May All Your Wishes Come True: A Study of Wishes and How to Recognize Them

Andrew B. Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson & Xiaojin Zhu

Computer Sciences Department
University of Wisconsin-Madison
Times Square  
Virtual Wishing Well

- In December 2007, Web users sent in their wishes for the new year
- Wishes were printed on confetti
- Released from the sky at midnight in sync with the famous “ball drop”
- Over 100,000 wishes collected to form the WISH corpus
## Sample New Year’s Wishes

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Wish</th>
</tr>
</thead>
<tbody>
<tr>
<td>514</td>
<td>peace on earth</td>
</tr>
<tr>
<td>351</td>
<td>peace</td>
</tr>
<tr>
<td>331</td>
<td>world peace</td>
</tr>
<tr>
<td>244</td>
<td>happy new year</td>
</tr>
<tr>
<td>112</td>
<td>love</td>
</tr>
<tr>
<td>76</td>
<td>health and happiness</td>
</tr>
<tr>
<td>75</td>
<td>to be happy</td>
</tr>
<tr>
<td>51</td>
<td>i wish for world peace</td>
</tr>
<tr>
<td>21</td>
<td>i wish for health and happiness</td>
</tr>
<tr>
<td>21</td>
<td>let there be peace on earth</td>
</tr>
<tr>
<td>16</td>
<td>to find my true love</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Wish</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>i wish for a puppy</td>
</tr>
<tr>
<td>7</td>
<td>for the war in iraq to end</td>
</tr>
<tr>
<td>6</td>
<td>peace on earth please</td>
</tr>
<tr>
<td>5</td>
<td>a free democratic venezuela</td>
</tr>
<tr>
<td>5</td>
<td>may the best of 2007 be the worst of 2008</td>
</tr>
<tr>
<td>5</td>
<td>to be financially stable</td>
</tr>
<tr>
<td>1</td>
<td>a little goodness for everyone would be nice</td>
</tr>
<tr>
<td>1</td>
<td>i hope i get accepted into a college that i like</td>
</tr>
<tr>
<td>1</td>
<td>i wish to get more sex in 2008</td>
</tr>
<tr>
<td>1</td>
<td>please let name be healthy and live all year</td>
</tr>
<tr>
<td>1</td>
<td>to be emotionally stable and happy</td>
</tr>
</tbody>
</table>
What is a wish?
What is a wish?

Formally:

wish (n.) “a desire or hope for something to happen”
What is a wish?

Formally:

wish (n.) “a desire or hope for something to happen”

• Open questions in NLP:
  • How are wishes expressed?
  • How can wishful expressions be automatically recognized?
What is a wish?

Formally:

_wish (n.) “a desire or hope for something to happen”_

- Open questions in NLP:
  - How are wishes expressed?
  - How can wishful expressions be automatically recognized?
- Our work:
  - Analyze this unique new collection of wishes
  - Leverage the WISH corpus to build general “wish detectors”
  - Demonstrate effectiveness on consumer product reviews and informal political discussion online
Outline
Outline

• Why study wishes? (relation to prior work)
  • Sentiment analysis
  • Psychology / cognitive science
Outline

• Why study wishes? (relation to prior work)
  • Sentiment analysis
  • Psychology / cognitive science

• Analysis of the WISH corpus
  • Topic and scope of wishes
  • Geographical differences
  • Latent topic modeling
Outline

• Why study wishes? (relation to prior work)
  • Sentiment analysis
  • Psychology / cognitive science

• Analysis of the WISH corpus
  • Topic and scope of wishes
  • Geographical differences
  • Latent topic modeling

• Building wish detectors
  • Key contribution: Automatically discovering wish templates
Outline

• Why study wishes? (relation to prior work)
  • Sentiment analysis
  • Psychology / cognitive science

• Analysis of the WISH corpus
  • Topic and scope of wishes
  • Geographical differences
  • Latent topic modeling

• Building wish detectors
  • Key contribution: Automatically discovering wish templates

• Experimental results
Why study wishes?
Why study wishes?

• Wishes add a novel dimension to sentiment analysis, opinion mining
  • What people explicitly want, not just what they like or dislike

>“Great camera. Indoor shots with a flash are not quite as good as 35mm. I wish the camera had a higher optical zoom so that I could take even better wildlife photos.”

