Low-resource Deep Entity Resolution with Transfer and Active Learning

Kun Qian

IBM Research

Joint work with

Jungo Kasai, Sairam Gurajada, Yunyao Li, Lucian Popa

Published at ACL 2019

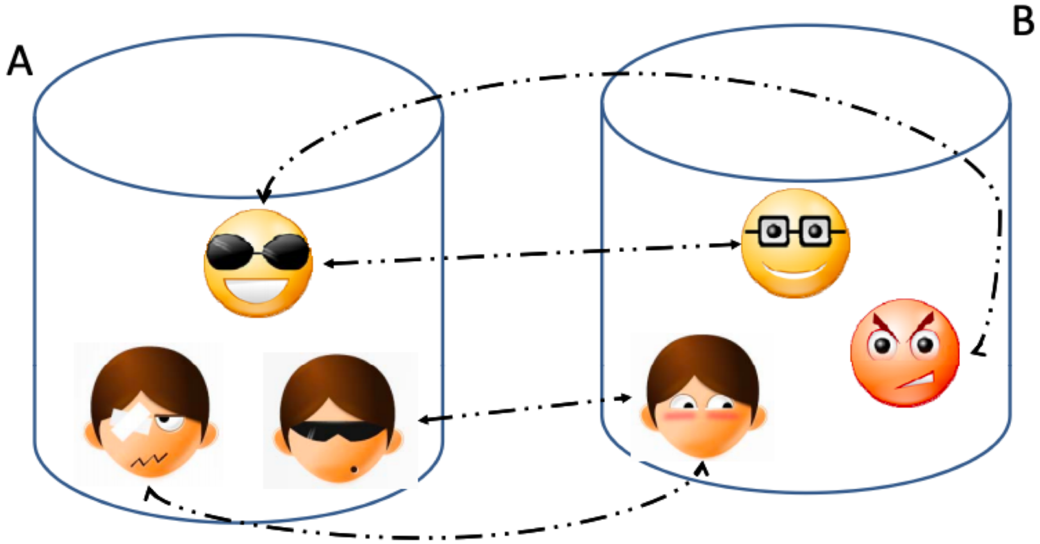
2019 CROSS Research Symposium & Oktoberfest

Contributions

- Our deep entity resolution model for easy transfer learning.
- Active learning algorithm for adapting to the target scenario.
- Extensive experiments over benchmarks and achieve SOTA performance **while using an order of magnitude fewer labels**.

Background & Introduction

Entity Resolution (ER)



Source: http://users.umiacs.umd.edu/~getoor/Tutorials/ER_VLDB2012.pdf

Entity Resolution (ER)

Active Learning for Large-Scale Entity Resolution

Full Text: PDF Get this Article

Authors: Kun Qian IBM Research - Almaden, San Jose, CA,
Lucian Popa IBM Research - Almaden, San Jose, CA,
Prithviraj Sen IBM Research - Almaden, San Jose, CA,

Published in:

· Proceeding
CIKM '17 Proceedings of the 2017 ACM on C
Information and Knowledge Management
Pages 1379-1388

Singapore, Singapore — November 06 - 10,
ACM New York, NY, USA ©2017
table of contents ISBN: 978-1-4503-4918-5
doi>10.1145/3132847.3132949

Semantic Scholar

Active Learning for Large-Scale Entity Resolution

Kun Qian, Lucian Popa, Prithviraj Sen • Published in CIKM 2017 • DOI: 10.1145/3132847.3132949

Entity resolution (ER) is the task of identifying different representations of the same real-world object across datasets. Designing and tuning ER algorithms is an error-prone, labor-intensive process, which can significantly benefit from data-driven, automated learning methods. Our focus is on "big data" scenarios where the primary challenges include 1) identifying, out of a potentially massive set, a small subset of informative examples to be labeled by the user, 2) using the labeled... CONTINUE READING

