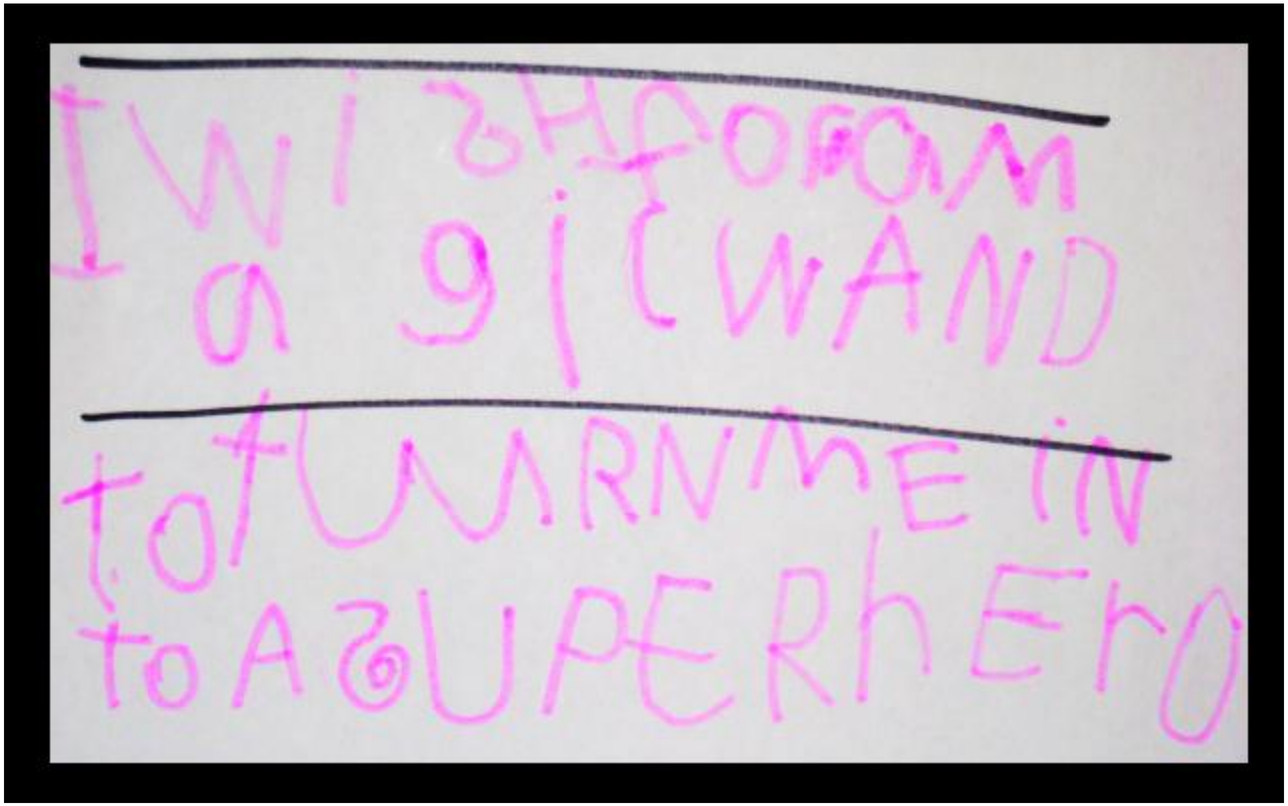
COMPUTERS DISCOVER WISHES AND CREATIVITY IN TEXT

