QALF: CONVERSATIONALLY BUILDING A RELATION EXTRACTOR IN 10 MINUTES VIA QUESTION ANSWERING

Braden Hancock, Robin Jia, Percy Liang, Christopher Ré

DAWN Retreat

19 Mar 2018

HIGHER BANDWIDTH SUPERVISION

Traditional Supervision

One bit of information per input…
(a binary label)

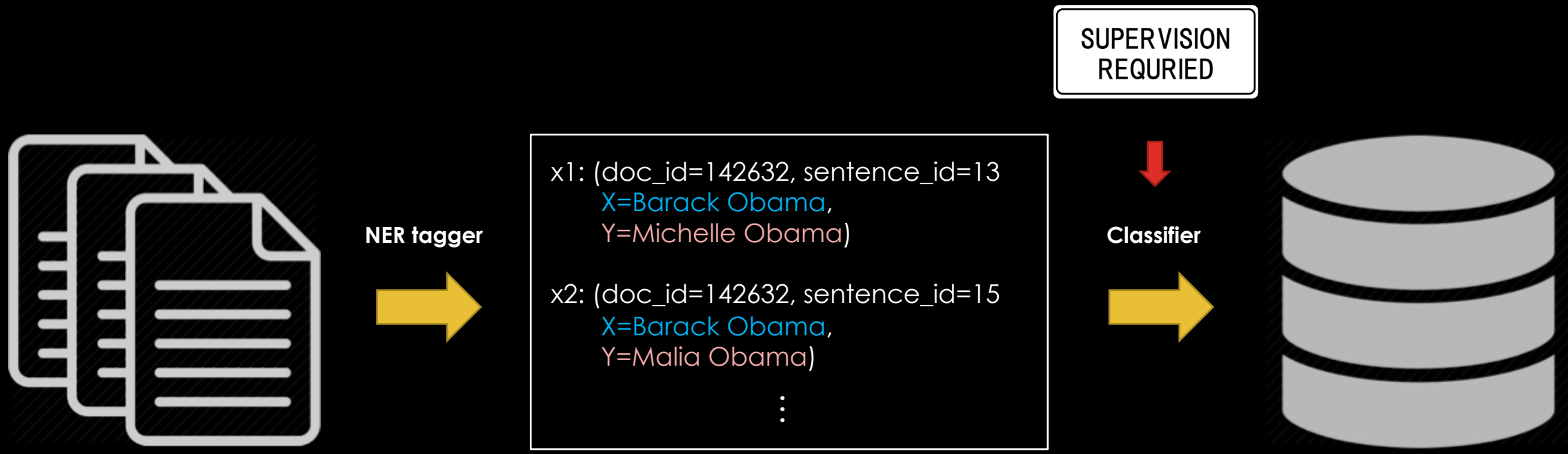
Weak Supervision

Many bits of information per input…
(very many weak labels)

Blog: https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/

EXAMPLE APPLICATION: KBC

GOAL: Build a knowledge base of spouses mentioned in the news.



LABELING FUNCTIONS (LFs)

Candidate (x1)

"Barack Obama and his wife Michelle attended their daughter's graduation hand-in-hand."

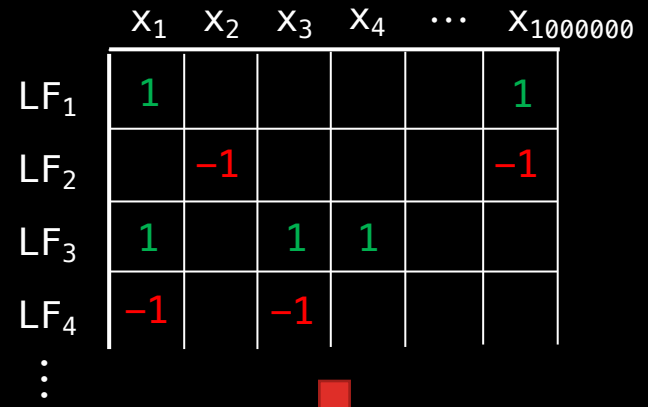
Labeling Functions

```
def LF1(x):
    return 1 if (re.search("his wife"), x.between)
        else 0

def LF2(x):
    return -1 if last_name(x.X) != last_name(x.Y)
        else 0

def LF3(x):
    return 1 if (x.X, x.Y) in set(known_spouses)
        else 0
```

Noisy Labels



Probabilistic Training Set

PRIOR WORK: BABBLE LABBLE

Candidate

Bob and his wife Alice visited Stanford University last Thursday.