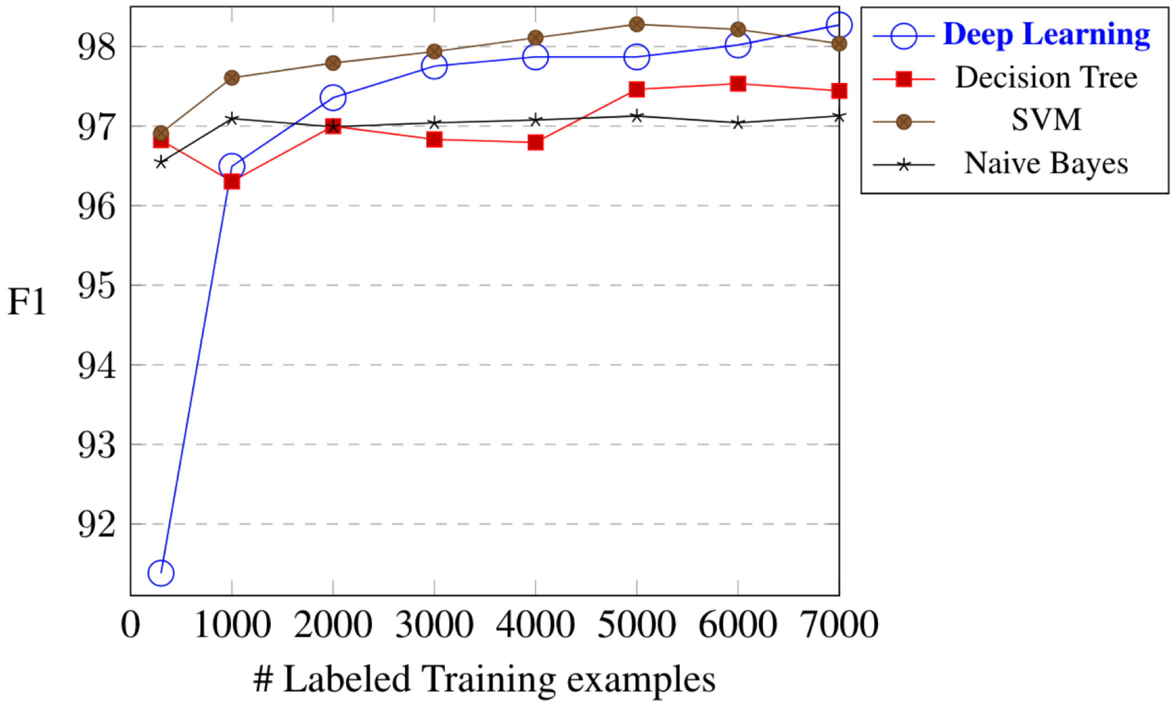
Entity Resolution (ER)

- Knowledge Base Creation, Text Mining (Zhao et al., 2014)
  - Papers, authors, ...
- Social Media Analysis (Campbell et al. 2016)
  - vs.
- Ironically also called...
  - Record Linkage (Fellegi and Sunter, 1969), Reference Reconciliation (Dong et al., 2005), Merge-Purge (Hernandez and Stolfo, 1995), and Entity Matching...

Deep Entity Resolution

- Rule-based vs. **ML-based**
- Recent Work (Ebraheem et al. 2018; Mudgal et al. 2018) proposed DL methods. No need to define features for every single ER scenario.
- However, ...

Deep ER is Data Hungry (DBLP-ACM)



Deep Entity Resolution

- Rule-based vs. **ML-based**
- Recent Work (Ebraheem et al. 2018; Mudgal et al. 2018) proposed DL methods. No need to define features for every single ER scenario.
- We establish **a novel framework for Low-resource Deep Entity Resolution**.

