learning language through interaction

Hal Daumé III | University of Maryland | me@hal3.name | @haldaume3

What is NLP?

- Fundamental goal: deep understanding of text
  - Not just string processing or keyword matching
- End systems that we want to build
  - Simple: Spelling correction, text categorization, etc.
  - Complex: Speech recognition, machine translation, information extraction, dialog interfaces, question answering
  - Unknown: human-level comprehension (more than just NLP?)

Simultaneous (machine) interpretation

Nuremburg Trials

- Dozens of defendants
- Judges from four nations (three languages)
- Status quo: speak, then translate
- After Nuremberg, simultaneous translations became the norm
- Long wait → bad conversation

Skype translator demo

Skype translator demo

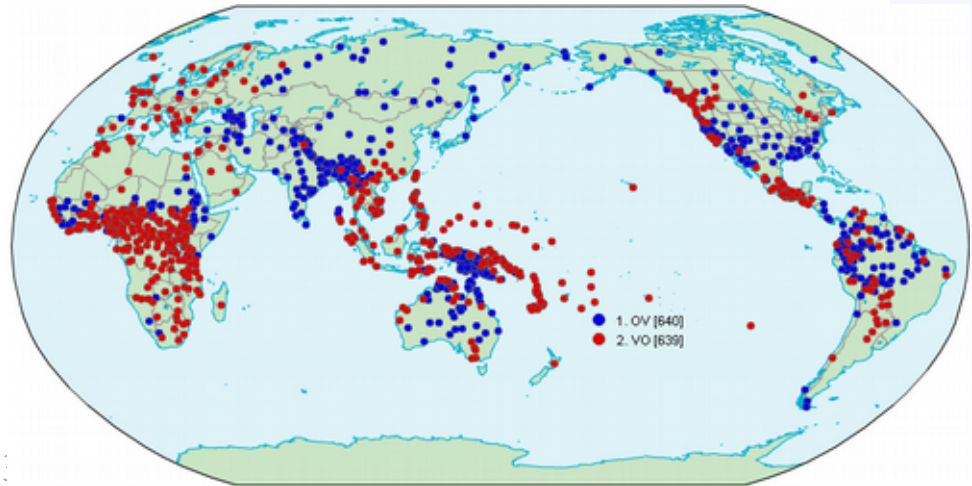
"Sorry Melanie,
say that again"

Why simultaneous interpretation is hard

- Human languages have vastly different word orders
  - About half are OV, the other half are VO
  - This comes with a lot more baggage than just verb-final

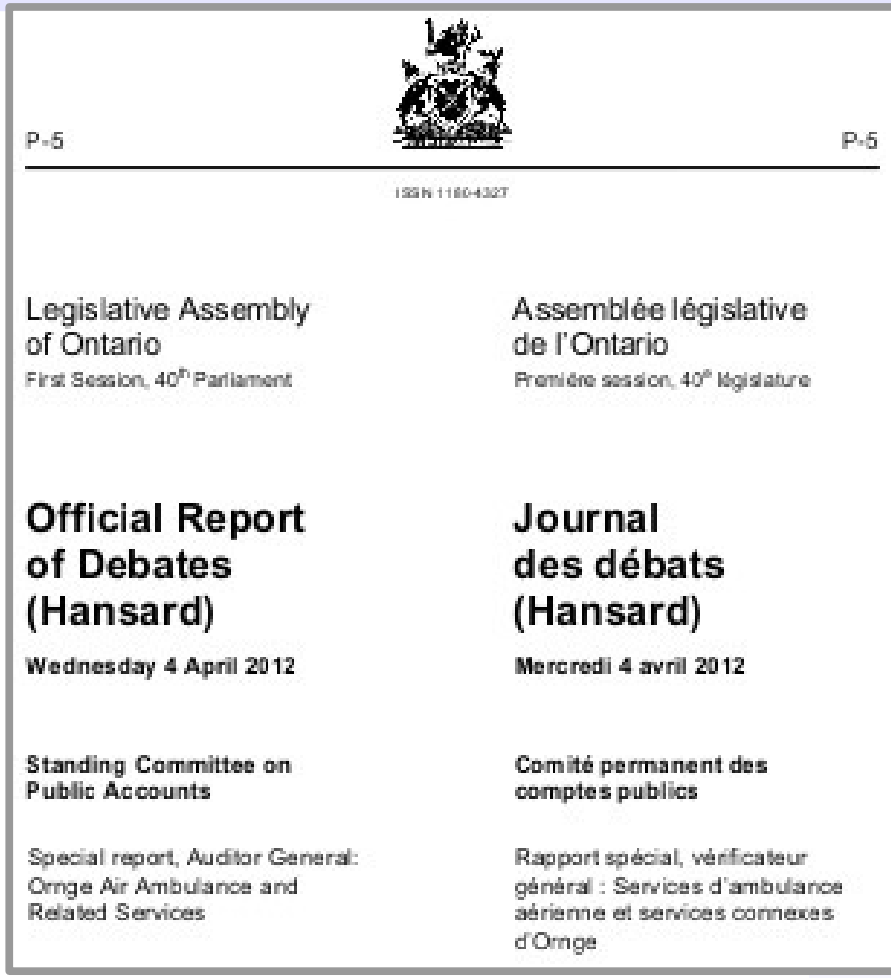
dinosaur-TOP store-LOC go-PAST
the dinosaur went to the store

food-OBJ buy-DESIRE dinosaur-TOP store-LOC go-PAST
the dinosaur who wanted to buy food went to the store



One slide on statistical machine *translation*

- Collect a bunch of text that's been translated by humans ("parallel corpus")
- Break it into "aligned" sentence pairs
- Learn statistical model to map from the "source" to "target"
- Apply that model to novel "unseen" sentences that you want to translate



P-5 P-5

ISSN 1180-4327

Legislative Assembly of Ontario
First Session, 40th Parliament

Assemblée législative de l'Ontario
Première session, 40e législature

Official Report of Debates (Hansard)

Wednesday 4 April 2012

Standing Committee on Public Accounts

Special report, Auditor General: Ornge Air Ambulance and Related Services

Journal des débats (Hansard)

Mercredi 4 avril 2012

Comité permanent des comptes publics

Rapport spécial, vérificateur général : Services d'ambulance aérienne et services connexes d'Ornge

Example

つまり例えばこの表現一は認識できますが二から四は認識できない

Batch
They might recognize expression one but not expression two to four.