• Automatic “wish detector” can provide political value & business intelligence
Why study wishes?

• Wishes add a novel dimension to sentiment analysis, opinion mining
  • What people explicitly want, not just what they like or dislike

  “Great camera. Indoor shots with a flash are not quite as good as 35mm. I wish the camera had a higher optical zoom so that I could take even better wildlife photos.”

• Automatic “wish detector” can provide political value & business intelligence

• Wishes can reveal a lot about people
  • Psychologists have studied wish content vs. location, gender, age, etc (Speer 1939, Milgram and Riedel 1969, Ehrlichman and Eichenstein 1992, King and Broyles 1997)
  • WISH corpus: much larger scale, from the entire globe
The WISH corpus
Analysis of the WISH corpus
Analysis of the WISH corpus

- Almost 100,000 wishes collected over 10 days in December 2007
  - We focus on the 89,574 wishes written in English
  - Remaining 10,000+ in Portuguese, Spanish, Chinese, French, etc
Analysis of the WISH corpus

- Almost 100,000 wishes collected over 10 days in December 2007
  - We focus on the 89,574 wishes written in English
  - Remaining 10,000+ in Portuguese, Spanish, Chinese, French, etc
- Many contain optional state/country location entered by the wisher
Analysis of the WISH corpus

- Almost 100,000 wishes collected over 10 days in December 2007
  - We focus on the 89,574 wishes written in English
  - Remaining 10,000+ in Portuguese, Spanish, Chinese, French, etc
- Many contain optional state/country location entered by the wisher
- Minimal preprocessing
  - TreeBank tokenization, downcasing, punctuation removal
Analysis of the WISH corpus

• Almost 100,000 wishes collected over 10 days in December 2007
  • We focus on the 89,574 wishes written in English
  • Remaining 10,000+ in Portuguese, Spanish, Chinese, French, etc
• Many contain optional state/country location entered by the wisher
• Minimal preprocessing
  • TreeBank tokenization, downcasing, punctuation removal
• Each wish is treated as a single entity (even if multiple sentences)
Analysis of the WISH corpus

- Almost 100,000 wishes collected over 10 days in December 2007
  - We focus on the 89,574 wishes written in English
  - Remaining 10,000+ in Portuguese, Spanish, Chinese, French, etc
- Many contain optional state/country location entered by the wisher
- Minimal preprocessing
  - TreeBank tokenization, downcasing, punctuation removal
- Each wish is treated as a single entity (even if multiple sentences)
- Average length of wishes is 8 tokens
WISH corpus: Scope and topic of wishes

Manually annotated random subsample of 5,000 wishes
WISH corpus: Scope and topic of wishes

Manually annotated random subsample of 5,000 wishes

Topic of wishes: what the wish is about

- career (4.6%)
- sex (4.5%)
- religion (3.4%)
- money (7.3%)
- political (2.8%)
- travel (1.4%)
- peace (12%)
- love (18%)
- happiness (15%)
- health (15%)
- other (16%)
WISH corpus: Scope and topic of wishes

Manually annotated random subsample of 5,000 wishes

Topic of wishes: what the wish is about

- career (4.6%)
- sex (4.5%)
- religion (3.4%)
- money (7.3%)
- political (2.8%)
- travel (1.4%)
- love (18%)
- happiness (15%)
- health (15%)
- peace (12%)
- other (16%)

individual requests: “I wish for a new puppy”
solicitations: “call me 555-1234”, “visit website.com”
sinister: “to take over the world”
WISH corpus: Scope and topic of wishes

Manually annotated random subsample of 5,000 wishes

Topic of wishes: what the wish is about
- career (4.6%)
- sex (4.5%)
- religion (3.4%)
- money (7.3%)
- peace (12%)
- political (2.8%)
- travel (1.4%)
- love (18%)
- health (15%)
- happiness (15%)
- other (16%)

Scope of wishes: who the wish is aimed at
- family (14%)
- name (14%)
- other (7.8%)
- country (5.8%)
- world (23%)
- self (36%)

individual requests: “I wish for a new puppy”
solicitations: “call me 555-1234”, “visit website.com”
sinister: “to take over the world”
WISH corpus: Geographical differences
WISH corpus: Geographical differences