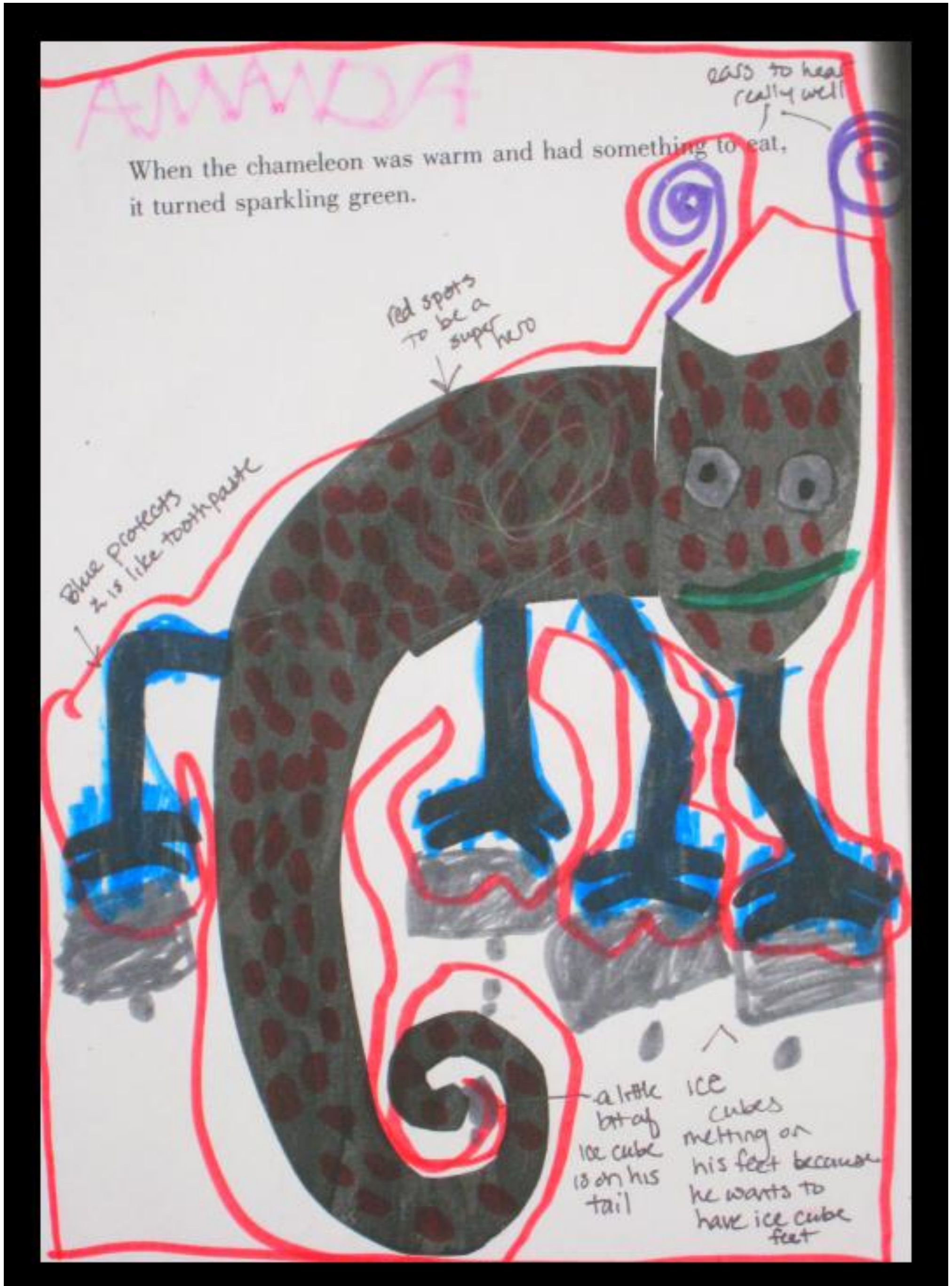
XIAOJIN ZHU

COLLABORATORS
David Andrzejewski, Nathanael Fillmore
Bryan Gibson, Andrew Goldberg
Tushar Khot, Zhiting Xu

DEPT. OF COMPUTER SCIENCES
UNIVERSITY OF WISCONSIN-MADISON



"I wish for a magic wand to turn me into a superhero"



"Blue protects & is like toothpaste"

"Ice cubes melting on his feet because he wants to have ice cube feet"

Machine learning in natural language processing

- Speech recognition
- Machine translation
- Information retrieval
- Text categorization
  - by topic (e.g., politics / sports), circa 1980
  - by content (e.g., spam filtering), circa 1990
  - by sentiment (e.g., thumbs up / thumbs down), circa 2000
  - this talk: more subjective frontiers (wish / not wish, creative / not creative)

Novel task 1

Identifying wishes in text

Why study wishes?