Interp
The phrase number one only is accepted and phrases two, three, four were not accepted.

(He et al., NAACL 2016)

General diffs of Interp vs Batch

- Inversion
  - Segmentation into multiple sentences
  - Passivization of single sentence
- Word generalization
  - (lower retrieval time)
- Summarization and omission
  - (to catch up)

(He et al., NAACL 2016)

Example (gen + segment)

(S) この日本語の待遇表現の特徴ですが英語から日本語へ直訳しただけでは表現できないといった特徴があります

(Batch) One of the characteristics of **honorific** Japanese is that it can not be **adequately** expressed when using a direct translation from English to Japanese.

(Interp) Now let me talk about the characteristic of the Japanese **polite** expressions. `<segment/>` And such expressions can not be expressed **enough** just by translating directly.

(He et al., NAACL 2016)

Example (gen + summarize)

(S) で三番目の特徴としてはですねえ出来る限り自然な日本語の話言葉とてその出力をするといったような特徴があります .

(Batch) Its third **characteristic** is that its output is, as much as possible, in the natural language of spoken ((Japanese)).

(Interp) And the third **feature** is that the translation could be produced in a very natural spoken language.

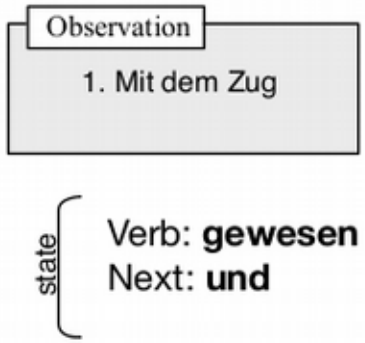
(He et al., NAACL 2016)

Model for interpretation decisions

- We have a set of actions (predict / translate)
  - Wait
  - Predict clause-verb
  - Predict next word
  - Commit ("speak")
- In a changing environment (state)
  - The words we've seen so far
  - Our models' internal predictions
- Have example data from human interpreters that demonstrates what to do!

(Grissom II et al., EMNLP 2014)

Example of interpretation trajectory



Ich	bin	mit	dem	Zug	nach	Ulm	gefahren
I	am	with	the	train	to	Ulm	traveled
I	*(……waiting……)*						traveled by train to Ulm

(Grissom II et al., EMNLP 2014)

DAgger: Dataset Aggregation

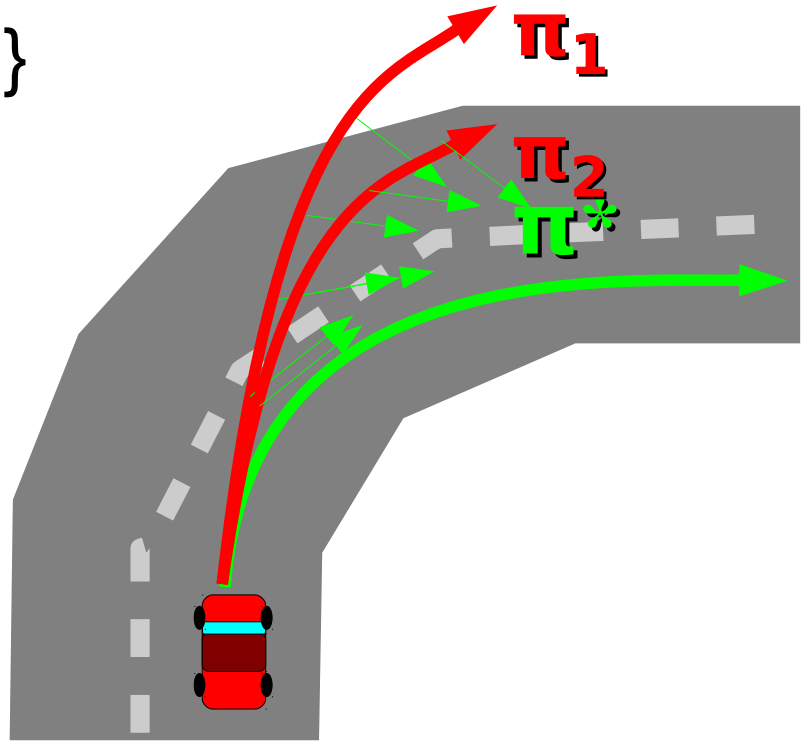
- Collect trajectories from expert π^*
- Dataset $\mathbf{D_0} = \{ (s, \pi^*(s)) \mid s \sim \pi^* \}$
- Train π_1 on $\mathbf{D_0}$
- Collect new trajectories from π_1
  - But let the *expert* steer!
- Dataset $\mathbf{D_1} = \{ (s, \pi^*(s)) \mid s \sim \pi_1 \}$
- Train π_2 on $\mathbf{D_0} \cup \mathbf{D_1}$

- In general:
  - $\mathbf{D_n} = \{ (s, \pi^*(s)) \mid s \sim \pi_n \}$
  - Train π_n on $\mathbf{U}_{i<n} \mathbf{D_i}$

If N = T log T,

$\mathbf{L(\pi_n) < T \epsilon_N + O(1)}$

for some n



(Ross+Gordon+Bagnell, AIStats 2011)

