positive model
  - Parallel corpus
  - No Sentiment Classification
Sentiment Translation Architecture

Parallel corpus

Sentiment Classification

Manual
  - Negative model
  - Neutral model
  - Positive model

Automatic
  - Negative model
  - Neutral model
  - Positive model

No Sentiment Classification
  - Baseline model
Sentiment Translation Architecture

- Sentiment Classification
  - Manual
    - Negative model
    - Neutral model
    - Positive model
  - Automatic
    - Negative model
    - Neutral model
    - Positive model
  - No Sentiment Classification
    - Baseline model

- Parallel corpus

- Translate
Sentiment Translation Architecture

Sentiment Classification

Manual
- Negative model
- Neutral model
- Positive model

Automatic
- Negative model
- Neutral model
- Positive model

Parallel corpus

No Sentiment Classification
- Baseline model

Translate
- Negative test
- Neutral test
- Positive test
Sentiment Translation Architecture

Sentiment Classification
- Manual
  - Negative model
  - Neutral model
  - Positive model
- Automatic
  - Negative model
  - Neutral model
  - Positive model

Parallel corpus
- No Sentiment Classification
  - Negative test
  - Neutral test
  - Positive test

Translate
- Negative model
- Neutral model
- Positive model
- Baseline model

A World Leading SFI Research Centre
Sentiment Translation Architecture

Sentiment Classification

Manual
- Negative model
- Neutral model
- Positive model

Automatic
- Negative model
- Neutral model
- Positive model

Parallel corpus

No Sentiment Classification
- Baseline model

Translate
- Negative test
- Neutral test
- Positive test

Whole test data
Sentiment Translation Architecture

- **Parallel corpus**
  - Manual
    - Negative model
      - Negative test
        - Negative translation
    - Neutral model
      - Neutral test
        - Neutral translation
    - Positive model
      - Positive test
        - Positive translation
  - Automatic
    - Negative model
      - Negative test
    - Neutral model
      - Neutral test
    - Positive model
      - Positive test
- **No Sentiment Classification**
  - Baseline model
  - whole test data
Sentiment Translation Architecture

Manual
- Negative model
- Neutral model
- Positive model

Automatic
- Negative model
- Neutral model
- Positive model

Parallel corpus

No Sentiment Classification

Baseline model
- whole test data

Sentiment Classification

Translate
- Negative test
- Neutral test
- Positive test

Negative translation
Neutral translation
Positive translation

Sentiment Translation Architecture
Sentiment Translation Architecture

- Sentiment Classification
  - Manual
    - Negative model
    - Neutral model
    - Positive model
    - Translate
    - Negative test
    - Neutral test
    - Positive test
    - Negative translation
    - Neutral translation
    - Positive translation
    - Output combination 1
  - Automatic
    - Negative model
    - Neutral model
    - Positive model
    - Parallel corpus
    - Negative test
    - Neutral test
    - Positive test
    - Negative translation
    - Neutral translation
    - Positive translation
    - Output combination 2
  - No Sentiment Classification
    - Baseline model
    - Whole test data

A World Leading SFI Research Centre
Sentiment Translation Architecture

- Parallel corpus
  - Sentiment Classification
    - Manual
      - Negative model
      - Neutral model
      - Positive model
    - Automatic
      - Negative model
      - Neutral model
      - Positive model
  - No Sentiment Classification
    - Baseline model

- Translate
  - Negative test
  - Neutral test
  - Positive test
  - Negative translation
  - Neutral translation
  - Positive translation

- Output combination1
- Output combination2
- Baseline translation

A World Leading SFI Research Centre
Sentiment Translation Architecture

- Manual Sentiment Classification
  - Negative model
  - Neutral model
  - Positive model
- Automatic Sentiment Classification
  - Negative model
  - Neutral model
  - Positive model
- Parallel corpus
- No Sentiment Classification
- Baseline model

- Sentiment Translation
  - Negative test
  - Neutral test
  - Positive test
  - Negative translation
  - Neutral translation
  - Positive translation
- Output combination 1
- Evaluate and measure sentiment preservation
- Output combination 2
- whole test data
- Baseline translation

www.adaptcentre.ie
## Results

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Data size</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>4k</td>
<td>✓</td>
<td>48.2</td>
<td>59.4</td>
<td>34.2</td>
<td>72.66%</td>
</tr>
<tr>
<td>Twitter (Baseline)</td>
<td></td>
<td>×</td>
<td>50.3</td>
<td>60.9</td>
<td>31.9</td>
<td>66.66%</td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>33k</td>
<td>✓</td>
<td>48.5</td>
<td>59.8</td>
<td>33.9</td>
<td>71.33%</td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td></td>
<td>×</td>
<td>50.7</td>
<td>62.0</td>
<td>31.3</td>
<td>62.66%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>✓</td>
<td>50.3</td>
<td>62.3</td>
<td>31.0</td>
<td>75.33%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>×</td>
<td>52.0</td>
<td>63.4</td>
<td>30.1</td>
<td>73.33%</td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>4k</td>
<td>✓</td>
<td>46.9</td>
<td>57.9</td>
<td>35.4</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

