“Best Dinner Ever!!”: Automatic Generation of Restaurant Reviews with LSTM-RNN

Alberto Bartoli    Andrea De Lorenzo
Eric Medvet    Dennis Morello    Fabiano Tarlao

Department of Engineering and Architecture
University of Trieste
Italy

October 16th, 2016

http://machinelearning.inginf.units.it
# Table of Contents

1. Motivation

2. The tool

3. Experimental evaluation
Product/service commerce

- People buy products/services online/offline
- When choosing seller, they trust other people’s opinion (*reviews*)

![Image of survey on checking online reviews before buying]

*Do you check reviews online before buying?*

- **YES**: 90%
- **NO**: 10%
Product/service commerce

- People buy products/services online/offline
- When choosing seller, they trust other people’s opinion (reviews)

A malicious seller might want to manipulate the choice (opinion spamming)
- Fabricating positive reviews for its products
- Fabricating negative reviews for competitors products
Review fabrication

Can be done “manually”:

“$100–$400” to “write and post a total or 10 reviews”, among which “5 good reviews about our hotel” and “5 very bad reviews about another hotel”!
Motivation

Review fabrication: the next level

Can be done automatically by a tool?
- much cheaper (≈free) for the single malicious seller
- much larger problem for the online retailer (Amazon, TripAdvisor, ...)
- (maybe) harder problem for opinion spamming researchers
Motivation

Review fabrication: the next level

Can be done automatically by a tool?

- much cheaper ($\approx$ free) for the single malicious seller
- much larger problem for the online retailer (Amazon, TripAdvisor, ...)
- (maybe) harder problem for opinion spamming researchers

Is that tool feasible?
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Motivation</td>
</tr>
<tr>
<td>2</td>
<td>The tool</td>
</tr>
<tr>
<td>3</td>
<td>Experimental evaluation</td>
</tr>
</tbody>
</table>
Problem statement

$r$ should:
- appear as generated by humans
- appear specific for restaurant
- express an overall rating $s$ (from ★ to ★★★★★★★)
Input: what’s a restaurant?

A set $C$ of categories: e.g., Italian, Cafe, International, Mediterranean
Method overview

Given an input $C, s$: 

1. generate many "human-like" reviews (NLG w/ LSTM-RNN)
2. discard those not consistent with categories ($C$, many binary classifiers)
3. discard those not consistent with ratings ($s$, one multiclass classifier)
4. select randomly one review $r$ among remaining reviews
Method overview

Given an input $C, s$:

1. generate many “human-like” reviews (NLG w/ LSTM-RNN)
Given an input $C, s$:

1. generate many “human-like” reviews (NLG w/ LSTM-RNN)
2. discard those not consistent with categories $C$ (many binary classifiers)
Method overview

Given an input $C, s$:

1. generate many “human-like” reviews (NLG w/ LSTM-RNN)
2. discard those not consistent with categories $C$ (many binary classifiers)
3. discard those not consistent with rating $s$ (one multiclass classifiers)
Method overview

Given an input $C, s$:

1. generate many “human-like” reviews (NLG w/ LSTM-RNN)
2. discard those not consistent with categories $C$ (many binary classifiers)
3. discard those not consistent with rating $s$ (one multiclass classifiers)
4. select randomly one review $r$ among remaining reviews
Generating human-like reviews

Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN)

- works at the character level
- char-rnn library with default settings (3 layers of 700 neurons)
- trained on a corpus of 500000 reviews ($\approx 1$ month)
- when generating, seed is a random sentence of a real review
- first generated review is discarded (influence of the seed)
 Experimental evaluation

Table of Contents

1 Motivation

2 The tool

3 Experimental evaluation
Aims

- Is an artificial review considered genuine? \((\textit{intrinsic evaluation})\)
- Can an artificial review influence the human subject? \((\textit{extrinsic evaluation})\)

Extrinsic performed first; 39 subjects involved, 3–4 forms each
Extrinsic evaluation

Simulates the restaurant choice by a user:
- each form with 3 reviews
- at least 1 artificial and 1 genuine

<table>
<thead>
<tr>
<th>Uncle Sam’s Meat &amp; Wine</th>
</tr>
</thead>
<tbody>
<tr>
<td>american, steakhouse</td>
</tr>
</tbody>
</table>

Review with ★★★★★
The atmosphere was very cozy. With small seating areas the noise is minimized. The service was good. […]
Useful? Y □ N □

Review with ★
This place is dimly lit and reminded me of a bad prom decorating. The waitress was nice, but a little over […]
Useful? Y □ N □

Review with ★★★★★
Great food and even better atmosphere. It is a quiet darker setting with no windows. The service […]
Useful? Y □ N □

Would you go to this restaurant? Y □ N □
Extrinsic evaluation

Simulates the restaurant choice by a user:

- each form with 3 reviews
- at least 1 artificial and 1 genuine

<table>
<thead>
<tr>
<th>Genuine</th>
<th>Artificial</th>
<th>Going [%]</th>
<th>Not going [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥★★★★★</td>
<td>≥★★★★★</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>≥★★★★★</td>
<td>≤★★★</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>≤★★★</td>
<td>≥★★★★★</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>≤★★★</td>
<td>≤★★★</td>
<td>23</td>
<td>77</td>
</tr>
</tbody>
</table>

Uncle Sam’s Meat & Wine
American, steakhouse

Review with ★★★★★
The atmosphere was very cozy. With small seating areas the noise is minimized. The service was good. […]
Useful? Y □ N □

Review with ★
This place is dimly lit and reminded me of a bad prom decorating. The waitress was nice, but a little over […]
Useful? Y □ N □

Review with ★★★★★
Great food and even better atmosphere. It is a quiet darker setting with no windows. The service […]
Useful? Y □ N □

Would you go to this restaurant? Y □ N □
Extrinsic evaluation

Simulates the restaurant choice by a user:
- each form with 3 reviews
- at least 1 artificial and 1 genuine

<table>
<thead>
<tr>
<th>Genuine</th>
<th>Artificial</th>
<th>Going [%]</th>
<th>Not going [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>⭐⭐⭐⭐⭐</td>
<td>⭐⭐⭐⭐⭐</td>
<td>23</td>
<td>77</td>
</tr>
</tbody>
</table>

Uncle Sam’s Meat & Wine
american, steakhouse

Review with ⭐⭐⭐⭐⭐
The atmosphere was very cozy. With small seating areas the noise is minimized. The service was good. [...] Use? Y □ N □

Review with ⭐⭐
This place is dimly lit and reminded me of a bad prom decorating. The waitress was nice, but a little over [...] Use? Y □ N □

Review with ⭐⭐⭐⭐⭐
Great food and even better atmosphere. It is a quiet darker setting with no windows. The service [...] Use? Y □ N □

Would you go to this restaurant? Y □ N □
Extrinsic evaluation

Simulates the restaurant choice by a user:

- each form with 3 reviews
- at least 1 artificial and 1 genuine

<table>
<thead>
<tr>
<th>Genuine</th>
<th>Artificial</th>
<th>Going [%]</th>
<th>Not going [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥★★★★</td>
<td>≥★★★★</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>≥★★★★</td>
<td>≤★★★</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>≤★★★</td>
<td>≥★★★★</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>≤★★★</td>
<td>≤★★★</td>
<td>23</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Useful [%]</th>
<th>Not useful [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genuine</td>
<td>80</td>
</tr>
<tr>
<td>Artificial</td>
<td>29</td>
</tr>
</tbody>
</table>

Uncle Sam’s Meat & Wine
american, steakhouse

Review with ★★★★★
The atmosphere was very cozy. With small seating areas the noise is minimized. The service was good. [...] Useful? Y □ N □

Review with ★
This place is dimly lit and reminded me of a bad prom decorating. The waitress was nice, but a little over [...] Useful? Y □ N □

Review with ★★★★★
Great food and even better atmosphere. It is a quiet darker setting with no windows. The service [...] Useful? Y □ N □

Would you go to this restaurant? Y □ N □
Intrinsic evaluation

Has this review been written by a human for this restaurant?
- forms with 5 reviews for each restaurant, 4 forms per user
- 4 kinds of reviews
  - $R_{gs}$ genuine for specific restaurant
  - $R_{gd}$ genuine for different restaurant
  - $R_{ad}$ artificial for specific restaurant
  - $R_{ad}$ artificial for different restaurant (no step 2)
Intrinsic evaluation

Has this review been written by a human for this restaurant?

- forms with 5 reviews for each restaurant, 4 forms per user
- 4 kinds of reviews
  - $R_{gs}$ genuine for specific restaurant
  - $R_{gd}$ genuine for different restaurant
  - $R_{as}$ artificial for specific restaurant
  - $R_{ad}$ artificial for different restaurant (no step 2)

<table>
<thead>
<tr>
<th></th>
<th>Yes [%]</th>
<th>No [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{gs}$</td>
<td>81</td>
<td>19</td>
</tr>
<tr>
<td>$R_{gd}$</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>$R_{as}$</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>$R_{ad}$</td>
<td>24</td>
<td>76</td>
</tr>
</tbody>
</table>
Automatic Generation of Restaurant Review:

- “Is that tool feasible?” Yes! (we did it!)

Machine generated reviews might become a real threat for (e-)commerce!
Conclusions

Automatic Generation of Restaurant Review:
- “Is that tool feasible?” Yes! (we did it!)
- “Is an artificial review considered genuine?” ≈ 1 on 4
- “Can an artificial review influence the human subject?” Unclear, deeper experiments needed
Automatic Generation of Restaurant Review:

- “Is that tool feasible?” Yes! (we did it!)
- “Is an artificial review considered genuine?” $\approx 1$ on 4
- “Can an artificial review influence the human subject?” Unclear, deeper experiments needed

Machine generated reviews might become a real threat for (e-)commerce!
Thanks!