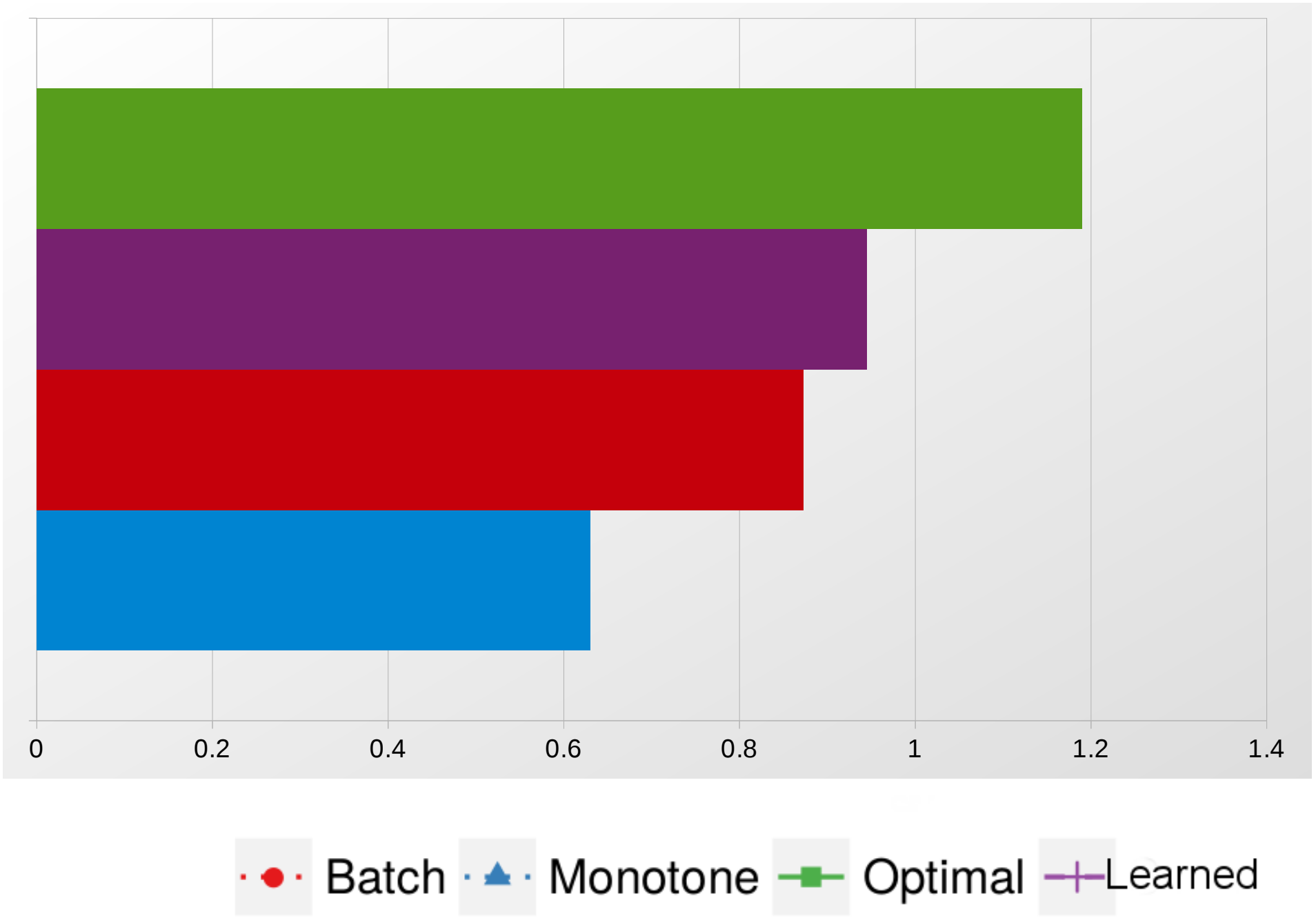
Training the policy

- Actions:
  - Commit `translate(revealed words)`
  - Predict (verb/next) `translate(revealed + predicted)`
  - Wait `get_next_words()`

- Features:
  - Output & confidence of predictors
  - Internal translation / language model scores
  - Previous decisions made by policy

(Grissom II et al., EMNLP 2014)

Evaluating performance



(Grissom II et al., EMNLP 2014)

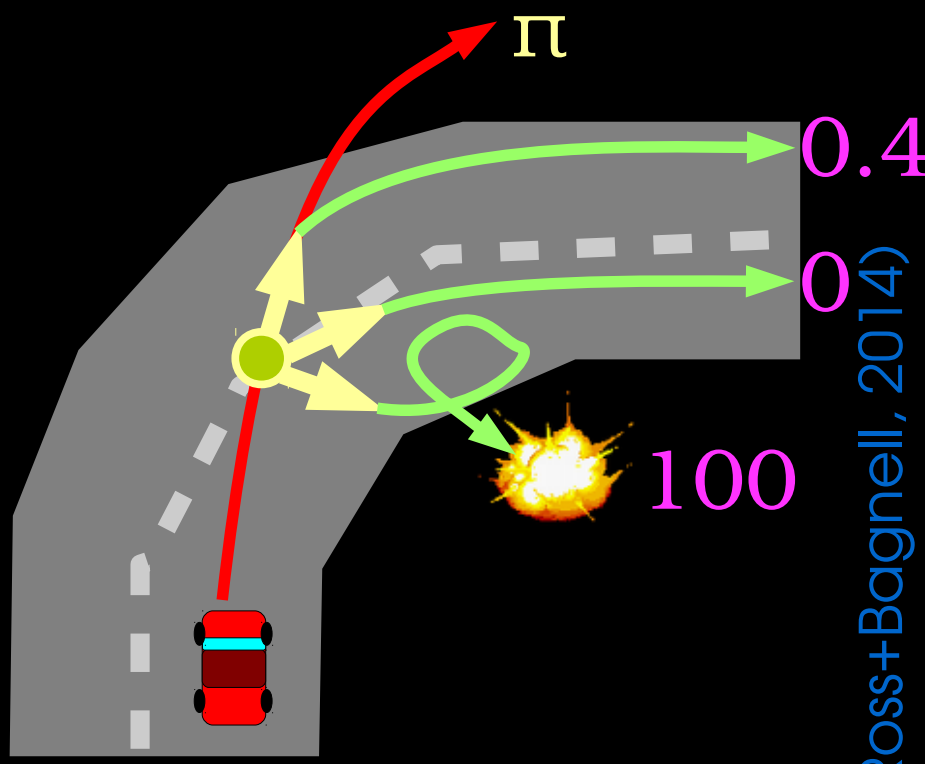
But wait, how good are humans....???

- Everyone makes errors:
  - Best human interpreters make some
  - Most human interpreters make many
- Mimicking human behavior is suboptimal!
- But then, what can we do?