*Experimental evaluation with data concatenation*
## Results

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Data size</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>4k</td>
<td>✓</td>
<td>48.2</td>
<td>59.4</td>
<td>34.2</td>
<td>72.66%</td>
</tr>
<tr>
<td>(Baseline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>33k</td>
<td>✓</td>
<td>48.5</td>
<td>59.8</td>
<td>33.9</td>
<td>71.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>✓</td>
<td>50.3</td>
<td>62.3</td>
<td>31.0</td>
<td>75.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>4k</td>
<td>✓</td>
<td>46.9</td>
<td>57.9</td>
<td>35.4</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

---

*Experimental evaluation with data concatenation*
## Results

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Data size</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>4k</td>
<td>✓</td>
<td>48.2</td>
<td>59.4</td>
<td>34.2</td>
<td>72.66%</td>
</tr>
<tr>
<td>Twitter (Baseline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>33k</td>
<td>✓</td>
<td>48.5</td>
<td>59.8</td>
<td>33.9</td>
<td>71.33%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>✓</td>
<td>50.3</td>
<td>62.3</td>
<td>31.0</td>
<td>75.33%</td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>4k</td>
<td>✓</td>
<td>46.9</td>
<td>57.9</td>
<td>35.4</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

Experimental evaluation with data concatenation
<table>
<thead>
<tr>
<th>Translation model</th>
<th>Data size</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>4k</td>
<td>✓</td>
<td>48.2</td>
<td>59.4</td>
<td>34.2</td>
<td>72.66%</td>
</tr>
<tr>
<td>Twitter (Baseline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>33k</td>
<td>✓</td>
<td>48.5</td>
<td>59.8</td>
<td>33.9</td>
<td>71.33%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>✓</td>
<td>50.3</td>
<td>62.3</td>
<td>31.0</td>
<td><strong>75.33%</strong></td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>4k</td>
<td>✓</td>
<td>46.9</td>
<td>57.9</td>
<td>35.4</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

**Experimental evaluation with data concatenation**

A World Leading SFI Research Centre
## Results

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Data size</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>4k</td>
<td>✓</td>
<td>48.2</td>
<td>59.4</td>
<td>34.2</td>
<td>72.66%</td>
</tr>
<tr>
<td>Twitter (Baseline)</td>
<td></td>
<td></td>
<td>50.3</td>
<td>60.9</td>
<td>31.9</td>
<td>66.66%</td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>33k</td>
<td>✓</td>
<td>48.5</td>
<td>59.8</td>
<td>33.9</td>
<td>71.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50.7</td>
<td>62.0</td>
<td>31.3</td>
<td>62.66%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>272k</td>
<td>✓</td>
<td>50.3</td>
<td>62.3</td>
<td>31.0</td>
<td><strong>75.33%</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>52.0</td>
<td>63.4</td>
<td>30.1</td>
<td>73.33%</td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>4k</td>
<td>✓</td>
<td>46.9</td>
<td>57.9</td>
<td>35.4</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

*Experimental evaluation with data concatenation*
## Examples

<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation model</th>
<th>Baseline model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Howard Web is a terrible ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
</tr>
<tr>
<td>2</td>
<td>injured Neymar out of World Cup 2014</td>
<td>verletzter Neymar out the WC2014</td>
<td>verletzter Neymar out of World Cup 2014</td>
</tr>
<tr>
<td>3</td>
<td>penalty shootouts are too intense !</td>
<td>penalty shoot is to intensiv !</td>
<td>penalties is to intensiv !</td>
</tr>
<tr>
<td>4</td>
<td>damn chile is nice !!!! #WorldCup</td>
<td>freeking Chile is good !!! #WorldCup</td>
<td>damn Chile is good !!! #WorldCup</td>
</tr>
<tr>
<td>5</td>
<td>a bit boring ...</td>
<td>a little boring ...</td>
<td>some boring ...</td>
</tr>
<tr>
<td>6</td>
<td>im with Germany</td>
<td>I stand to Germany’s side</td>
<td>I stand to Deutschlands side</td>
</tr>
<tr>
<td>7</td>
<td>as getting I, GO CHILE !</td>
<td>completely mache I it GO CHILE !</td>
<td>as getting I, GO CHILE !</td>
</tr>
</tbody>
</table>