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

"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
    (Speer 1939, Milgram and Riedel 1969, Ehrlichman and Eichenstein 1992, King and Broyles 1997)

What is a wish?

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

- Open questions in NLP:
  - How are wishes expressed?
  - How can wishes be automatically recognized?
- Our work:
  - Analyze a unique collection of wishes
  - Build a general “wish detectors”

A unique WISH corpus

A unique wish corpus

Times Square
Virtual Wishing Wall

- 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

Virtual New Year's Wishing Wall 2009

Questions marked with an asterisk (*) are mandatory.

Share your hopes, dreams and resolutions for 2009 – then watch them flutter down as confetti in the heart of Times Square on New Year's Eve!

1 Name:

2 City:

3 Country:

4 * New Year's Wish (please note that wishes containing phone numbers, email addresses, websites or other contact information will be deleted):

5 Please add me to the Times Square Alliance e-newsletter list, so that I can be among the first to learn about Times Square news, events, deals and promotions.

YES NO

SUBMIT

Quiz

- What is the most frequent new year’s wish?

Sample New Year's wishes

Freq.	Wish
514	peace on earth
351	peace
331	world peace
244	happy new year
112	love
76	health and happiness
75	to be happy
51	i wish for world peace
21	i wish for health and happiness
21	let there be peace on earth
16	to find my true love

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

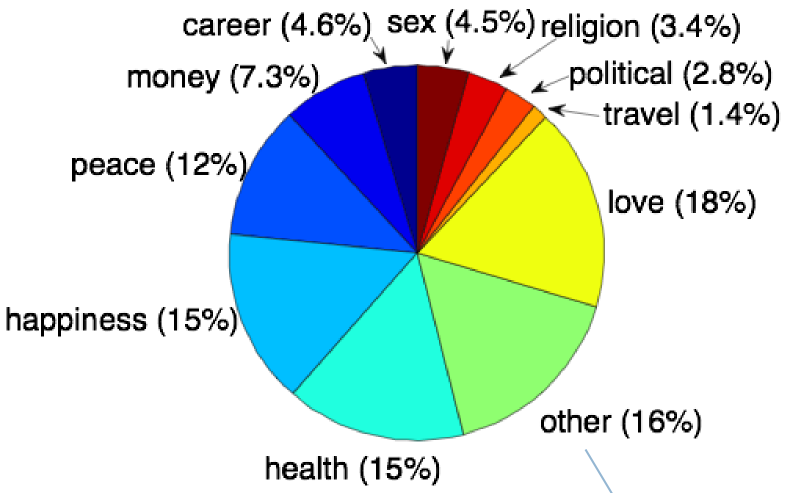
Analysis of the WISH corpus

- Almost 100,000 wishes collected over 10 days in Dec. 2007
  - 89,574 wishes written in English
  - Remaining 10,000+ in Portuguese, Spanish, Chinese, etc .
- Many contain optional location entered by the wisher
- Minimal preprocessing
  - TreeBank tokenization, downcasing, punctuation removal
- Each wish is treated as a single entity
- Average length of wishes is 8 tokens