Is X the spouse of Y?

Label

Explanation Why do you think so?

Because the words 'his wife' are immediately to the left of Y.

PRIOR WORK: BABBLE LABBLE

Explanation

"Yes, Because the words 'his wife' are immediately to the left of Y."

Babble Labble

Labeling Function

```
def LF1(x):
    return 1 if re.search(
     "his wife", left(x.Y)) else 0
```

Result

We achieved the same F1 score with classifiers trained using up to **100×** fewer inputs when users provide explanations instead of labels!

"On Bob and Alice's honeymoon, it rained all week."

Prefers Semantics
(e.g., "it says so")

Yes, because it says they went on a honeymoon together.

#$*@&!

Do you think X is the spouse of Y? Why?

Do you mean "because the word 'honeymoon' occurs within 10 characters to the right of person1 or person 2?"

Prefers Syntax
(e.g., keywords, word distance, capitalization, etc.)

RAISING THE LEVEL OF ABSTRACTION

QALF	Supervise at the level of ***semantics*** with **language**
Babble Labble	Supervise at the level of **syntax** with **language**
Snorkel	Supervise at the level of **syntax** with **code**
Traditional	Supervise at the level of **examples** with **labels**

QA MODELS

Document

Santa Cruz (Spanish: Holy Cross) is the county seat and largest city of Santa Cruz County, California. As of 2013 the U.S. Census Bureau estimated Santa Cruz's population at 62,864.

Santa Cruz is known for its moderate climate, natural environment, coastline, redwood forests, alternative community lifestyles, and socially liberal leanings. It is also home to the University of California, Santa Cruz, a premier research institution and educational hub, as well as the Santa Cruz Beach Boardwalk, an oceanfront amusement park operating continuously since 1907.

Questions	Answers
What state is Santa Cruz in?	California (char. 92-101)
What university is located in Santa Cruz?	University of California, Santa Cruz (char. 366-401)
When did the Santa Cruz Beach Boardwalk first open?	1907 (char. 557-561)

Strengths

- Implicit typing (Who -> people, Where -> places, etc.)
- Improved robustness to paraphrases via word embeddings
- Generic relation recognizer ("X is the R of Y" → (X, Y) ∈ R)

THE QA ARMS RACE

- SQuAD Leaderboard
  - 17 of the top 20 scores are from 2018
  - Leaderboard populated with top companies and universities around the world

SQuAD
The Stanford Question Answering Dataset

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Jan 22, 2018	Hybrid AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.482	**89.281**
1 Mar 06, 2018	QANet (ensemble) *Google Brain & CMU*	82.744	89.045
1 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (ensemble model) *Microsoft Research Asia & NUDT*	**82.849**	88.764
2 Feb 02, 2018	Reinforced Mnemonic Reader (ensemble model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798	82.283	88.533
2 Jan 03, 2018	r-net+ (ensemble) *Microsoft Research Asia*	82.650	88.493
2 Jan 05, 2018	SLQA+ (ensemble) *Alibaba iDST NLP*	82.440	88.607
3 Dec 17, 2017	r-net (ensemble) *Microsoft Research Asia* http://aka.ms/rnet	82.136	88.126
3 Dec 22, 2017	AttentionReader+ (ensemble) *Tencent DPDAC NLP*	81.790	88.163
4 Feb 27, 2018	QANet (single model) *Google Brain & CMU*	80.929	87.773
4 Nov 17, 2017	BiDAF + Self Attention + ELMo (ensemble) *Allen Institute for Artificial Intelligence*	81.003	87.432

QALF INTERFACE

Candidate

John Stamos and new bride Caitlin McHugh celebrated their honeymoon at Disney World Resort.