Learning to search: AggraVaTe

1. Let learned policy π drive for t timesteps to obs. o
2. For each possible action a:
   - Take action a, and let expert π^{ref} drive the rest
   - Record the overall loss, c_a
3. Update π based on example:

   $(o, \langle c_1, c_2, \ldots, c_K \rangle)$
4. Goto (1)



(Ross+Bagnell, 2014)

Learning to search: AggraVaTe

On the bright side:

Doesn't need immediate feedback on every decision

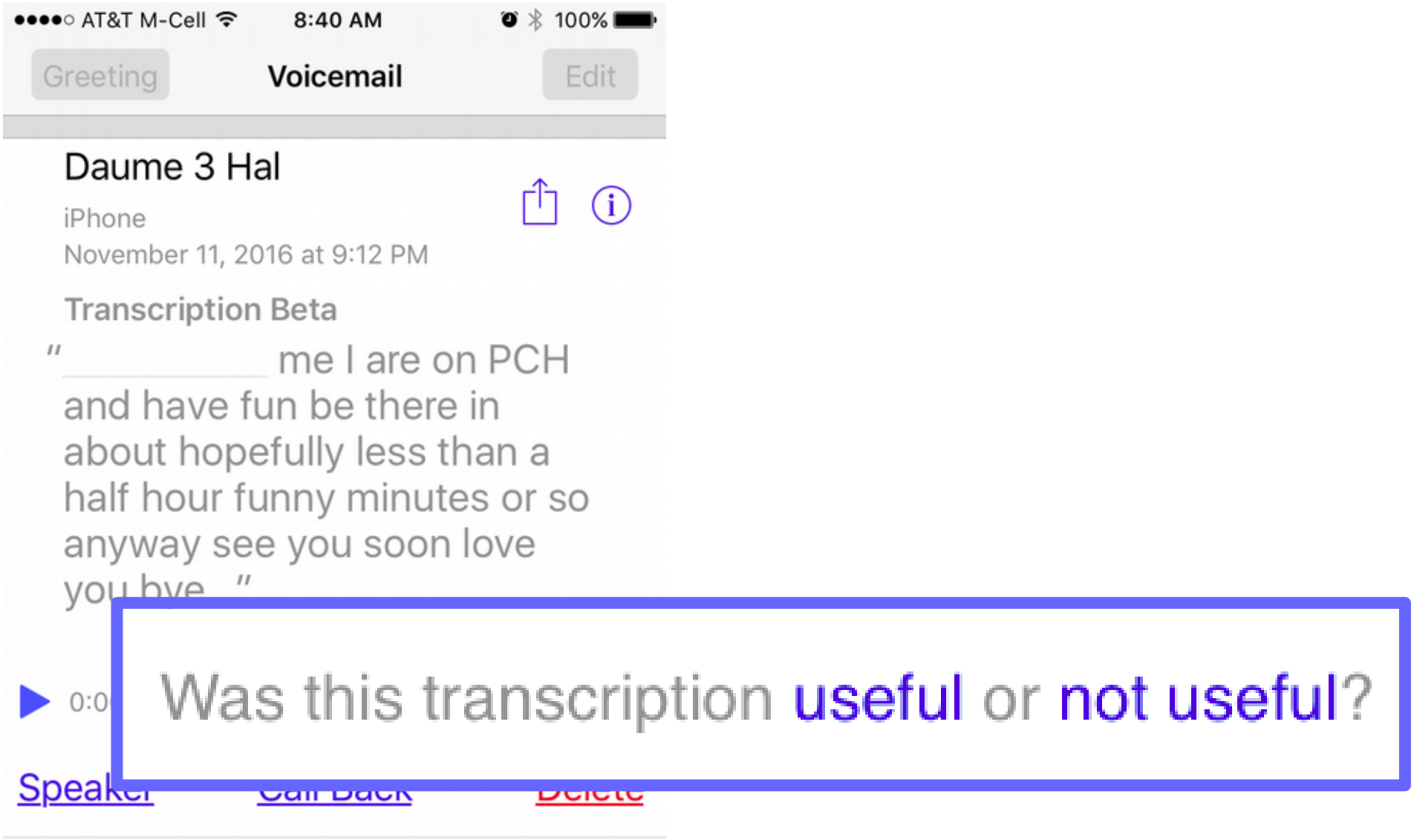
On the not-so-bright side:

Have to evaluate a *ton* of trajectories

Need to assume driver is optimal

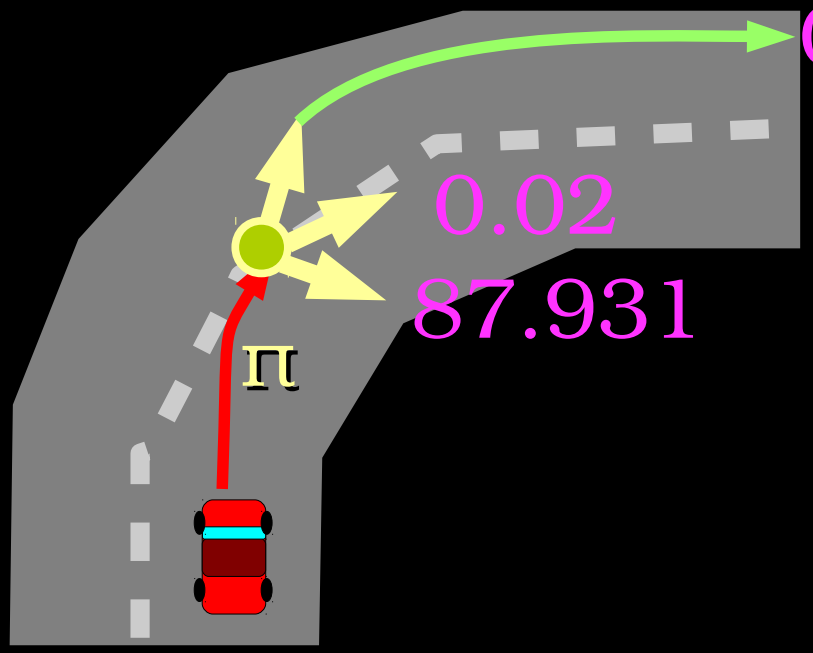
(Ross+Bagnell, 2014)

Motivating setting



BanditLOLS: learning from bandit feedback

1. Let learned policy π drive for t timesteps to obs. o
2. Choose a single "deviation" action a:
   - Take action a, and let policy π^{out} drive the rest
   - Record the overall loss, c_a
3. Estimate the cost of all $a' \neq a$ by regression
4. Update π based on example: $(o, \langle c_1, c_2, \ldots, c_K \rangle)$
5. Goto (1)



(Sharaf+D, AIStats 2017)

BanditLOLS: learning from bandit feedback

- Human feedback (thumbs up/thumbs down) only required on one prediction per input