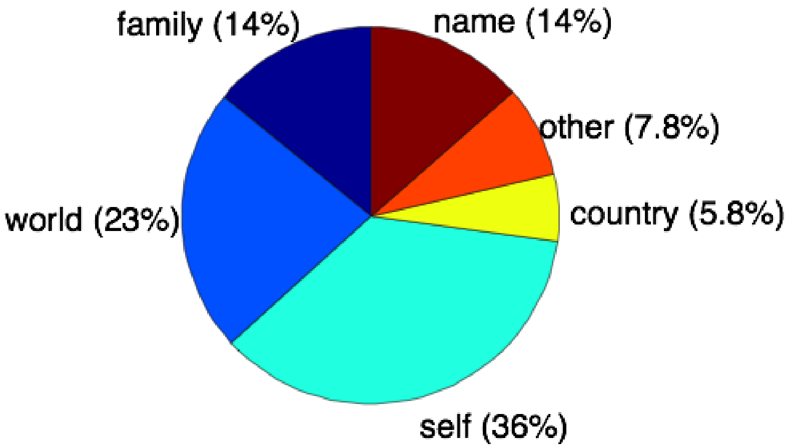
WISH corpus: Scope and topic

Manually annotated random subsample of 5,000 wishes

Topic of wishes:
what the wish is about



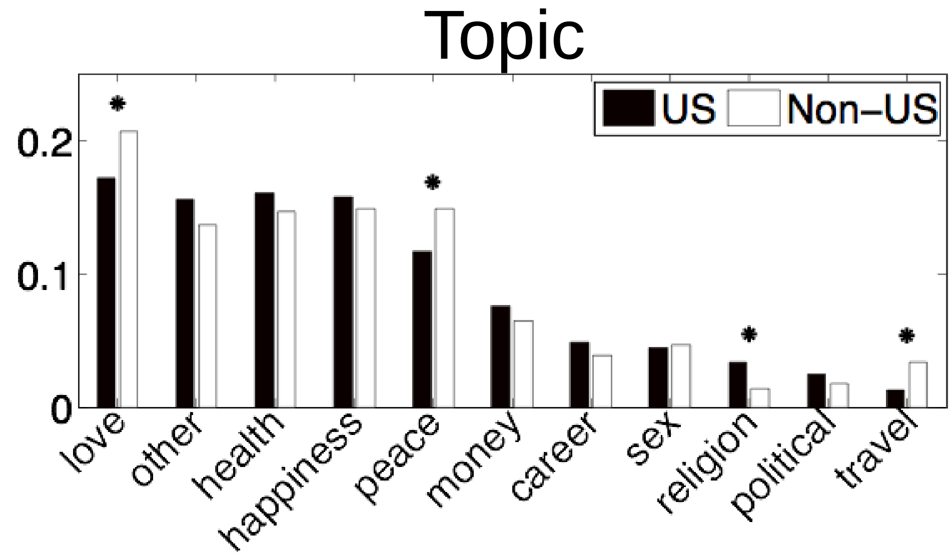
Scope of wishes:
who the wish is aimed at

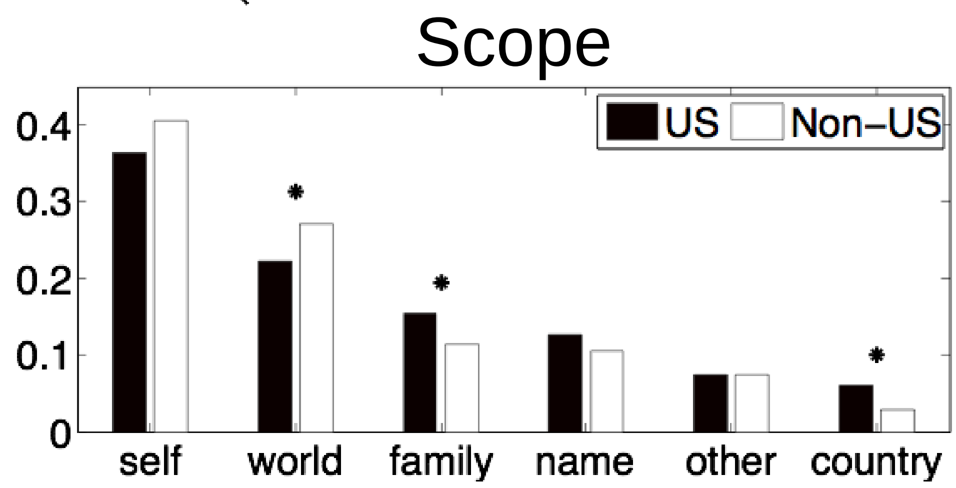


individual requests: “i wish for a new puppy”
solicitations: “call me *555-1234*”, “visit *website*.com”
sinister: “to take over the world”

WISH corpus: Geographical differences

- 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 (Pearson X^2-test, $p < 0.01$)
- *But* no significant difference between red vs. blue states





WISH corpus: Latent topic modeling

- Previous 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 α=0.5, β=0.1

WISH corpus: Latent topic modeling

Topic	Top words, sorted by p(word|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

world peace and my friends in iraq to come home

Building wish detectors

Wish detection

- Novel NLP task: 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"
- This is an initial study
  - Assume some labeled data in target domains
  - Try to beat some standard baselines by exploiting the WISH corpus to learn wish templates