Is X the spouse of Y?

Label

Question

Write a question that contains X whose answer is Y.

Who celebrated a honeymoon with X?

MAKING QA-LFs

Polarity Positive

Question “Who is X’s daughter?”

User-Provided

Auto-Generated

QA-based LF

```
def QA_LF1(x):
    question = “Who celebrated a honeymoon with” + str(x.X) + “?”
    return 1 if QAModel(x.doc, question, x.Y) > 0.8 else 0
```

APPLYING QA-LFs

Intuition

If the QA Model is very confident that Y is the answer to the question based on X, return a label of the user-provided polarity; otherwise, abstain.

QA-based LF

```
def QA_LF1(x):
    question = "Who celebrated a honeymoon with" + str(x.X) + "?"
    return 1 if QAModel(x.doc, question, x.Y) > 0.8 else 0
```

Candidates

```
x1 (Aaron, Alice)    QAModel(x1.doc, "Who celebrated a honeymoon with Aaron?", Alice)
x2 (Billy, Betty)    QAModel(x2.doc, "Who celebrated a honeymoon with Billy?", Betty)
x3 (Chris, Carol)    QAModel(x3.doc, "Who celebrated a honeymoon with Chris?", Carol)
```

MAKING QA-LFs

(+) Positive Questions

"Who is the spouse of X?"

"Who has a child with X?"

"Who tied the knot with X?

(-) Negative Questions

"Who works with X?"

"Who is X's daughter?"

"Who played X in a movie?"

QALF

= Possible X

"Benny promised Jane he wouldn't miss their daughter Amy's recital this year."

"Molly Brown and Bob Johnson went over their notes on the way to the meeting."

"Jenny and Dave, her boyfriend of 10 years, finally tied the knot this weekend."

NER tagger

x1: (doc_id=142632, sentence_id=13
X=Benny,
Y=Jane)

x2: (doc_id=142632, sentence_id=15
X=Benny,
Y=Amy)

⋮

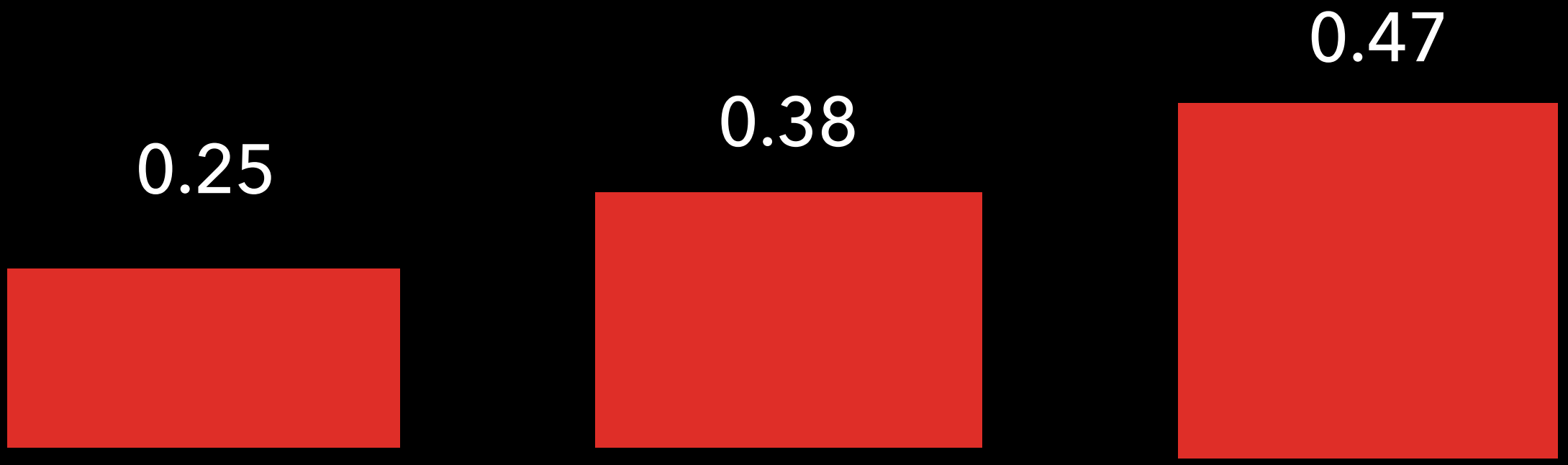
Classifier

Documents **Candidates** **Knowledge Base**

EXPERIMENTAL RESULTS

(F1 scores)