ER Problem

D_1, D_2 : Data record collections from Databases. Classify into matches or non-matches.

$$\langle e_1, e_2 \rangle, \; \forall e_1 \in D_1, e_2 \in D_2$$

Author	Title	Venue	Year
Alan Turing	Low resource entity resolution...	Proceedings of the Association...	1940
Turing, Alan M.	Low resource entity resolution...	ACL	1940

Two Steps in Entity Resolution

- **Blocking**
  - Reduce $D_1 \times D_2$
  - Eliminate obvious non-matches by blocking functions (predicates)
  - E.g. are they published in the same year?
- **Matching**
  - Classify the remaining pairs (candidate set).
  - Our work focuses on this step.

Our Framework for Low-resource Deep ER

- **Transfer Learning**
  - From a high resource scenario to a target that has no labeled data.
  - Simply take the parameters from source training.
- **Active Learning**
  - Further refine the model to the target by labeling a small number of informative samples.

Transfer Learning

Authors	Title	Venue
Alan Turing	Computing Machinery and Intelligence	Mind

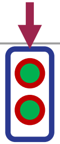
dblp
computer science bibliography

Google Scholar

Authors	Title	Venue
Turing, Alan M.	Computing Machinery and Intelligence	Mind

Architecture for record comparison

Authors	Title	Venue
Alan Turing	Computing Machinery and Intelligence	Mind



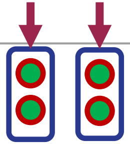
dblp computer science bibliography

Google Scholar

Authors	Title	Venue
Turing, Alan M.	Computing Machinery and Intelligence	Mind

Architecture for record comparison

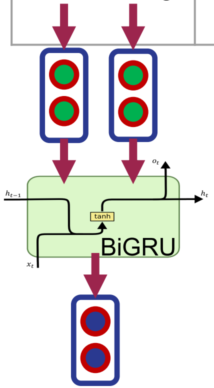
Authors	Title	Venue
Alan Turing	Computing Machinery and Intelligence	Mind



Authors	Title	Venue
Turing, Alan M.	Computing Machinery and Intelligence	Mind

Architecture for record comparison

Authors	Title	Venue
Alan Turing	Computing Machinery and Intelligence	Mind



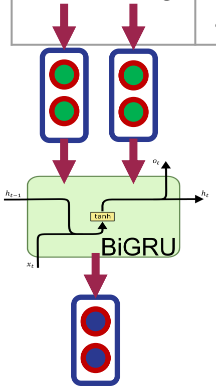
dblp computer science bibliography

Google Scholar

Authors	Title	Venue
Turing, Alan M.	Computing Machinery and Intelligence	Mind

Architecture for record comparison

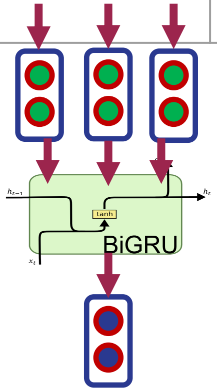
Authors	Title	Venue
Alan Turing	Computing Machinery and Intelligence	Mind