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

Words
- Linear Support Vector Machine
- Train on labeled training set from the target domain
- binary word-indicator vector
- normalized to sum to 1
- Natural first baseline for a new text classification task

Expect high recall, low precision

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

- Can discover non-obvious templates

  world peace, peace on earth → **let there be** ___

  become rich, win the lottery → **to finally** ___

  get a job, save the environment → ___ **please**

The graph

- Formally, 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:

From left to right:

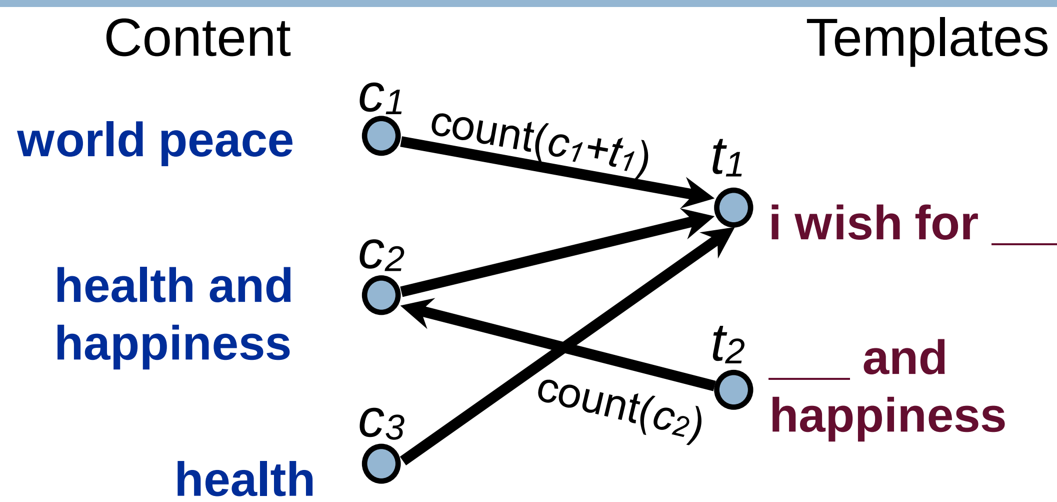
- If a wish w_j contains another wish w_i

  w_i = "world peace"

  w_j = "i wish for world peace"

then create content node $c = w_i$
and
template node t = "i wish for __"

- Create edge from $c \rightarrow t$
- Weight the edge by count(c+t)



From right to left:

- Some false-positive template nodes arise from nested contents
- Template t = "__ and happiness" is "bad" because it matches $w_j \in C$
- Create edge from $t \rightarrow c = w_j$
- Weight the edge by count(c)

The algorithm

- Intuition: useful templates match many complete wishes but few content-only wishes
- Score template nodes *t by score(t)* = *in(t) - out(t)*
- Apply threshold *score(t)* ≥ *5* to obtain 811 wish templates

Wish templates

Some of the 811 wish templates selected by our algorithm

Top 10	Others in Top 200
___ in 2008	i want to ___
i wish for ___	___ for everyone
i wish ___	i hope ___
i want ___	my wish is ___
i want my ___	___ please
___ this year	wishing for ___
i wish ___ in 2008	may you ___
i wish to ___	i wish i had ___
i wish ___ this year	to finally ___
___ in the new year	for my family to have ___