*Comparison of translations by sentiment translation models and Baseline model*
<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation model</th>
<th>Baseline model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Howard Web is a terrible ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
</tr>
<tr>
<td>2</td>
<td>injured Neymar out of World Cup 2014</td>
<td>verletzter Neymar out the WC2014</td>
<td>verletzter Neymar out of World Cup 2014</td>
</tr>
<tr>
<td>3</td>
<td>penalty shootouts are too intense !</td>
<td>penalty shoot is to intensiv !</td>
<td>penalties is to intensiv !</td>
</tr>
<tr>
<td>4</td>
<td>damn chile is nice !!!!! #WorldCup</td>
<td>freeking Chile is good !!! #WorldCup</td>
<td>damn Chile is good !!! #WorldCup</td>
</tr>
<tr>
<td>5</td>
<td>a bit boring ...</td>
<td>a little boring ...</td>
<td>some boring ...</td>
</tr>
<tr>
<td>6</td>
<td>im with Germany</td>
<td>I stand to Germany’s side</td>
<td>I stand to Deutschlands side</td>
</tr>
<tr>
<td>7</td>
<td>as getting I, GO CHILE !</td>
<td>completely mache I it GO CHILE !</td>
<td>as getting I, GO CHILE !</td>
</tr>
</tbody>
</table>

Comparison of translations by sentiment translation models and Baseline model
<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation model</th>
<th>Baseline model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Howard Web is a terrible ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
</tr>
<tr>
<td>2</td>
<td>injured Neymar out of World Cup 2014</td>
<td>verletzter Neymar out the WC2014</td>
<td>verletzter Neymar out of World Cup 2014</td>
</tr>
<tr>
<td>3</td>
<td>penalty shootouts are too intense!</td>
<td>penalty shoot is to intensiv!</td>
<td>penalties is to intensiv!</td>
</tr>
<tr>
<td>4</td>
<td>damn chile is nice!!!! #WorldCup</td>
<td>freeking Chile is good !!! #WorldCup</td>
<td>damn Chile is good !!! #WorldCup</td>
</tr>
<tr>
<td>5</td>
<td>a bit boring ...</td>
<td>a little boring ...</td>
<td>some boring ...</td>
</tr>
<tr>
<td>6</td>
<td>im with Germany</td>
<td>I stand to Deutschlands side</td>
<td>I stand to Germany’s side</td>
</tr>
<tr>
<td>7</td>
<td>as getting I, GO CHILE!</td>
<td>completely mache I it GO CHILE!</td>
<td>as getting I, GO CHILE!</td>
</tr>
</tbody>
</table>

Comparison of translations by sentiment translation models and Baseline model
<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation model</th>
<th>Baseline model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Howard Web is a terrible ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
<td>Howard Web is a schrecklicher ref #WorldCup</td>
</tr>
<tr>
<td>2</td>
<td>injured Neymar out of World Cup 2014</td>
<td>verletzter Neymar out the WC2014</td>
<td>verletzter Neymar out of World Cup 2014</td>
</tr>
<tr>
<td>3</td>
<td>penalty shootouts are too intense !</td>
<td>penalty shoot is to intensiv !</td>
<td>penalties is to intensiv !</td>
</tr>
<tr>
<td>4</td>
<td>damn chile is nice !!!! #WorldCup</td>
<td>freeking Chile is good !!! #WorldCup</td>
<td>damn Chile is good !!! #WorldCup</td>
</tr>
<tr>
<td>5</td>
<td>a bit boring ...</td>
<td>a little boring ...</td>
<td>some boring ...</td>
</tr>
<tr>
<td>6</td>
<td>im with Germany</td>
<td>I stand to Germany’s side</td>
<td>I stand to Deutschlands side</td>
</tr>
<tr>
<td>7</td>
<td>as getting I, GO CHILE !</td>
<td>completely mache I it GO CHILE !</td>
<td>as getting I, GO CHILE !</td>
</tr>
</tbody>
</table>

Comparison of translations by sentiment translation models and Baseline model
Examples

