Panel: Open Problems in MT
Challenges in MT for the 2020’s

Ondřej Bojar moderating
Alex Fraser, Yvette Graham,
Marcin Junczys-Dowmunt, Alon Lavie

WMT19 Florence; Aug 1, 2019
Outline

▶ Funding.
▶ User model.
▶ Speech translation.
▶ (Vision and other modalities.)
▶ Evaluation and human parity.
▶ Training Superhuman.
▶ Better training methods.
▶ Understanding for MT.
▶ What’s on the next plateau?
MT is becoming very, very good in many settings. Some might say even better than human performance. esp. if ref. translators have an axe over their heads.

Is the current MT quality becoming sufficient?

Are we going to see the end of big MT research labs at companies and academic institutions?
Funding

MT is becoming very, very good in many settings. Some might say even better than human performance. esp. if ref. translators have an axe over their heads.

Is the current MT quality becoming sufficient?

Are we going to see the end of big MT research labs at companies and academic institutions?

Explain to your boss or funding agency:

why is your work on MT important.

why it will be important in 5 years from now.
User Model

- Current APIs: Sentence in, sentence out.
- Current research: Document-level coherence.
- Our aim: LLL FAHQMT.
  - (Life-Long Learning Fully Automatic High Quality MT).
- Successful use cases of MT now:
  - Gisting for end users, without any usable feedback.
  - Post-edited at translation agencies feeds into TMs.
User Model

- Current APIs: Sentence in, sentence out.
- Current research: Document-level coherence.

- Our aim: LLL FAHQMT.
  - (Life-Long Learning Fully Automatic High Quality MT).

- Successful use cases of MT now:
  - Gisting for end users, without any usable feedback.
  - Post-edited at translation agencies feeds into TMs.

- How should the APIs be changed?
- How can we best interact with real users?
Speech Translation

- Speech translation $\neq$ ASR + MT + TTS.
- Human interpreters can sustain only for $\sim$20 minutes at a time.
- What is the potential of speech translation?
- What are the hardest problems?
Speech Translation

- Speech translation \( \neq \) ASR + MT + TTS.
- Human interpreters can sustain only for \( \sim \)20 minutes at a time.

- What is the potential of speech translation?
- What are the hardest problems?
- How simultaneous can we go?
  - Can we beat interpreters in guessing the German verb to come?

Caveat: Interpreting is *way more* than just translating the uttered words.
Humans arguably use all relevant available signal.

An image is worth 1000 words...
Vision and Other Modalities

- Humans arguably use all relevant available signal.
- An image is worth 1000 words...
  - ...but only in 1/1000 cases!
Vision and Other Modalities

- Humans arguably use all relevant available signal.
- An image is worth 1000 words...
  - ...but only in 1/1000 cases!
- What properties should multi-modal tasks have?
- Is the evaluation going to be harder or easier?
Evaluation and Human Parity

- In some settings, we are really running into our limits in assessing output quality.
- Can we provide both a simple and encompassing picture?
  - WMT DA ranking is simple.
  - Many test suites together are encompassing.
- Why do we trust assessors more than translators?
Evaluation and Human Parity

- In some settings, we are really running into our limits in assessing output quality.

- Can we provide both a simple and encompassing picture?
  - WMT DA ranking is simple.
  - Many test suites together are encompassing.

- Why do we trust assessors more than translators?

- How big of a problem is this for the field?

- How can we address it?
Once human parity is reached, how do we improve further?

In Funding, we discussed if anyone will care.
Here we discuss how to tackle it.

A natural option:
Train only on higher and higher quality data.

My proof that parity was reached: back-translated data often serves better than genuine!

What else has to be changed?
Training Superhuman_{Adv}

Once human parity is reached, how do we improve further?

In Funding, we discussed if anyone will care. Here we discuss how to tackle it.

A natural option:

Train only on higher and higher quality data.

My proof that parity was reached: back-translated data often serves better than genuine!

What else has to be changed?
Training Superhuman_{Adv}

- Once human parity is reached, how do we improve further?
  - In Funding, we discussed if anyone will care.
  - Here we discuss *how* to tackle it.
- A natural option:
  - Train only on higher and higher quality data.
Once human parity is reached, how do we improve further?

- In Funding, we discussed if anyone will care.
- Here we discuss how to tackle it.

A natural option:

- Train only on higher and higher quality data.
- My proof that parity was reached: back-translated data often serves better than genuine!
Training Superhuman_{Adv}

▶ Once human parity is reached, how do we improve further?
  ▶ In Funding, we discussed if anyone will care.
  ▶ Here we discuss *how* to tackle it.
▶ A natural option:
  ▶ Train only on higher and higher quality data.
  ▶ My proof that parity *was* reached:
    back-translated data often serves better than genuine!
▶ What else has to be changed?
Better Training Methods

NMT training has to be improved in multiple ways:

▶ Document-level training.
Better Training Methods

NMT training has to be improved in multiple ways:

▶ Document-level training. Solved by Marcin’s paper??

▶ Low-resource conditions.

▶ Life-long learning vs. Chinese-soup models.

▶ What should our systems ideally learn from?

- Unlabelled inputs?
- Scarce feedback?
- Or everything available, collected in endless loop?

▶ From alchemy to chemistry.

▶ Reintroduce composability to our models?

▶ Make use of large pre-trained models/components.

▶ Know how the neural network reacts to our training.
Better Training Methods

NMT training has to be improved in multiple ways:

- Document-level training. Solved by Marcin’s paper??
- Low-resource conditions.
Better Training Methods

NMT training has to be improved in multiple ways:

▶ Document-level training. Solved by Marcin’s paper??
▶ Low-resource conditions.
▶ Life-long learning vs. Chinese-soup models.
  ▶ What should our systems ideally learn from?
    Unlabelled inputs? Scarce feedback?
    Or everything available, collected in endless loop?

From alchemy to chemistry.
▶ Reintroduce composability to our models?
▶ Make use of large pre-trained models/components.
▶ Know how the neural network reacts to our training.
Better Training Methods

NMT training has to be improved in multiple ways:

▶ Document-level training. Solved by Marcin’s paper??
▶ Low-resource conditions.
▶ Life-long learning vs. Chinese-soup models.
   ▶ What should our systems ideally learn from?
     Unlabelled inputs? Scarce feedback?
     Or everything available, collected in endless loop?
▶ From alchemy to chemistry.
   ▶ Reintroduce composability to our models?
   ▶ Make use of large pre-trained models/components.
   ▶ *Know* how the neural network reacts to our training.
Understanding in MT

- NMT is not understanding any better than PBMT.
- Yes, fluency improved a lot, adequacy improved too but catastrophic errors still happen.
  - Document-specific terminology totally unhandled (see the Audit/Agreements Test Suite).
- How far can we get without understanding?
- How to define understanding for our purposes to make it somehow practically useful?
The Next Plateau

- Some people saw a plateau with PBMT and friends.
  - At that point, rich morphology was treated but far from solved.
- NMT came and the problems of word order and rich inflectional morphology more or less vanished.
The Next Plateau

- Some people saw a plateau with PBMT and friends.
  - At that point, rich morphology was treated but far from solved.
- NMT came and the problems of word order and rich inflectional morphology more or less vanished.
- What are cool problems that are still difficult now?
The Next Plateau

- Some people saw a plateau with PBMT and friends. At that point, rich morphology was treated but far from solved.
- NMT came and the problems of word order and rich inflectional morphology more or less vanished.
- What are cool problems that are still difficult now?
- What problems do you expect to be “current” ones when NMT reaches a similar plateau?