• About 4,000 of the manually annotated wishes included valid location information
  • Covered all 50 U.S. states and all continents except Antarctica
WISH corpus: Geographical differences

- About 4,000 of the manually annotated wishes included valid location information
  - Covered all 50 U.S. states and all continents except Antarctica
- We compared topic and scope distributions between U.S. and non-U.S. wishes
WISH corpus: Geographical differences

- About 4,000 of the manually annotated wishes included valid location information
  - Covered all 50 U.S. states and all continents except Antarctica
- We compared topic and scope distributions between U.S. and non-U.S. wishes
WISH corpus: Geographical differences

- About 4,000 of the manually annotated wishes included valid location information
  - Covered all 50 U.S. states and all continents except Antarctica
- We compared topic and scope distributions between U.S. and non-U.S. wishes
WISH corpus: Geographical differences

- About 4,000 of the manually annotated wishes included valid location information
  - Covered all 50 U.S. states and all continents except Antarctica
- We compared topic and scope distributions between U.S. and non-U.S. wishes
- Statistically significant differences in both cases (Pearson $X^2$-test, $p < 0.01$)
WISH corpus: Geographical differences

• About 4,000 of the manually annotated wishes included valid location information
  • Covered all 50 U.S. states and all continents except Antarctica
• We compared topic and scope distributions between U.S. and non-U.S. wishes
• Statistically significant differences in both cases (Pearson X²-test, p < 0.01)
• But no significant difference between red vs. blue states
WISH corpus: Latent topic modeling
WISH corpus: Latent topic modeling

• So far analysis was of 5,000 manually labeled wishes
WISH corpus: Latent topic modeling

- So far analysis was of 5,000 manually labeled wishes
- We automatically analyzed all ~90,000 using Latent Dirichlet Allocation
  - Each wish is treated as a short document
  - 12 topics
  - Inference performed by collapsed Gibbs sampling
  - Hyperparameters set to $\alpha=0.5$, $\beta=0.1$
WISH corpus: Latent topic modeling

| Topic | Top words, sorted by $p(\text{word}|\text{topic})$ | Subjective Label |
|-------|-----------------------------------------------|------------------|
| 1     | year, new, happy, 2008, best, everyone, great, wishing, hope | New Year         |
| 2     | all, god, home, come, safe, us, bless, troops, bring, iraq, return | Troops           |
| 3     | end, no, more, 2008, war, president, paul, ron, less, bush, vote | Election         |
| 4     | more, better, life, one, live, time, make, people, than, day, every | Life             |
| 5     | health, happiness, good, family, friends, prosperity, wealth, success | Prosperity       |
| 6     | love, find, true, life, meet, want, man, marry, someone, boyfriend | Love             |
| 7     | get, job, out, hope, school, better, house, well, back, college | Career           |
| 8     | win, 2008, money, want, make, become, lottery, more, great, lots | Money            |
| 9     | peace, world, love, earth, happiness, everyone, joy, 2008, around | Peace            |
| 10    | love, forever, jesus, know, together, u, always, best, mom, christ | Religion         |
| 11    | healthy, family, baby, life, children, safe, husband, stay, marriage | Family           |
| 12    | me, lose, please, let, cancer, weight, cure, mom, mother, visit, dad | Health           |
Building wish detectors
Building wish detectors

**Novel NLP task: Wish Detection**
Given sentence $S$, classify $S$ as wish or non-wish
Building wish detectors

**Novel NLP task:** Wish Detection
Given sentence $S$, classify $S$ as *wish* or *non-wish*

- Want an approach that will extend beyond New Year’s wishes
  - Target domains: product reviews, political discussions
Building wish detectors

**Novel NLP task: Wish Detection**
Given sentence $S$, classify $S$ as *wish* or *non-wish*

- Want an approach that will extend beyond New Year’s wishes
  - Target domains: product reviews, political discussions
- Wishes are highly domain dependent
  - New Year’s eve: “I wish for world peace”
  - Product review: “I want to have instant access to the volume”
Building wish detectors

**Novel NLP task: Wish Detection**
Given sentence $S$, classify $S$ as *wish* or *non-wish*

- Want an approach that will extend beyond New Year’s wishes
  - Target domains: product reviews, political discussions
- Wishes are highly domain dependent
  - New Year’s eve: “I wish for world peace”
  - Product review: “I want to have instant access to the volume”
- Initial study
  - Assume some labeled data in target domains
  - Try to beat some standard baselines by exploiting the WISH corpus to learn patterns of wish expressions (wish templates)
Two simple baseline wish detectors