Learning with wish templates

- 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

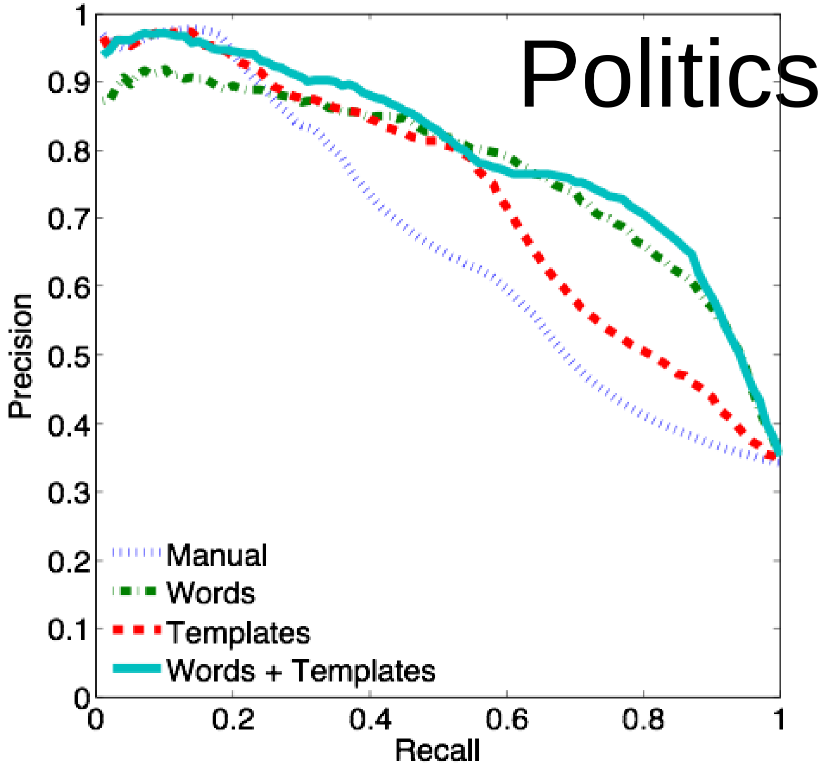
Test corpora

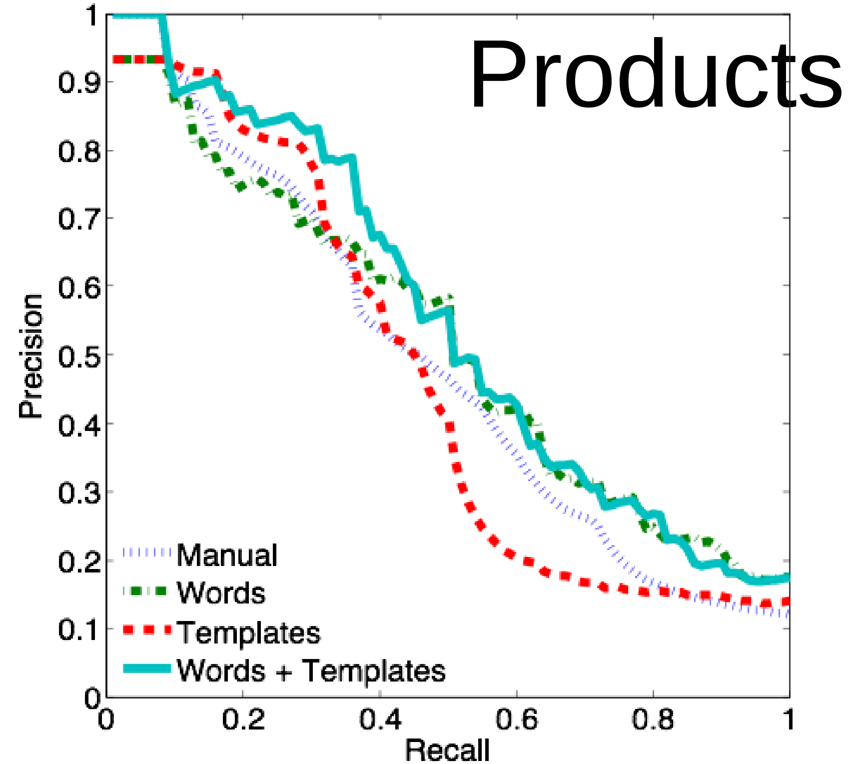
- Two test corpora, manually labeled sentences as wish vs. non-wish
  - Consumer product reviews
    - 1,235 sentences from amazon.com and cnet.com reviews
      (selected from data used in Hu and Liu, 2004; Ding et al., 2008)
    - 12% wishes
  - Political discussion board postings
    - 6,379 sentences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





AUC

Corpus	Manual	Words	Templates	Words + Templates
Politics	0.67 ± 0.03	0.77 ± 0.03	0.73 ± 0.03	0.80 ± 0.03
Products	0.49 ± 0.13	0.52 ± 0.16	0.47 ± 0.16	0.56 ± 0.16

What features are important?

