Will Computers Ever Understand Us?

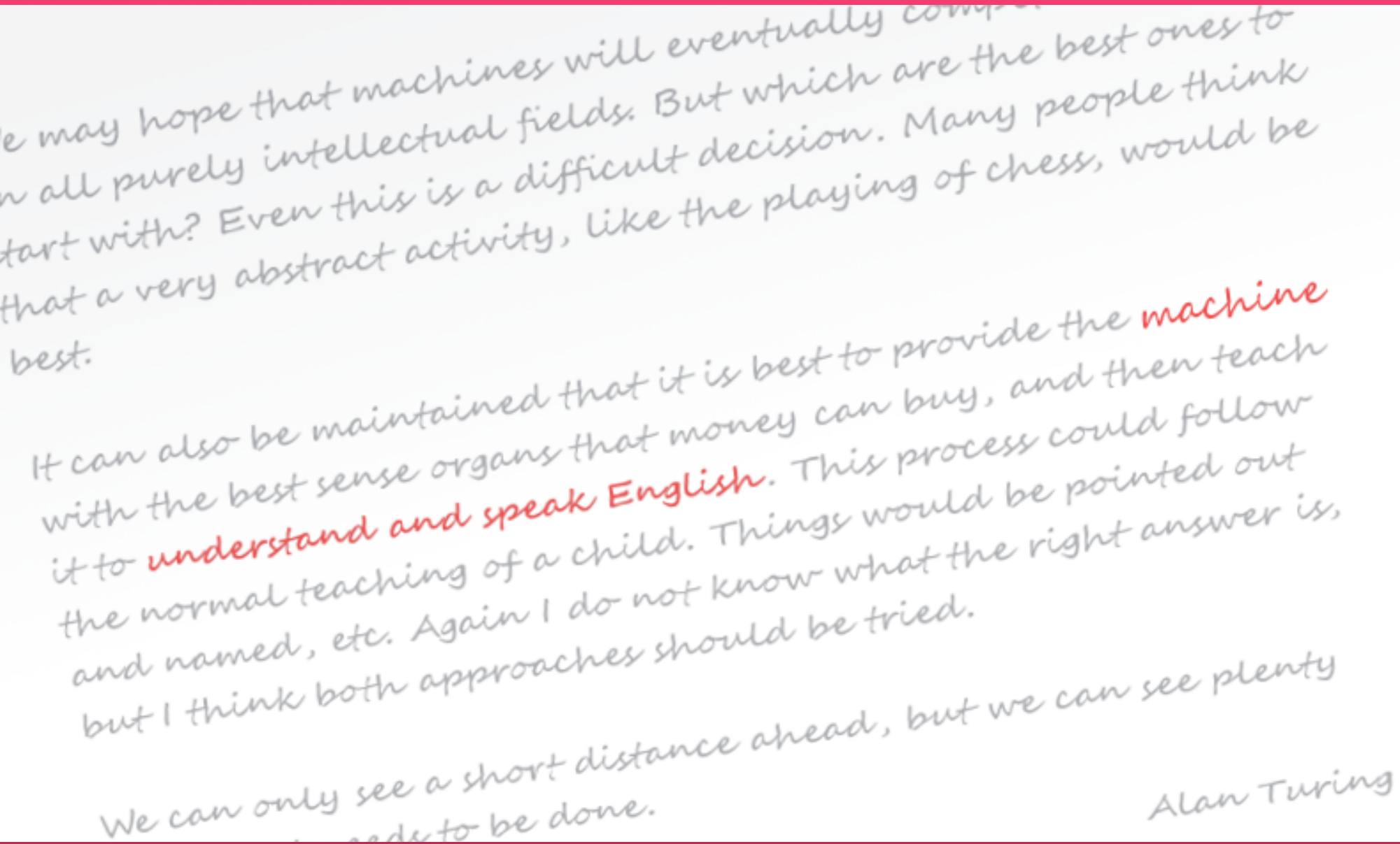
Zuzana Nevěřilová

Natural Language Processing Centre, Faculty of Informatics
Masaryk University, Brno, Czech Republic

xpopelk@fi.muni.cz

Workshop of Natural Language Processing Centre
May 28, 2013

Will computers ever understand us?