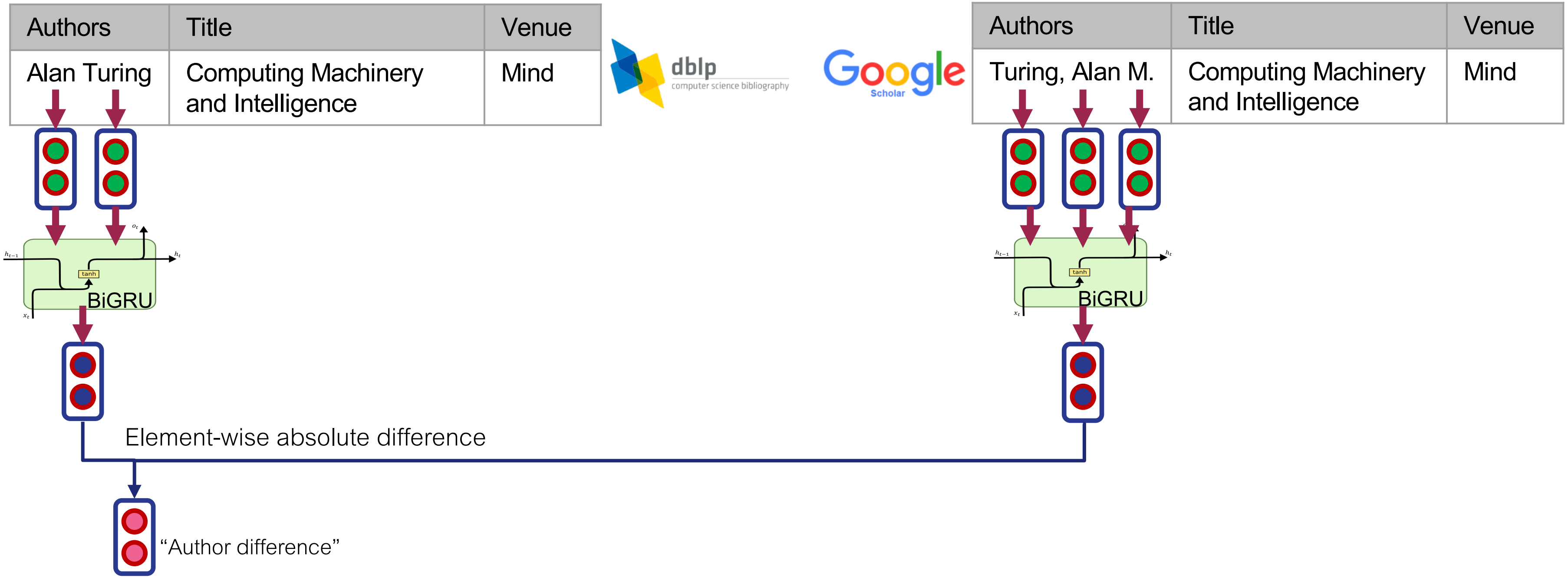
dblp computer science bibliography

Google Scholar

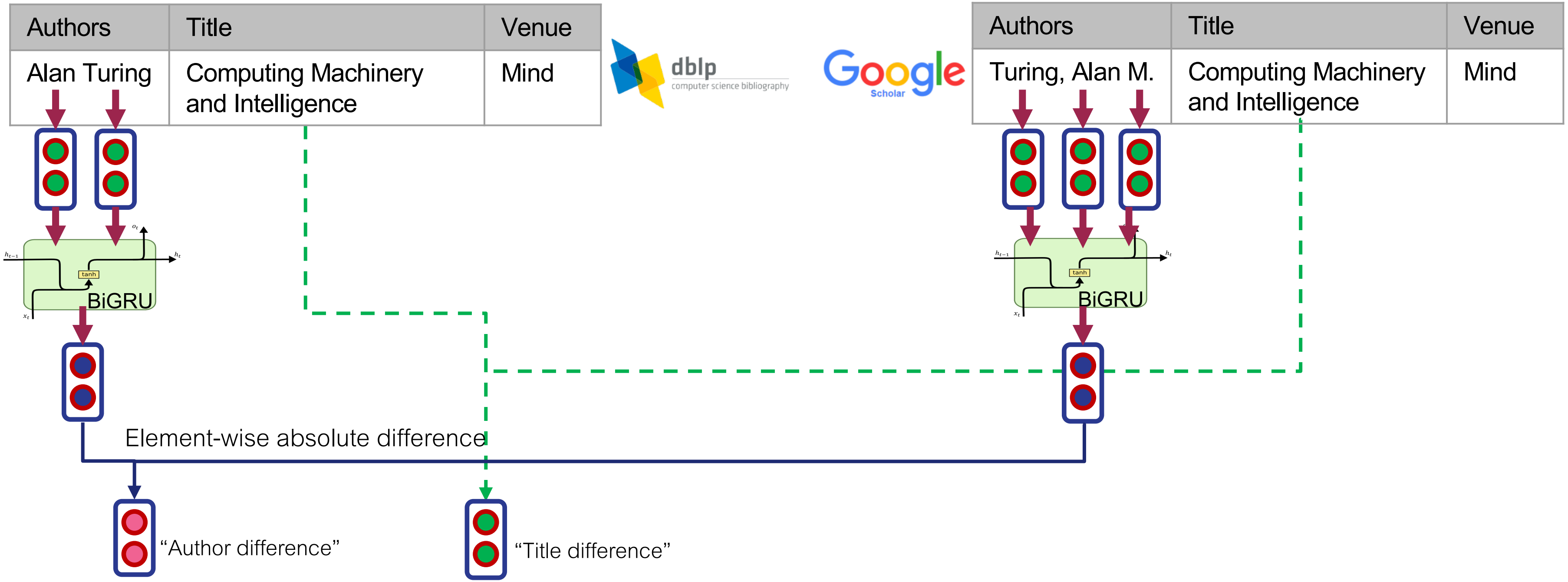
Authors	Title	Venue
Turing, Alan M.	Computing Machinery and Intelligence	Mind