- Evaluation balances exploration & exploitation

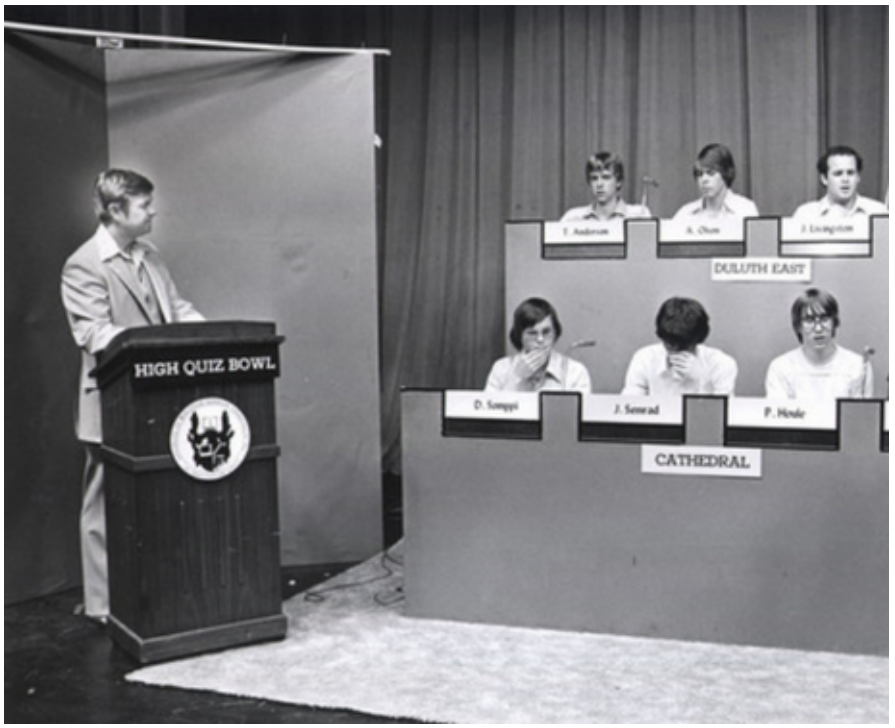
Algorithm	Exploration	POS Accuracy	Dependency UAS	Chunking F-Score
Reference	-	47.24	44.15	74.73
LOLS	ϵ-greedy	2.29	18.55	31.76

all code available, integrated into Vowpal Wabbit (`github.com/hal3/`)

(Sharaf+D, AIStats 2017?)

Humans are not optimal at *games* either!

- Game called "quiz bowl"
- Two teams play each other
  - Moderator reads a question
  - When a team knows the answer, they buzz in
  - **If right**, they get points; **otherwise**, rest of the question is read to the other team
- Hundreds of teams in the US alone
- Example …



(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous

(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory

(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so

(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients

(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by

(Iyyer et al., ACL 2014)

Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating

(Iyyer et al., ACL 2014)

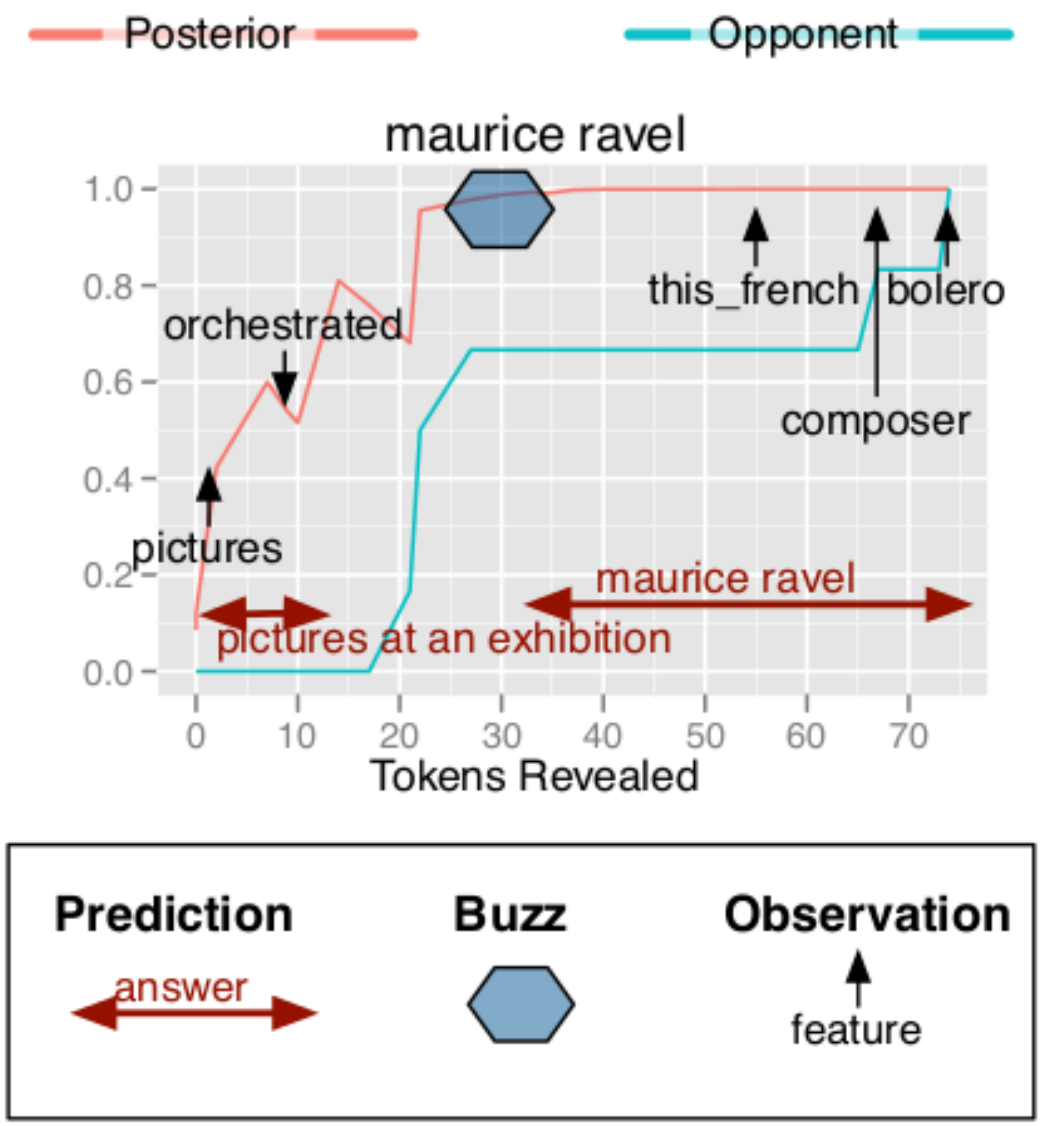
Quizbowl example

With Leo Szilard, he invented a doubly-eponymous refrigerator with no moving parts. He did not take interaction with neighbors into account when formulating his theory of heat capacity, so Debye adjusted the theory for low temperatures. His summation convention automatically sums repeated indices in tensor products. His name is attached to the A and B coefficients for spontaneous and stimulated emission, the subject of one of his multiple groundbreaking 1905 papers. He further developed the model of statistics sent to him by Bose to describe particles with integer spin. For 10 points, who is this German physicist best known for formulating the special and general theories of relativity?

