Machine Translation

Computational Linguistics: Jordan Boyd-Graber
University of Maryland

NEURAL MODELS

Adapted from material by Thang Luong
What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent!
What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent! (Looks better, hard to diagnose errors)
What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent! (Looks better, hard to diagnose errors)
- More popular, seems to work better (with enough data)
RNNs to the rescue

I am a student
Je suis étudiant
RNNs to the rescue

Encoder

I
am
a
student
RNNs to the rescue

Encoder

I am a student

Je suis étudiant

Boundary marker
RNNs to the rescue

Encoder

I am a student

Je suis étudiant

• Recurrent Neural Networks:
  – Model $P(\text{target} | \text{source})$ directly.
  – Can be trained end-to-end.

Neural Machine Translation (NMT)

I

Encoder

I am a student

Je suis étudiant
RNNs to the rescue

Encoder

Decoder

Boundary marker

I am a student

Je suis étudiant

• Recurrent Neural Networks: Can be trained end-to-end.
Test time . . .

étudiant
  _
  Je
  moi
  suis

moi

0.1
0.1
0.3
0.1
0.4

I am a student _
Test time …
Test time . . .

éudiant

Je

moi

suis

moi

suis étudiant

0.1

0.1

0.3

0.4

0.1

0.1

0.1

0.1

0.6

0.6

0.5

0.2

0.1

0.1

0.1

0.1

0.1
Test time . . .

\[
\text{étudiant} \quad 0.1 \quad 0.1 \\
\text{Je} \quad 0.3 \quad 0.1 \\
\text{moi} \quad \text{0.4} \quad 0.1 \\
\text{suis} \quad 0.1 \quad \text{0.6}
\]

\[
\text{moi} \quad 0.1 \\
\text{suis} \quad 0.5 \\
\text{étudiant} \quad 0.1 \\
\]

I am a student . . .

moi

suis

étudiant

Computational Linguistics: Jordan Boyd-Graber | UMD
Test time ...
Problems

- Out of vocabulary words
- Complicated morphology
- Long sentences
Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology
- Long sentences
Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology: character-level models
- Long sentences
Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology: character-level models
- Long sentences: attention
Attention

知识就是力量 <end>
Attention

知识就是力量 <end>
Attention

Encoder

知识 → 就是力量 <end>
Attention
Attention
Attention
Attention
Attention
Attention
Attention

```
Knowledge
  └── e0 ─── e1 ─── e2 ─── e3 ─── e4 ─── e5 ─── e6 ─── <end>
      └── d0 ─── d1 ─── d2 ─── d3
        └── is ─── power
```
Attention

Encoder

知识 → e₁ → e₂ → e₃ → e₄ → e₅ → e₆

Decoder

Knowledge → d₀ → d₁ → d₂ → d₃

is → power → <end>
Beyond simple models

- Active area of research
- Implementations in all frameworks
- Reinforcement learning is huge component
- Low data