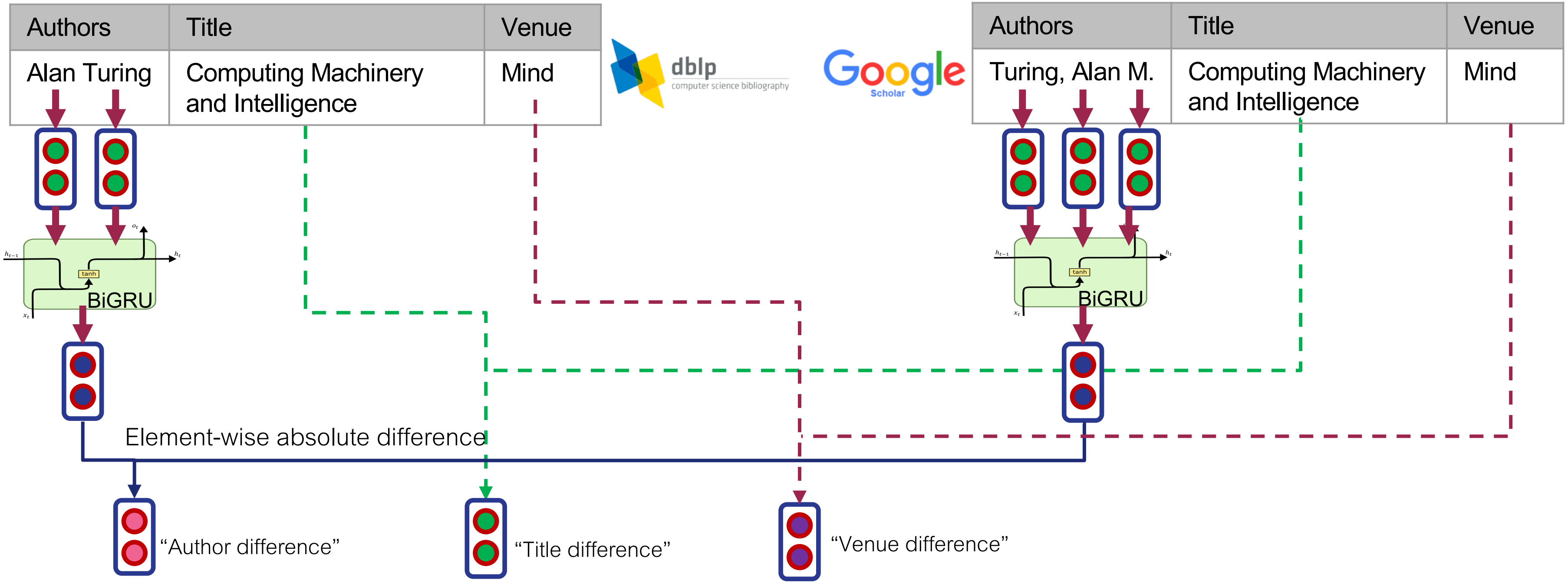
Architecture for record comparison