(Iyyer et al., ACL 2014)

Evaluation methodology

- Mechanical Turk to collect human data
- 7000 questions were answered in the first day
- Over 43000 questions were answered in the space of two weeks
- Total of 461 unique users
- Leaderboard to encourage users



(Iyyer et al., ACL 2014)

Deep Q-learning

- Q-learning
  - Estimate $Q(s, a)$: expected future reward starting from (s,a)
  - Reward function: quiz bowl scoring

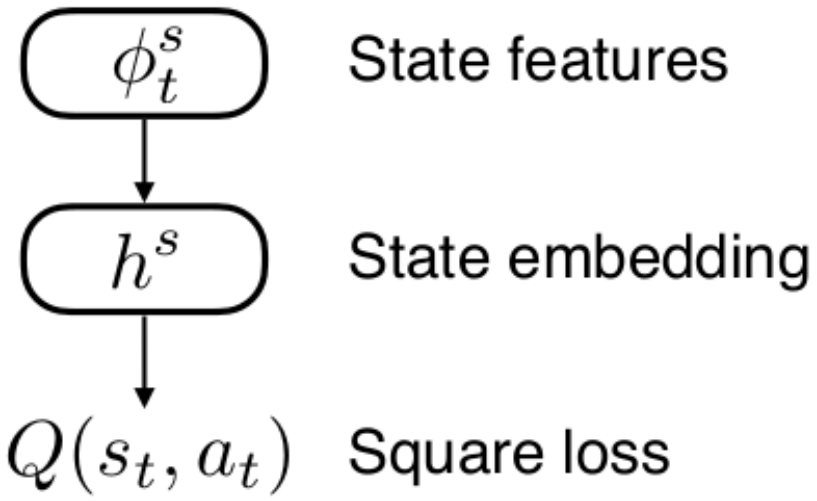
After the death of this man, The Liberators' Civil War was fought in

0.11 Mithridates
0.09 Julius Caesar
0.07 Sulia

Q-function →

What is the final score I can get if I BUZZ / WAIT now

- Deep Q-learning
  - Use a neural network to approximate the Q-function



(He et al., ICML 2016)

Opponent as part of the world

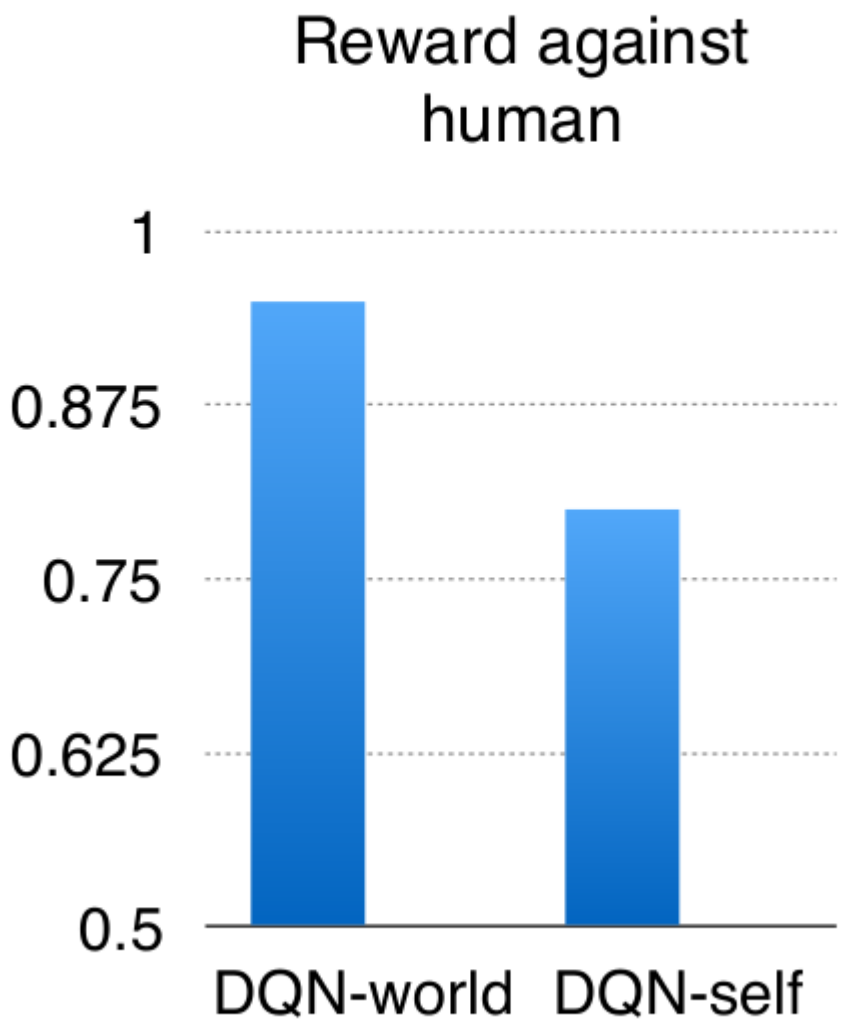
DQN-self:

imitate the expert: buzz whenever the current prediction is correct

DQN-world:

reinforcement learning from long term feedback

Interacting with opponent helps learn!