Computer “understanding”: Use cases

- inappropriate discussion posts detection
- text summarization
- opinion mining
- content targeting
- question answering

Inappropriate discussion posts detection

That BOY aint done growing and fcuking so she would be stooopid to tie HERSELF down wit a BABY and a tattoo is just as worse!!!

Inappropriate discussion posts detection

That BOY aint done growing and fcuking so she would be stooopid to tie HERSELF down wit a BABY and a tattoo is just as worse!!!

Inappropriate discussion posts detection

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Inappropriate discussion posts detection

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Inappropriate discussion posts detection

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Inappropriate discussion posts detection

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Inappropriate discussion posts detection

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Inappropriate discussion posts detection

fcuking *(adjective)*

Lemma	Score	Frequency
phucking	0.618	67
out-an-out	0.456	22
god-damned	0.445	357
self-deceiving	0.36	133
frigging	0.21	1296
sleazebags	0.181	93
whingeing	0.162	105
bold-faced	0.158	471
complusive	0.155	69
bald-faced	0.137	870
self-deluded	0.106	401
blithering	0.105	844
bare-faced	0.1	343
barefaced	0.097	740
cking	0.089	2365

Use case: discussion forum, automatic detection of inappropriate posts

Common solution: word list

But: users use concealed words that are difficult to detect (f*king, f.u.c.k, f..k, fcuking)

Better solution: word list + concealing rules

But: users invent new words and concealing patterns

Even better solution: word list + automatically generated thesaurus + concealing rules + metarules

Text summarization

When she got her first cellphone about a decade ago, one of the first things she did was go to her service provider's website and stock up on ringtones.

"I think I downloaded about 10 of them, and at that point they were $5 or $6 each," May, now 22, recalls of her adolescent shopping spree. "And then my dad got the phone bill."

At the time, the expense seemed worth it to have access to fresh music on her phone.

"When new songs came out, I would download them ... whatever was popular at the time," she told CNN. And yet, "the majority of the ones I downloaded, I never really used. I remember listening to them, but I don't remember using them, because most of the time I was in school or work."

Text summarization

When she got her first cellphone about a decade ago, one of the first things she did was go to her service provider's website and stock up on ringtones.

"I think I downloaded about 10 of them, and at that point they were $5 or $6 each," May, now 22, recalls of her adolescent shopping spree. "And then my dad got the phone bill."

At the time, the expense seemed worth it to have access to fresh music on her phone.

"When new songs came out, I would download them ... whatever was popular at the time," she told CNN. And yet, "the majority of the ones I downloaded, I never really used. I remember listening to them, but I don't remember using them, because most of the time I was in school or work."

A decade ago a girl spent money for ringtones.

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Text summarization

Use case: automatic abstract generation, multiple document digest, are these documents stating similar or opposite arguments?

Naive solution: take every first sentence in a paragraph

Common solution: take every sentence containing a keyword

But: not really scalable, difficult to detect the main message

Better approach:

1. analyse text on several levels
   - whole document (sections, paragraphs, consistency)
   - sequence of sentences (each having a structure)
   - bag of words and keywords (in different forms, synonyms, abbreviations etc.)
2. generate a summary

Opinion mining

The **iPhone 5** price was predictably high and continues to be so, so consumers will need to bear that in mind too when looking for their next smartphone.

...

Well, all of those picking up the iPhone 5 will have the same reaction: this thing is amazingly light. You've probably heard the numbers by now (20 per cent lighter than the predecessor, as well as beating most of the opposition too at 112g.)

Opinion mining

The **iPhone 5** price was predictably high and continues to be so, so consumers will need to bear that in mind too when looking for their next smartphone.

...