<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation system</th>
<th>Baseline system</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bosnia and Herzegovina really got f*** over man</td>
<td>Bosnia and Herzegovina eliminated echt demolished</td>
<td>Bosnia and Herzegovina were a abgezogen</td>
</tr>
<tr>
<td>2</td>
<td>when USA lost, but were still moving onto the next round</td>
<td>even if USA today we in the next round</td>
<td>could usa loses the next round</td>
</tr>
<tr>
<td>3</td>
<td>Brazil 5 WorldCup championship Argentina 2 WorldCup championship so Ill go with Brazil</td>
<td>Brazil 5 time world champion Argentina 2 time world champion so Im for Brazil</td>
<td>Brazil 5 time world champions Argentina 2 time world champions so for Brazil</td>
</tr>
</tbody>
</table>

*Examples where sentiment is altered by the Baseline system*
### Examples where sentiment is altered by the Baseline system

<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Sentiment translation system</th>
<th>Baseline system</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bosnia and Herzegovina really got f*** over man</td>
<td>Bosnia and Herzegovina eliminated echt demolished</td>
<td>Bosnia and Herzegovina were a abgezogen</td>
</tr>
<tr>
<td>2</td>
<td>when USA lost, but were still moving onto the next round</td>
<td>even if USA today we in the next round</td>
<td>could usa loses the next round</td>
</tr>
<tr>
<td>3</td>
<td>Brazil 5 WorldCup championship Argentina 2 WorldCup championship so Ill go with Brazil</td>
<td>Brazil 5 time world champion Argentina 2 time world champion so Im for Brazil</td>
<td>Brazil 5 time world champions Argentina 2 time world champions so for Brazil</td>
</tr>
</tbody>
</table>
Comparison between sentiment polarities using the right and wrong MT engine

<table>
<thead>
<tr>
<th>Example</th>
<th>Reference</th>
<th>Right MT engine</th>
<th>Wrong MT engine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>little break on the #WorldCup for an amazing #Wimbledon final!</td>
<td>small Pause from the #WorldCup for a amazing #Wimbledon final!</td>
<td>kleine Pause of the #WorldCup for a erstaunliches #Wimbledon final!</td>
</tr>
<tr>
<td>2</td>
<td>yes !!!!!</td>
<td>yes !!!!!</td>
<td>so !!!!!</td>
</tr>
<tr>
<td>3</td>
<td>a bit boring ...</td>
<td>a little boring ...</td>
<td>some was ...</td>
</tr>
</tbody>
</table>
Discussion

- MT scores are better when no sentiment classification is used
- Sentiment classification approach performs better than the systems where it is switched off
## Discussion

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>✓</td>
<td>48.2</td>
<td>72.66% (+6%)</td>
</tr>
<tr>
<td>Twitter (Baseline)</td>
<td>×</td>
<td>50.3 (+2.1)</td>
<td>66.66%</td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>✓</td>
<td>48.5</td>
<td>71.33% (+8.67%)</td>
</tr>
<tr>
<td>Twitter + Flickr</td>
<td>×</td>
<td>50.7 (+2.2)</td>
<td>62.66%</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>✓</td>
<td>50.3</td>
<td>75.33% (+2%)</td>
</tr>
<tr>
<td>Twitter + Flickr + News</td>
<td>×</td>
<td>52.0 (+1.7)</td>
<td>73.33%</td>
</tr>
</tbody>
</table>

MT quality VS sentiment preservation
Discussion

- In most cases, the Baseline system produces better outputs in terms of BLEU score
- In some cases, interestingly, sentiment classification approach produces better outputs
In most cases, the Baseline system produces better outputs in terms of BLEU score.

In some cases, interestingly, sentiment classification approach produces better outputs.

Using specific-sentimented MT model to translate a text with the same sentiment is better in both ways.

<table>
<thead>
<tr>
<th>Translation model</th>
<th>Sentiment Classification</th>
<th>BLEU</th>
<th>Sentiment Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter (Right MT engine)</td>
<td>√</td>
<td>48.2 (+1.3)</td>
<td>72.66% (+25%)</td>
</tr>
<tr>
<td>Twitter (Wrong MT engine)</td>
<td>√</td>
<td>46.9</td>
<td>47.33%</td>
</tr>
</tbody>
</table>

MT quality VS sentiment preservation
Conclusions

- Despite a small deterioration in translation quality, our approach significantly improves sentiment preservation.

- It is essential to carefully select the proper MT engine conveying the same sentiment polarity as that of the UGC.
Future work

➢ To apply to other language pairs and also other forms of UGC such as customer feedback, blogs etc.

➢ Further refine the sentiment classes (strong positive, strong negative etc.,) in order to build more specific translation models
Thank you

pintu.lohar@adaptcentre.ie