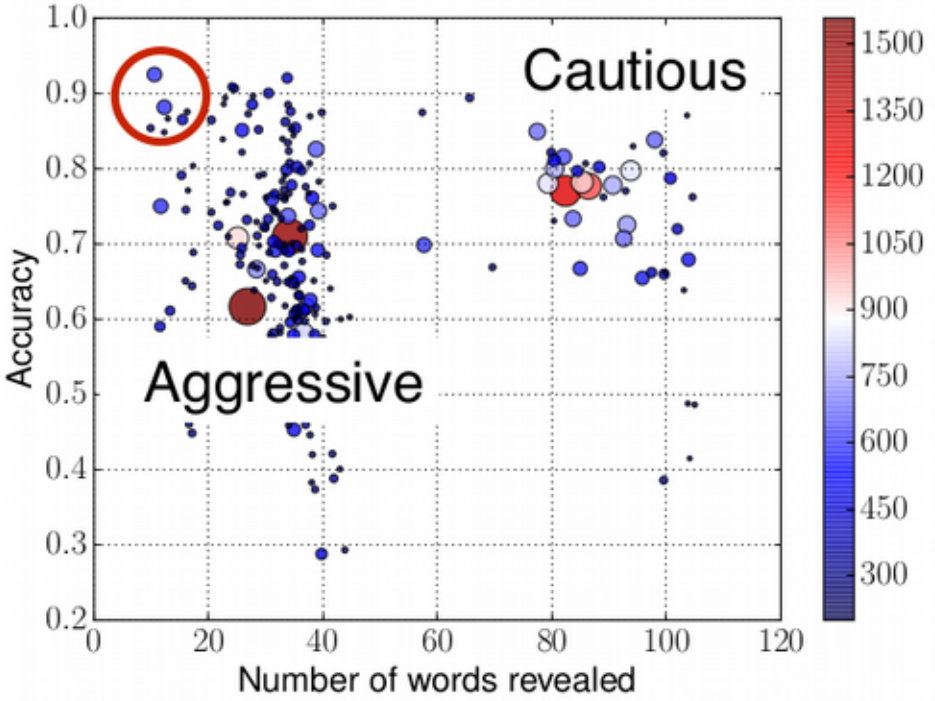


(He et al., ICML 2016)

Exploit the opponent?