Well, all of those picking up the iPhone 5 will have the same reaction: this thing is amazingly light. You've probably heard the numbers by now (20 per cent lighter than the predecessor, as well as beating most of the opposition too at 112g.)


high price


weight

Opinion mining

Use case: what are people thinking about a particular product/company/idea X?

Solution: search X, find evaluative words

But: opinions are expressed by non-evaluative words

Better solution:

- extract useful attributes of X (noise, weight, price, appearance)
- generate thesauri of evaluative words:
  thin iPhone 5 $\times$ thin tasteless burger

Opinion mining

Use case: what are people thinking about a particular product/company/idea X?

Solution: search X, find evaluative words

But: opinions are expressed by non-evaluative words

Better solution:

- extract useful attributes of X (noise, weight, price, appearance)
- generate thesauri of evaluative words:
  thin iPhone 5 $\times$ thin tasteless burger

Opinion mining

Use case: what are people thinking about a particular product/company/idea X?

Solution: search X, find evaluative words

But: opinions are expressed by non-evaluative words

Better solution:

- extract useful attributes of X (noise, weight, price, appearance)
- generate thesauri of evaluative words:
  thin iPhone 5 $\times$ thin tasteless burger

Opinion mining

Use case: what are people thinking about a particular product/company/idea X?

Solution: search X, find evaluative words

But: opinions are expressed by non-evaluative words

Better solution:

- extract useful attributes of X (noise, weight, price, appearance)
- generate thesauri of evaluative words:
  thin iPhone 5 $\times$ thin tasteless burger

Question answering

Do you have a bike for a 4-year-old girl?

Search results for “bike”, “girl”

. . .

Do you have a bike for a 4-year-old girl?

If she is under 110 cm tall I will recommend Maggie, Princess or Misty. If she is taller I would recommend Miss B or Kellie. If she does not insist on bike for girls I would also recommend Racer or Mr. Lightning. How tall is she?

About 105 cm.

Do you have some other constraints?

I look for something cheaper.

Then I would recommend Princess. It is a popular bike.

Question answering

Use case: chatbot providing basic support

Solution: patterns, keyword detection, searching

But: no real dialogue, no real answers, just searching

Better solution: sentence structure analysis, keyword detection, coreference resolution, dialogue strategy

Question answering

Use case: chatbot providing basic support

Solution: patterns, keyword detection, searching

But: no real dialogue, no real answers, just searching

Better solution: sentence structure analysis, keyword detection, coreference resolution, dialogue strategy

Question answering

Use case: chatbot providing basic support

Solution: patterns, keyword detection, searching

But: no real dialogue, no real answers, just searching

Better solution: sentence structure analysis, keyword detection, coreference resolution, dialogue strategy

Question answering

Use case: chatbot providing basic support

Solution: patterns, keyword detection, searching

But: no real dialogue, no real answers, just searching

Better solution: sentence structure analysis, keyword detection, coreference resolution, dialogue strategy

Conclusions: Understanding of *understanding*

Is this real understanding?

Probably not.

We do not know what understanding is but we know how it looks like when someone understands.

Computer programs that can discover a vulgar text, summarize a text, recognize someone's feelings or answer questions **look like** they understand our language . . . .

Conclusions: Understanding of *understanding*

Is this real understanding?

Probably not.

We do not know what understanding is but we know how it looks like when someone understands.

Computer programs that can discover a vulgar text, summarize a text, recognize someone's feelings or answer questions **look like** they understand our language . . . .

Conclusions: Understanding of *understanding*

Is this real understanding?

Probably not.

We do not know what understanding is but we know how it looks like when someone understands.

Computer programs that can discover a vulgar text, summarize a text, recognize someone's feelings or answer questions **look like** they understand our language . . . .

Conclusions: Understanding of *understanding*

Is this real understanding?

Probably not.

We do not know what understanding is but we know how it looks like when someone understands.

Computer programs that can discover a vulgar text, summarize a text, recognize someone's feelings or answer questions **look like** they understand our language . . . .