Decoding with sampling and nonlocal features

Wilker Aziz
Team: Federico, Hamid, Vanessa, Bill, Vito, Massi

September 13, 2014
Sampling and MT

1. Simpler DP-friendly model
Sampling and MT

1. Simpler DP-friendly model
2. translation features + partial LM
Sampling and MT

1. Simpler DP-friendly model
2. translation features + partial LM
3. MC sampling
Sampling and MT

1. Simpler DP-friendly model
2. translation features + partial LM
3. MC sampling
4. independent samples
Sampling and MT

1. Simpler DP-friendly model
2. translation features + partial LM
3. MC sampling
4. independent samples
5. Expectations computed using importance sampling can incorporate complex features at this point
Sampling and MT

1. Simpler DP-friendly model
2. translation features + partial LM
3. MC sampling
4. independent samples
5. Expectations computed using importance sampling
   can incorporate complex features at this point
6. Decode with MBR
   because we know how to estimate expectations
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   - design
     negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   - implement
   - test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   ▶ design
   negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   ▶ implement
   ▶ test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   ▶ design
   negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   ▶ implement
   ▶ test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   ▶ design
      negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   ▶ implement
   ▶ test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   - design
     - negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   - implement
   - test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   ▶ design
      negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   ▶ implement
   ▶ test
4. Minimum risk training
This week

1. Train a baseline Chinese-English experiment with n-gram LMs
2. Framework for nonlocal features
3. Features
   - design
     - negation (Federico), reordering (Hamid and Massi), distributed representation (Vanessa), questions (Bill), NE (Vito)
   - implement
   - test
4. Minimum risk training
Remarks

Preliminary findings
Remarks

Preliminary findings

▶ rejection sampling: very slow :(

> implemented consensus decoding (DeNero et al, 2009)

Code
https://github.com/wilkeraziz/chisel-features

Coming soon

▶ paper and complete code

5 sampling algorithms for PB- and HPB-SMT

▶ ack: MODIST project
Remarks

Preliminary findings

- rejection sampling: very slow :(  
- importance sampling: fast :) high variance :( 

5 sampling algorithms for PB- and HPB-SMT

ack: MODIST project

implemented consensus decoding (DeNero et al, 2009)

Code https://github.com/wilkeraziz/chisel-features

Coming soon

▶ paper and complete code
Remarks

Preliminary findings

- rejection sampling: very slow :(  
- importance sampling: fast :) high variance :(  
- more samples (to cope with variance) → MBR is slow $O(n^2)$
Remarks

Preliminary findings

- rejection sampling: very slow :( 
- importance sampling: fast :) high variance :( 
- more samples (to cope with variance) $\rightarrow$ MBR is slow $O(n^2)$ 
  implemented consensus decoding (DeNero et al, 2009)
Remarks

Preliminary findings

▶ rejection sampling: very slow :(  
▶ importance sampling: fast :) high variance :(  
▶ more samples (to cope with variance) → MBR is slow $O(n^2)$  
implemented consensus decoding (DeNero et al, 2009)

Code
https://github.com/wilkeraziz/chisel-features
Remarks

Preliminary findings

▶ rejection sampling: very slow :(
▶ importance sampling: fast :) high variance :(  
▶ more samples (to cope with variance) → MBR is slow $O(n^2)$  
implemented consensus decoding (DeNero et al, 2009)

Code
https://github.com/wilkeraziz/chisel-features

Coming soon

▶ paper and complete code  
5 sampling algorithms for PB- and HPB-SMT
▶ ack: MODIST project
Thanks!
References 1