- Do not use WISH corpus
Two simple baseline wish detectors

- Do not use WISH corpus

**Manual**

- Rule-based classifier
- If part of a sentence matches a template, classify it as a wish
- Some of the 13 templates created by two native English speakers:
  - i wish __ if only __
  - i hope __ would be better if __
  - i want __ would like if __
  - hopefully __ should __

*Expect high precision, low recall*
Two simple baseline wish detectors

- Do not use WISH corpus

<table>
<thead>
<tr>
<th>Manual</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule-based classifier</td>
<td>Linear Support Vector Machine</td>
</tr>
<tr>
<td>If part of a sentence matches a template, classify it as a wish</td>
<td>Train on labeled training set from the target domain</td>
</tr>
<tr>
<td>Some of the 13 templates created by two native English speakers:</td>
<td>Representation:</td>
</tr>
<tr>
<td>i wish __ if only __</td>
<td>- binary word-indicator vector</td>
</tr>
<tr>
<td>i hope __ would be better if __</td>
<td>- normalized to sum to 1</td>
</tr>
<tr>
<td>i want __ would like if __</td>
<td>- Natural first baseline for a new text classification task</td>
</tr>
<tr>
<td>hopefully __ should __</td>
<td>Expect high recall, low precision</td>
</tr>
</tbody>
</table>

Expect high precision, low recall

Expect high recall, low precision
Learning wish templates

Key idea: Exploit redundancy in how wishes are expressed
Learning wish templates

Key idea: Exploit redundancy in how wishes are expressed

Many entries in the WISH corpus contain only a short “wish content”

  world peace     health and happiness
Learning wish templates

Key idea: Exploit redundancy in how wishes are expressed

Many entries in the WISH corpus contain only a short “wish content”

world peace          health and happiness

These “wish contents” appear within longer wishes with a common prefix/suffix:

  i wish for world peace          i wish for health and happiness
Learning wish templates

Key idea: Exploit redundancy in how wishes are expressed

Many entries in the WISH corpus contain only a short “wish content”
world peace  health and happiness

These “wish contents” appear within longer wishes with a common prefix/suffix:

i wish for world peace  i wish for health and happiness

Intuitively, popular content appears within popular templates.
Learning wish templates

Key idea: Exploit redundancy in how wishes are expressed

Many entries in the WISH corpus contain only a short “wish content”

world peace  health and happiness

These “wish contents” appear within longer wishes with a common prefix/suffix:

i wish for world peace  i wish for health and happiness

Intuitively, popular content appears within popular templates.

Can discover non-obvious templates, too:

world peace, peace on earth  →  let there be ___
become rich, win the lottery  →  to finally ___
get a job, save the environment  →  ___ please
Learning wish templates

Formally, we build a bipartite graph
Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right
Two kinds of edges: 
• $c \rightarrow t$ (weighted by # times content appears in the template)
• $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph.

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right.

Two kinds of edges:

- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes \( c \in C \) on left, Template nodes \( t \in T \) on right

Two kinds of edges:

- \( c \rightarrow t \) (weighted by # times content appears in the template)
- \( t \rightarrow c \) (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right

Two kinds of edges:

- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)

Content

world peace

Template

i wish for ___

$C_1 \xrightarrow{\text{#("i wish for world peace")}} t_1$
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right

Two kinds of edges:

- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right

Two kinds of edges:
- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph.

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right.

Two kinds of edges:
- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right

Two kinds of edges:
- $c \rightarrow t$ (weighted by # times content appears in the template)
- $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph

Two kinds of nodes: Content nodes \( c \in C \) on left, Template nodes \( t \in T \) on right

Two kinds of edges:

- \( c \rightarrow t \) (weighted by \# times content appears in the template)
- \( t \rightarrow c \) (weighted by \# times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph
Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right
Two kinds of edges:  
• $c \rightarrow t$ (weighted by # times content appears in the template)  
• $t \rightarrow c$ (weighted by # times template matches a content node)
Learning wish templates

Formally, we build a bipartite graph
Two kinds of nodes: Content nodes $c \in C$ on left, Template nodes $t \in T$ on right
Two kinds of edges:  
• $c \rightarrow t$ (weighted by # times content appears in the template)
• $t \rightarrow c$ (weighted by # times template matches a content node)
Ranking template nodes

world peace

health and happiness

health

C1

#(“i wish for world peace”)

C2

#(“i wish for health and ...”)