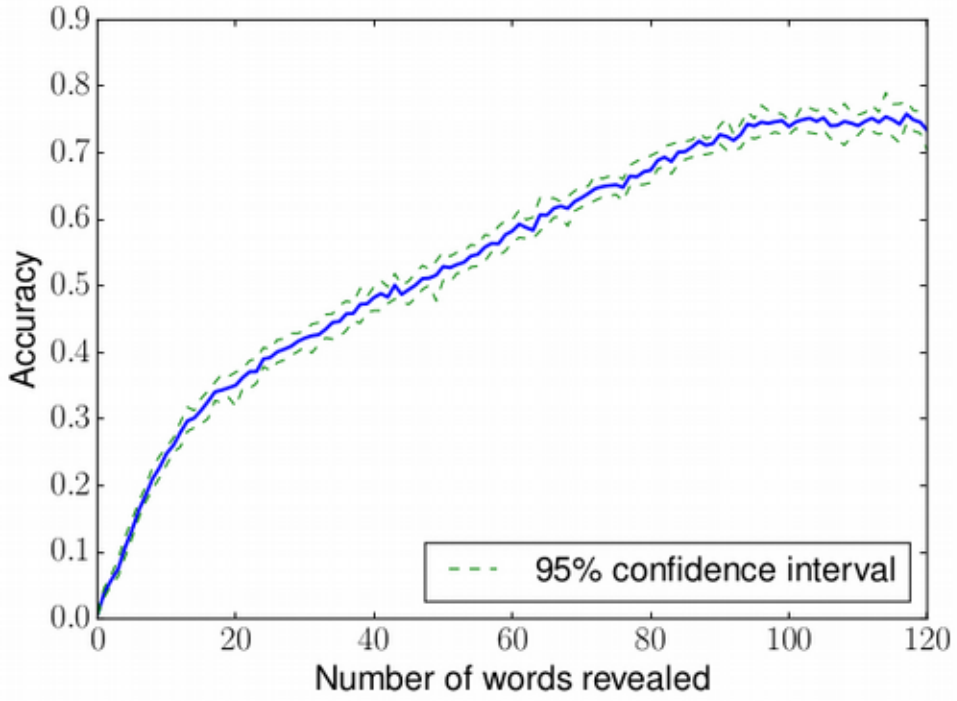
Human performance

Fast and good



Machine performance

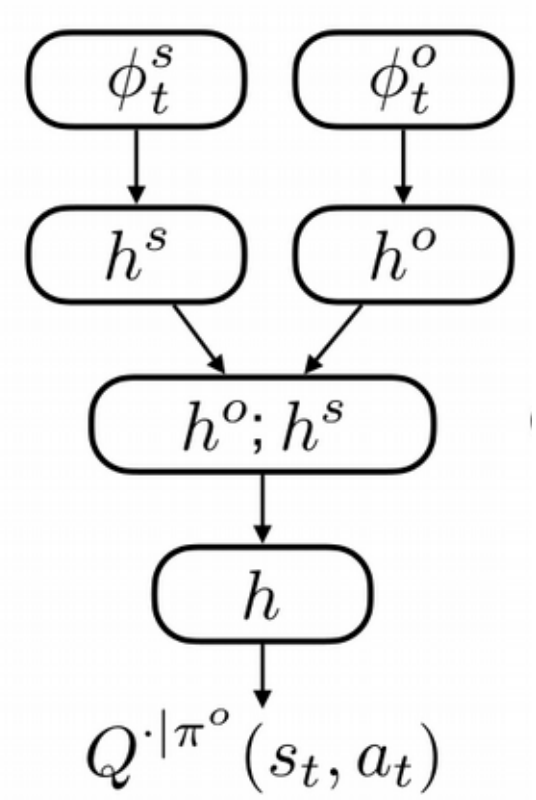
More words seen, higher accuracy



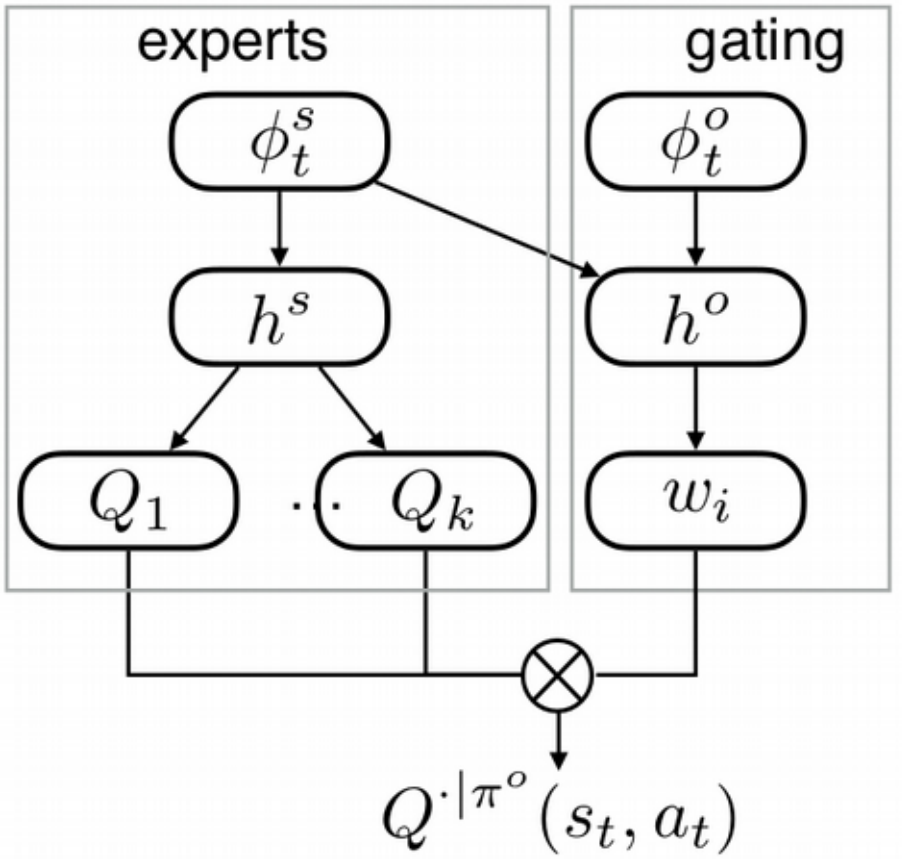
(He et al., ICML 2016)

How to incorporate opponent?

Concatenation

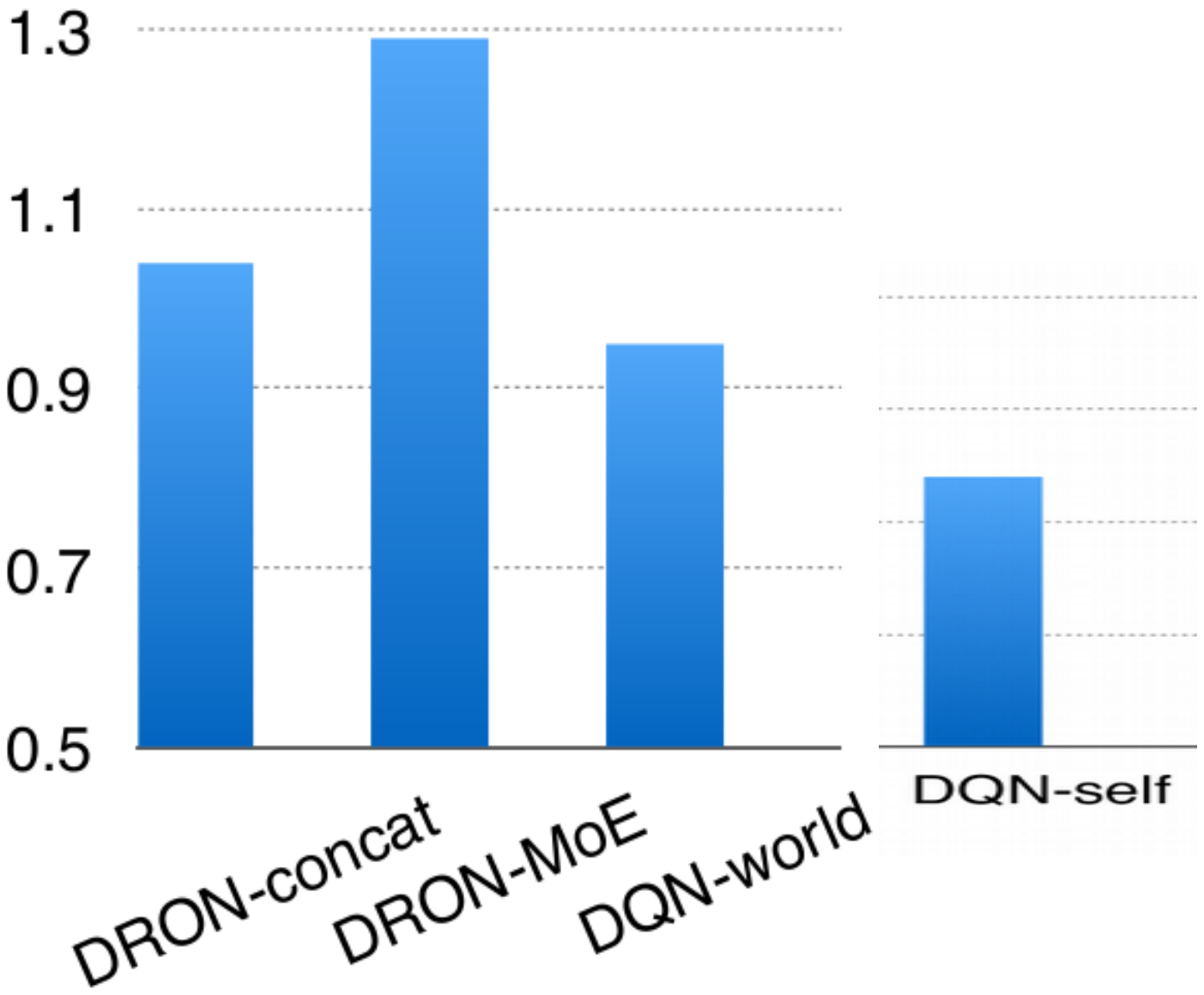


Mixture-of-Experts



(He et al., ICML 2016)

Overall reward against human players



(He et al., ICML 2016)

Students

Leonardo Claudino

Alvin Grissom II

He He

Mohit Iyyer

Sudha Rao

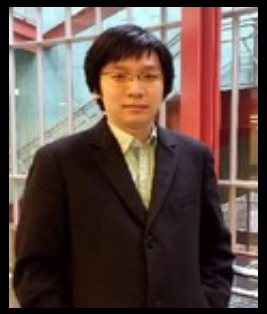
Amr Sharaf

Other Collaborators

Alekh Agarwal

Jordan Boyd-Graber

Alina Beygelzimmer



Kai-Wei Chang

Akshay Krishnamurthy

John Langford

Paul Mineiro

Stéphane Ross

Discussion

- Learning from interaction requires:
  - Build a system and let (simulated) people us it...
  - They provide some feedback (implicit/explicit)...
  - System improves performance over time...

- Open questions:
  - How can we best use offline-acquired data?
  - Can we (substantially) reduce the amount of interaction required to learn something interesting?
  - How can we convert implicit feedback into an implicit reward signal?

THANKS! Questions?