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

Sign	Words	Templates	Words + Templates
+	wish	i hope ___	hoping ___
+	hope	i wish ___	i hope ___
+	hopefully	hoping ___	i just want ___
+	hoping	i just want ___	i wish ___
+	want	i would like ___	i would like ___
-	money	family ___	micro
-	find	___ forever	about
-	digital	let me ___	fix
-	again	___ d	digital
-	you	___ for my dad	you

Novel task 2

Measuring creativity in text

An example

- Consider the word "hamster"
- Which sentence is more creative?
  1. She asked if I had any pets, so I told her I once did until I discovered that I liked the taste of hamster.
  2. A hamster has four legs.
- Humans can assign a numerical creativity score to each sentence (9.25 vs. 0)
- **Our contribution:** A machine learning algorithm that automatically predicts the subjective creativity score.

The scope of our study

- Measuring human creativity in composing a single sentence, when the sentence is constrained by a **given keyword**.
  - limited.
  - A first step towards automatically measuring creativity in more complex natural language text.
  - Assume the sentence is meaningful, then

$$\text{creativity} \approx \text{outlier}$$

But what is creativity anyway?

- Subjective.
- Difficult to write down rules.
- Humans recognize creativity when they see it.
- We circumvent the definition problem by predicting human judgment scores:
  - don’t care what creativity is.
  - goal is to accurately predict scores from a training set.

The Creative Writing dataset

Procedure

1. We give a keyword z (e.g., “hamster”) to a human writer.
2. The human writer composes a sentence x about z.
3. Human judges assign the sentence x a creative score y.
4. Given a training set of m triples {(zi, xi, yi)} for i=1...m, we develop a statistical machine learning predictor Y(x, z) that predicts the score y.

The Wisconsin Creative Writing dataset

- m = 105 keywords
- 21 writers, 5 random keywords each
  - compose a not-so-creative sentence about one randomly selected keyword
    - two non-creative examples given to all writers:
    - “Iguana has legs” for “Iguana”
    - “Anvil can get rusty” for “Anvil”
  - compose four creative sentences about the other four keywords
    - no creative examples given, to avoid bias

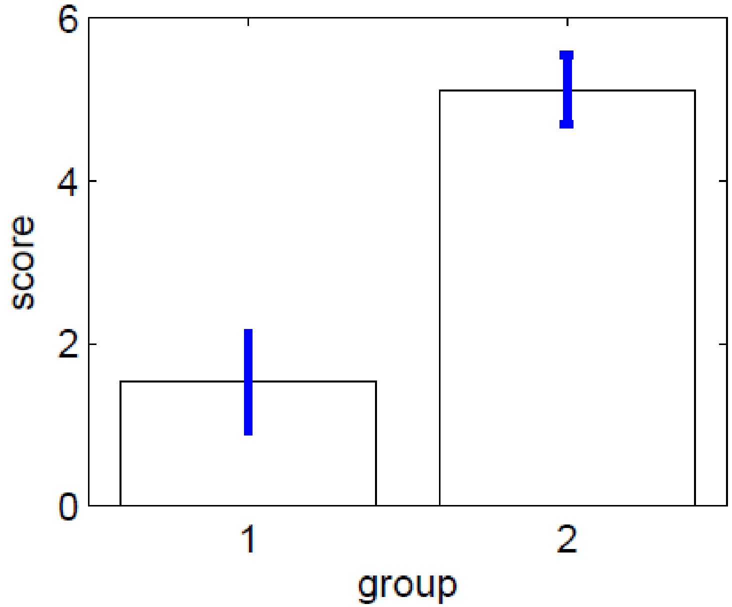
Judging

- All sentences and their keywords are given to 4 judges
- Each judge independently assigns a creativity score in 0–10 to each sentence-keyword pair
  - 0: not creative at all; judges are given the Iguana and Anvil examples
  - 10: the most creative
- General agreement on creativity: statistically significant ($p < 10^{-8}$) linear correlation among the four judges' score

correlation coefficient	judge 2	judge 3	judge 4
judge 1	0.68	0.61	0.74
judge 2		0.55	0.74
judge 3			0.61

Further validation

- The judges didn't know which sentences were instructed to be non-creative (group 1) or creative (group 2).
- Still, their scores are significantly lower on group 1 ($p < 10^{-11}$)



Examples from the dataset

average score y	keyword z	sentence $\mathbf{x}$
9.25	hamster	She asked if I had any pets, so I told her I once did until I discovered that I liked taste of hamster.
9.0	wasp	The wasp is a dinosaur in the ant world.
8.5	dove	Dove can still bring war by the information it carries.
		...
0.25	guitar	A Guitar has strings.
0.25	leech	Leech lives in the water.
0.25	elephant	Elephant is a mammal.