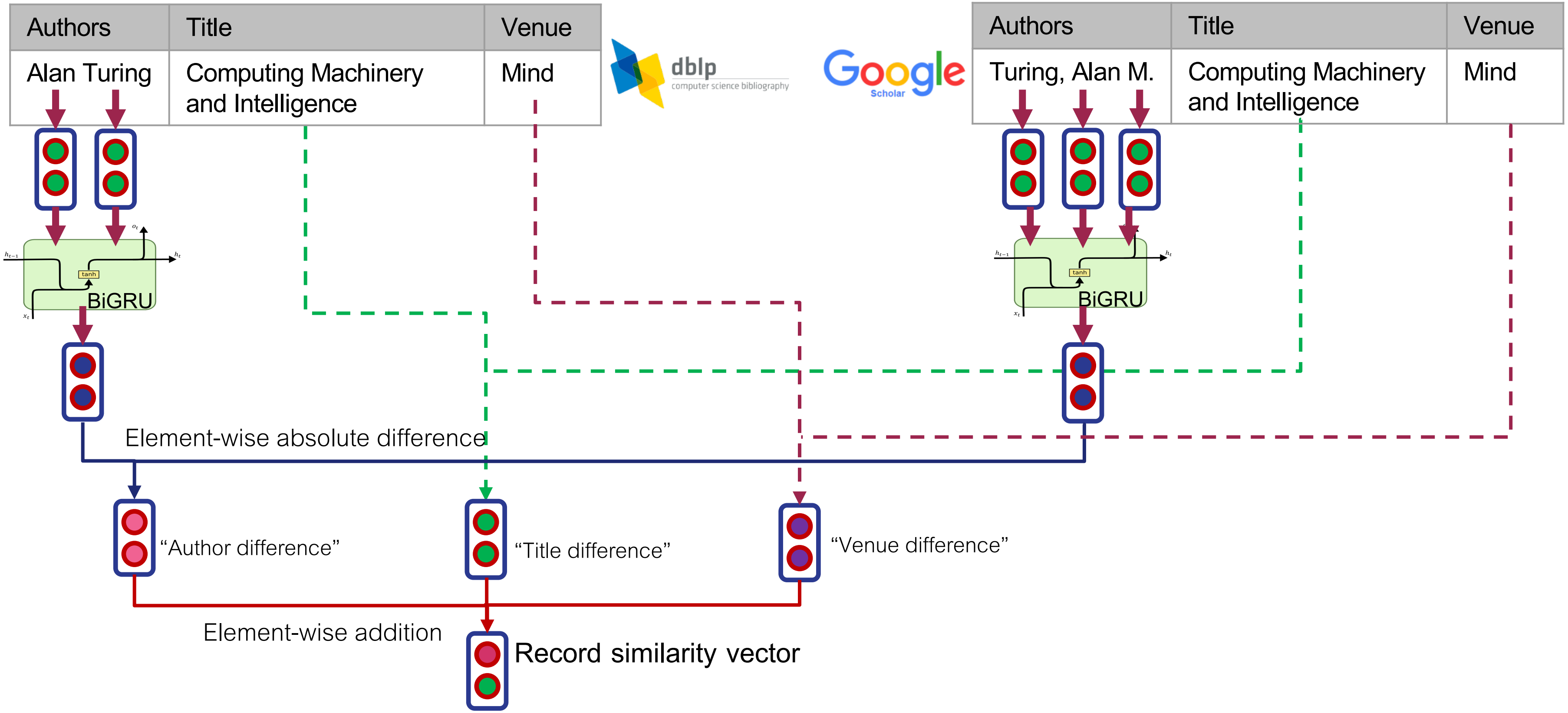
Architecture for record comparison