C3

#(“i wish for health”)

#(“health and happiness”)

#(“health and happiness”)

#(“health and happiness”)

i wish for ___

___ and happiness
Ranking template nodes

• Useful templates match many complete wishes but few content-only wishes
Ranking template nodes

• Useful templates match many complete wishes but few content-only wishes
• We rank all template nodes $t$ by $score(t) = in(t) - out(t)$
Ranking template nodes

- Useful templates match many complete wishes but few content-only wishes
- We rank all template nodes \( t \) by \( \text{score}(t) = \text{in}(t) - \text{out}(t) \)
- Subtracting the out-degree eliminates “bad” templates that contain specific topical content (e.g., “___ and happiness”)
Ranking template nodes

• Useful templates match many complete wishes but few content-only wishes
• We rank all template nodes $t$ by $score(t) = \text{in}(t) - \text{out}(t)$
• Subtracting the out-degree eliminates “bad” templates that contain specific topical content (e.g., “___ and happiness”)
• Apply threshold $score(t) \geq 5$ to obtain 811 top templates for use as features
## Wish template features

Some of the top 811 template features selected by our algorithm

<table>
<thead>
<tr>
<th>Top 10</th>
<th>Others in Top 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>___ in 2008</td>
<td>i want to ___</td>
</tr>
<tr>
<td>i wish for ___</td>
<td>___ for everyone</td>
</tr>
<tr>
<td>i wish ___</td>
<td>i hope ___</td>
</tr>
<tr>
<td>i want ___</td>
<td>my wish is ___</td>
</tr>
<tr>
<td>i want my ___</td>
<td>___ please</td>
</tr>
<tr>
<td>___ this year</td>
<td>wishing for ___</td>
</tr>
<tr>
<td>i wish ___ in 2008</td>
<td>may you ___</td>
</tr>
<tr>
<td>i wish to ___</td>
<td>i wish i had ___</td>
</tr>
<tr>
<td>i wish ___ this year</td>
<td>to finally ___</td>
</tr>
<tr>
<td>___ in the new year</td>
<td>for my family to have ___</td>
</tr>
</tbody>
</table>
Learning with wish template features
Learning with wish template features

• We use the templates as features for classification in target domains
Learning with wish template features

• We use the templates as features for classification in target domains
• Each template leads to 2 features depending on level of matching in sentence:
  • Whole-sentence match: “I wish this mp3 player had more storage”
  • Partial-sentence match: “most of all I wish this camera was smaller”
Learning with wish template features

- We use the templates as features for classification in target domains.
- Each template leads to 2 features depending on level of matching in sentence:
  - Whole-sentence match: “i wish this mp3 player had more storage”
  - Partial-sentence match: “most of all i wish this camera was smaller”

Models using templates:
- [Templates] uses only these features in a linear SVM
- [Words+Templates] combines unigram and template features in a linear SVM
Test corpora
Test corpora

• Recall goal of discovering wishes in interesting text domains
Test corpora

• Recall goal of discovering wishes in interesting text domains

• Two test corpora, manually labeled sentences as wish vs. non-wish
Test corpora

• Recall goal of discovering wishes in interesting text domains

• Two test corpora, manually labeled sentences as wish vs. non-wish
  • Consumer product reviews
    • 1,235 sentences from amazon.com and cnet.com reviews (selected from data used in Hu and Liu, 2004; Ding et al., 2008)
    • 12% wishes
Test corpora

- Recall goal of discovering wishes in interesting text domains

- Two test corpora, manually labeled sentences as wish vs. non-wish
  - Consumer product reviews
    - 1,235 sentences from amazon.com and cnet.com reviews
      (selected from data used in Hu and Liu, 2004; Ding et al., 2008)
    - 12% wishes
  - Political discussion board postings
    - 6,379 sentences selected from politics.com (Mullen and Malouf, 2008).
    - 34% wishes
Test corpora