Task:
Build a KB of spouses



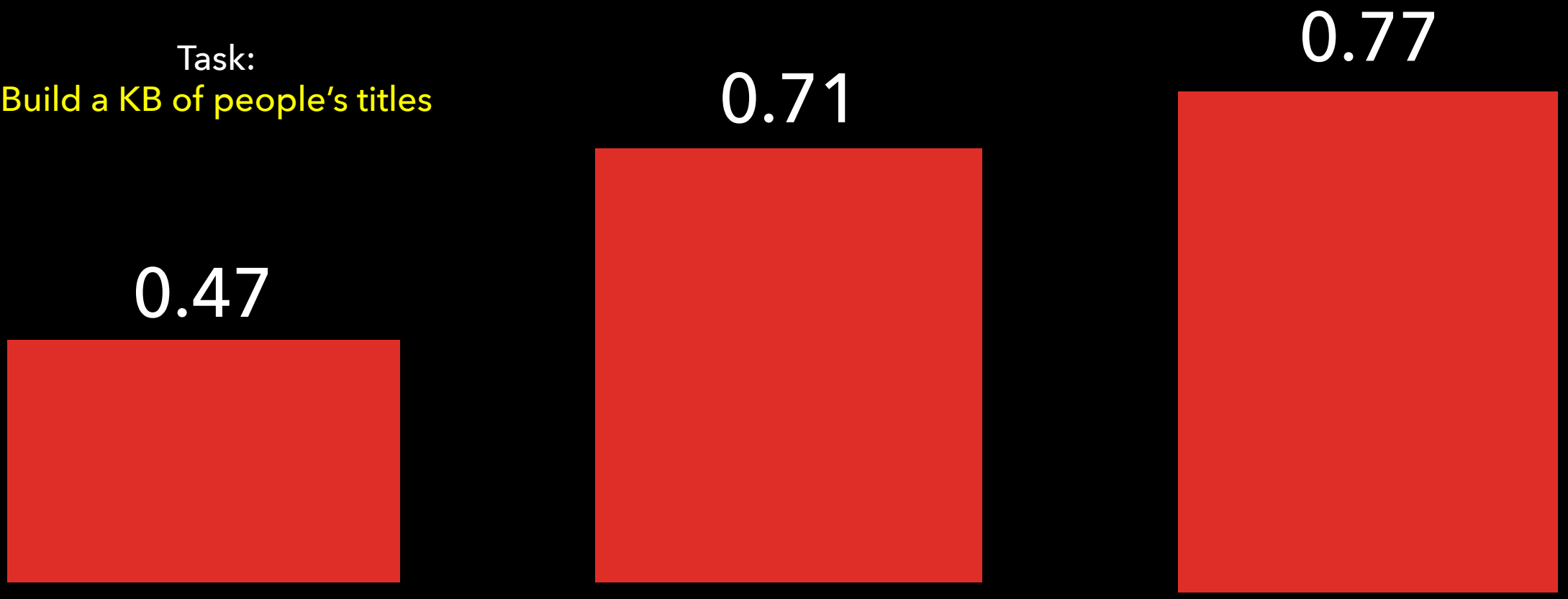
Ask the question on the test set
"Who is the spouse of X?"

Use the question to make a training set from unlabeled data

Aggregate the noisy labels from multiple LFs

EXPERIMENTAL RESULTS

Task:
Build a KB of people's titles



Ask the question on the test set
"What is X's title?"

Use the question to make a training set from unlabeled data

Aggregate the noisy labels from multiple LFs

CONCLUSION

- LFs allow higher bandwidth supervision

- QA LFs allow semantic-level supervision

Yes, because it says they went on a honeymoon together.

- Initial results suggest big wins (up to 30 F1 points) over direct QA model use