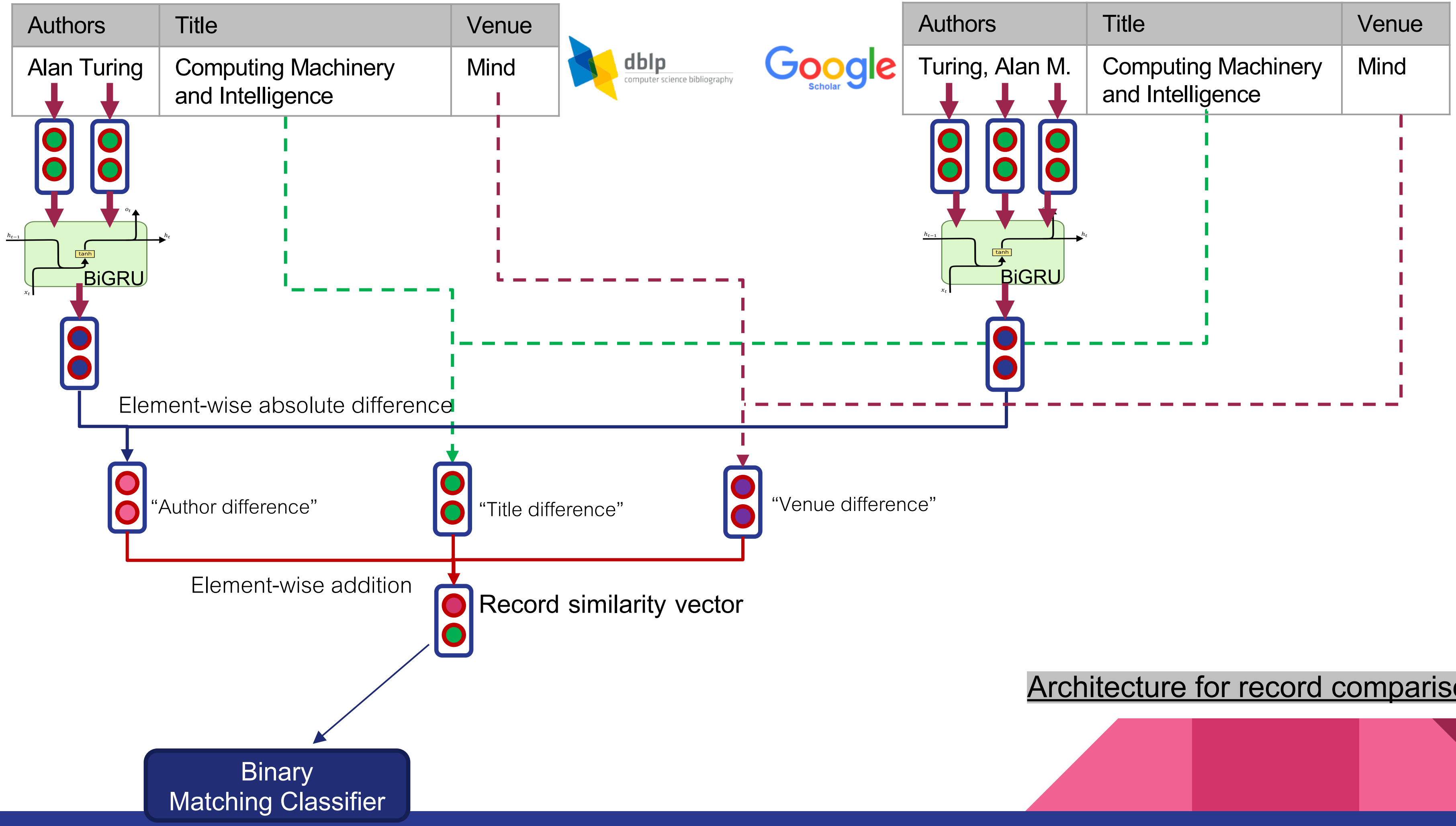
Architecture for record comparison