- Recall goal of discovering wishes in interesting text domains

- Two test corpora, manually labeled sentences as wish vs. non-wish
  - Consumer product reviews
    - 1,235 sentences from amazon.com and cnet.com reviews (selected from data used in Hu and Liu, 2004; Ding et al., 2008)
    - 12% wishes
  - Political discussion board postings
    - 6,379 sentences selected from politics.com (Mullen and Malouf, 2008).
    - 34% wishes

Download from http://pages.cs.wisc.edu/~goldberg/wish_data
Experimental results

10-fold cross validation, linear classifier (SVM$^\text{light}$ using default parameters)
Experimental results

10-fold cross validation, linear classifier ($\text{SVM}^{\text{light}}$ using default parameters)

Politics

Precision vs. Recall for different features:
- Manual
- Words
- Templates
- Words + Templates
Experimental results

10-fold cross validation, linear classifier (SVM\textsuperscript{light} using default parameters)

Politics

Products
Experimental results

10-fold cross validation, linear classifier ($\text{SVM}^{\text{light}}$ using default parameters)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Manual</th>
<th>Words</th>
<th>Templates</th>
<th>Words + Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>0.67 ± 0.03</td>
<td>0.77 ± 0.03</td>
<td>0.73 ± 0.03</td>
<td>0.80 ± 0.03</td>
</tr>
<tr>
<td>Products</td>
<td>0.49 ± 0.13</td>
<td>0.52 ± 0.16</td>
<td>0.47 ± 0.16</td>
<td>0.56 ± 0.16</td>
</tr>
</tbody>
</table>
What features are important?

Features with largest magnitude weights for one fold of the Products corpus

<table>
<thead>
<tr>
<th>Sign</th>
<th>Words</th>
<th>Templates</th>
<th>Words + Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>wish</td>
<td>i hope ___</td>
<td>hoping ___</td>
</tr>
<tr>
<td>+</td>
<td>hope</td>
<td>i wish ___</td>
<td>i hope ___</td>
</tr>
<tr>
<td>+</td>
<td>hopefully</td>
<td>hoping ___</td>
<td>i just want ___</td>
</tr>
<tr>
<td>+</td>
<td>hoping</td>
<td>i just want ___</td>
<td>i wish ___</td>
</tr>
<tr>
<td>+</td>
<td>want</td>
<td>i would like ___</td>
<td>i would like ___</td>
</tr>
<tr>
<td>-</td>
<td>money</td>
<td>family ___</td>
<td>micro</td>
</tr>
<tr>
<td>-</td>
<td>find</td>
<td>___ forever</td>
<td>about</td>
</tr>
<tr>
<td>-</td>
<td>digital</td>
<td>let me ___</td>
<td>fix</td>
</tr>
<tr>
<td>-</td>
<td>again</td>
<td>___ d</td>
<td>digital</td>
</tr>
<tr>
<td>-</td>
<td>you</td>
<td>___ for my dad</td>
<td>you</td>
</tr>
</tbody>
</table>
Conclusions & Future Work
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
• Introduced and analyzed the WISH corpus of ~90,000 wishes
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
• Introduced and analyzed the WISH corpus of ~90,000 wishes
• Proposed new wish detection task and simple baselines
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
• Introduced and analyzed the WISH corpus of ~90,000 wishes
• Proposed new wish detection task and simple baselines
• Generated wish templates that transfer to new domains
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
• Introduced and analyzed the WISH corpus of ~90,000 wishes
• Proposed new wish detection task and simple baselines
• Generated wish templates that transfer to new domains
• Built wish-annotated test corpora in product review and political domains
Conclusions & Future Work

• Studied wishes from an NLP perspective for the first time
• Introduced and analyzed the WISH corpus of ~90,000 wishes
• Proposed new wish detection task and simple baselines
• Generated wish templates that transfer to new domains
• Built wish-annotated test corpora in product review and political domains

• Much future work in wish detection remains:
  • Additional wish-sensitive features
  • Annotated training data is expensive ➔ semi-supervised learning
Acknowledgements

We’d like to thank:

Times Square Alliance for providing the WISH corpus
Wisconsin Alumni Research Foundation
Yahoo! Key Technical Challenges Program
&
you!

Download test corpora at
http://pages.cs.wisc.edu/~goldberg/wish_data