Dataset available at http://pages.cs.wisc.edu/~jerryzhu/pub/WisconsinCreativeWriting.txt

Learn to measure creativity

The machine learning problem

- Regression Y(x, z) = y
- Linear regression with 17 features (a few will be discussed next)
- Features from computer science perspective: language models
- Features from psychology perspective: word norms, WordNet

Feature f_3: language model

- Creative: other words x_{-z} in the sentence should be difficult to predict from the keyword z
- Feature is length-normalized "context" log-probability

$$f_3(\mathbf{x}, z) = \frac{1}{n-1} \log p(\mathbf{x}_{-z}|z) = \frac{1}{n-1} \log \frac{p(\mathbf{x})}{p(z)}$$

where n = length(x).

- p(x) estimated from Google 1T 5-gram corpus, using a simplified Jelinek-Mercer smoothed 5-gram language model.
- p(z) estimated from a unigram language model.

Feature f_4: word norms

- Norms: the set of words humans “think of” when given a keyword z
- z=dog, norms(z)={animal, bark, tail, bone, ...}
- The Leuven norms, collected in a psychology study
- Creative: not too many norm words in the sentence
- Feature is n

$$f_4(\mathbf{x}, z) = \frac{1}{n} \sum_{i=1}^{n} 1_{x_i \in \text{norms}(z)}$$

Features f_5 and f_{12}: WordNet

- Creativity related to similarity between keyword z and other words
- $s(z, x_i)$: WordNet path similarity between z and x_i by NLTK
- Mean similarity: $f_5(\mathbf{x}, z) = \frac{1}{n}\sum_{i=1}^{n} s(z, x_i)$
- $f_{12}(x, z)$ = the 5th largest similarity

Regression analysis

Linear correlation between each feature and label y

$$\rho_i = \frac{\mathrm{Cov}(f_i, y)}{\sigma_{f_i}\sigma_y}$$

	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$	$f_{1,5}$	f_2	f_3*	f_4*	f_5*
ρ	0.09	0.09	0.17	0.06	-0.04	-0.07	-0.32	-0.48	-0.41

	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}*	f_{13}
ρ	-0.19	-0.25	-0.02	0.06	0.23	0.30	0.47	-0.01

Language model features

Word norms feature

WordNet features

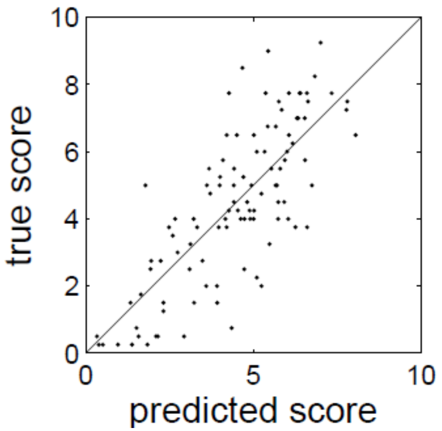
* features we discussed

Linear regression model

- Stepwise regression: a technique for feature selection
  - iteratively including / excluding candidate features based on statistical significance tests
  - results in a linear regression model with a small number of salient features

$$Y(\mathbf{x}, z) = -5.06 f_4 + 1.80 f_{12} - 0.76 f_3 - 3.39 f_5 + 0.92$$

- Final model
- Root mean squared error (RM
  Baseline constant predictor RI



Conclusions

- Two novel natural language processing tasks enabled by machine learning
  - Identifying wishes
  - Measuring creativity
- “Shallow” computation, not deep understanding (AI-complete)
- What else can we do?

Thank you