TACKLING NATURAL LANGUAGE GENERATION CHALLENGES AT NARRATIVE SCIENCE

Mike Pham & Clayton Norris

Oct 19, 2017

Narrative Science

OVERVIEW

Quill is an Advanced Natural Language Generation (NLG) platform

NLG A form of artificial intelligence (AI) that automatically produces language from structured data.

intent-driven Advanced NLG uses intent, or what you want to know, as its guide from the very beginning.

So what?

So what?

- How is this different than Amazon sending me a templated email receipt of my recent purchases?

So what?

- · How is this different than Amazon sending me a templated email receipt of my recent purchases?
- · What about all those neural nets generating facebook posts that sound eeriely like my previous posts?

So what?

- · How is this different than Amazon sending me a templated email receipt of my recent purchases?
- · What about all those neural nets generating facebook posts that sound eeriely like my previous posts?
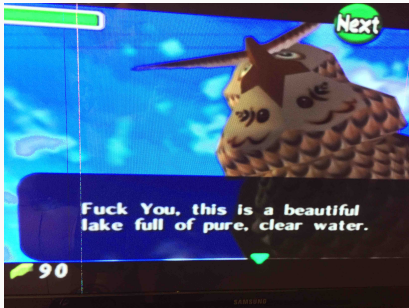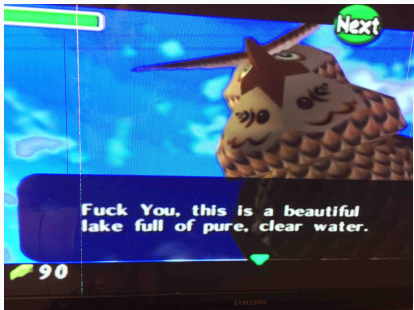
trigger warning: offensive language

trigger warning: offensive language

trigger warning: offensive language

- What seems off here?

· Video games conversations have complex decision trees

- Video games conversations have complex decision trees
    - Can result in very good and/or appropriate language
    - ...but often is mad-libby
    - Flexibility and linguistic creativity is limited and/or unscaleable in production

- Video games conversations have complex decision trees
    - Can result in very good and/or appropriate language
    - ...but often is mad-libby
    - Flexibility and linguistic creativity is limited and/or unscaleable in production
- Neural nets can learn from data to generate new language

· Video games conversations have complex decision trees
  · Can result in very good and/or appropriate language
  · ...but often is mad-libby
  · Flexibility and linguistic creativity is limited and/or unscaleable in production

· Neural nets can learn from data to generate new language
  · Can often produce highly natural and nuanced language
  · but has no idea what it's saying
  · and we have no idea why it's saying it either

VERB CONJUGATION

· A single verb can have various **word forms**:

(1) CREATE

    a. create, creates, created, creating

    b. creator, creation, creative, creatively

· (1a) is an example of **inflectional morphology**
  · expresses grammatical features
  · (usually) doesn't change basic meaning or part of speech

- **Grammatical features** are properties that the grammar of any language tracks and manifests
- Some features that English is sensitive to:
    - **number**: dog, dogs
    - **tense**: create, created
    - **gender**: he, she
    - **person**: we, yall, they
    - **mass/count**: 3 books, *3 bloods
    - **case**: I, me, my, mine

- Word forms can track multiple features at once
- This can be tracked within an **inflectional paradigm**

CREATE

Present		
	singular	plural
1	create	create
2	create	create
3	creates	create

Past		
	singular	plural
1	created	created
2	created	created
3	created	created

- Word forms can track multiple features at once
- This can be tracked within an **inflectional paradigm**

CREATE

Present		
	singular	plural
1	create	create
2	create	create
3	creates	create

Past		
	singular	plural
1	created	created
2	created	created
3	created	created

- Only 3rd person singular is different – this looks easy!
  - Just add -s to the 3.sg present form and -d to all past forms!

Unfortunately, we all know there are **irregular verbs** in English

BE

Present		
	singular	plural
1	am	are
2	are	are
3	is	are

Past		
	singular	plural
1	was	were
2	were	were
3	was	were

Unfortunately, we all know there are **irregular verbs** in English

BE

Present					Past				
	singular	plural					singular	plural	
1	am	are			1		was	were	
2	are	are			2		were	were	
3	is	are			3		was	were	

· Darn, how do we get am or was from be?

PRONOUNS

SENTENCE SELECTION

Grammaticality: only grammatical and accurate sentences should be **generated**

Sentence selection: the stylistically best sentence from the set of grammatical candidate sentences should be **selected**

Grammaticality: only grammatical and accurate sentences should be **generated**

Sentence selection: the stylistically best sentence from the set of grammatical candidate sentences should be **selected**

· but what determines a stylistically 'good' sentence?

- Most native speakers will agree when a sentence is grammatical
- But style is vague and elusive, varying from person to person

- Most native speakers will agree when a sentence is grammatical
- But style is vague and elusive, varying from person to person
- Which do think is the best sentence?

(2)   Aaron Young generated $3M in revenue in 2016.

(3)   Aaron Young's revenue was $3M in 2016.

(4)   Revenue for Aaron Young was $3M in 2016.

(5)   In 2016, Aaron Young generated $3M in revenue.

(6)   Aaron Young's 2016 generated revenue was $3M.

CONCLUSION

ELEMENTS

```
The theme provides sensible defaults to \emph{emphasis}
text, \alert{accent} parts or show \textbf{bold} results.
```

becomes

The theme provides sensible defaults to emphasis text, accent parts or show **bold** results.

CONCLUSION

Get the source of this theme and the demo presentation from

github.com/matze/mtheme

The theme itself is licensed under a Creative Commons
Attribution-ShareAlike 4.0 International License.

QUESTIONS?