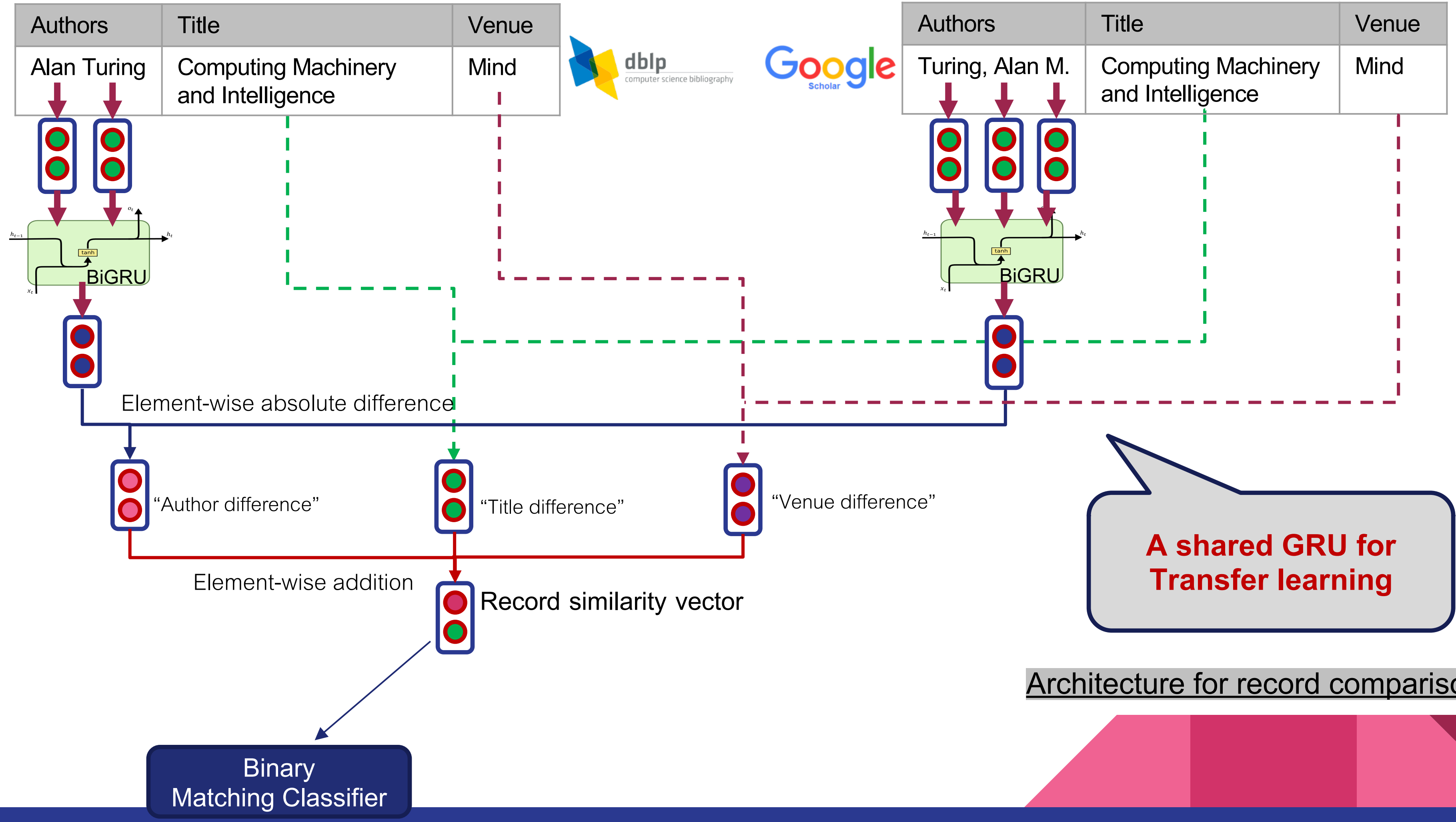
Architecture for record comparison



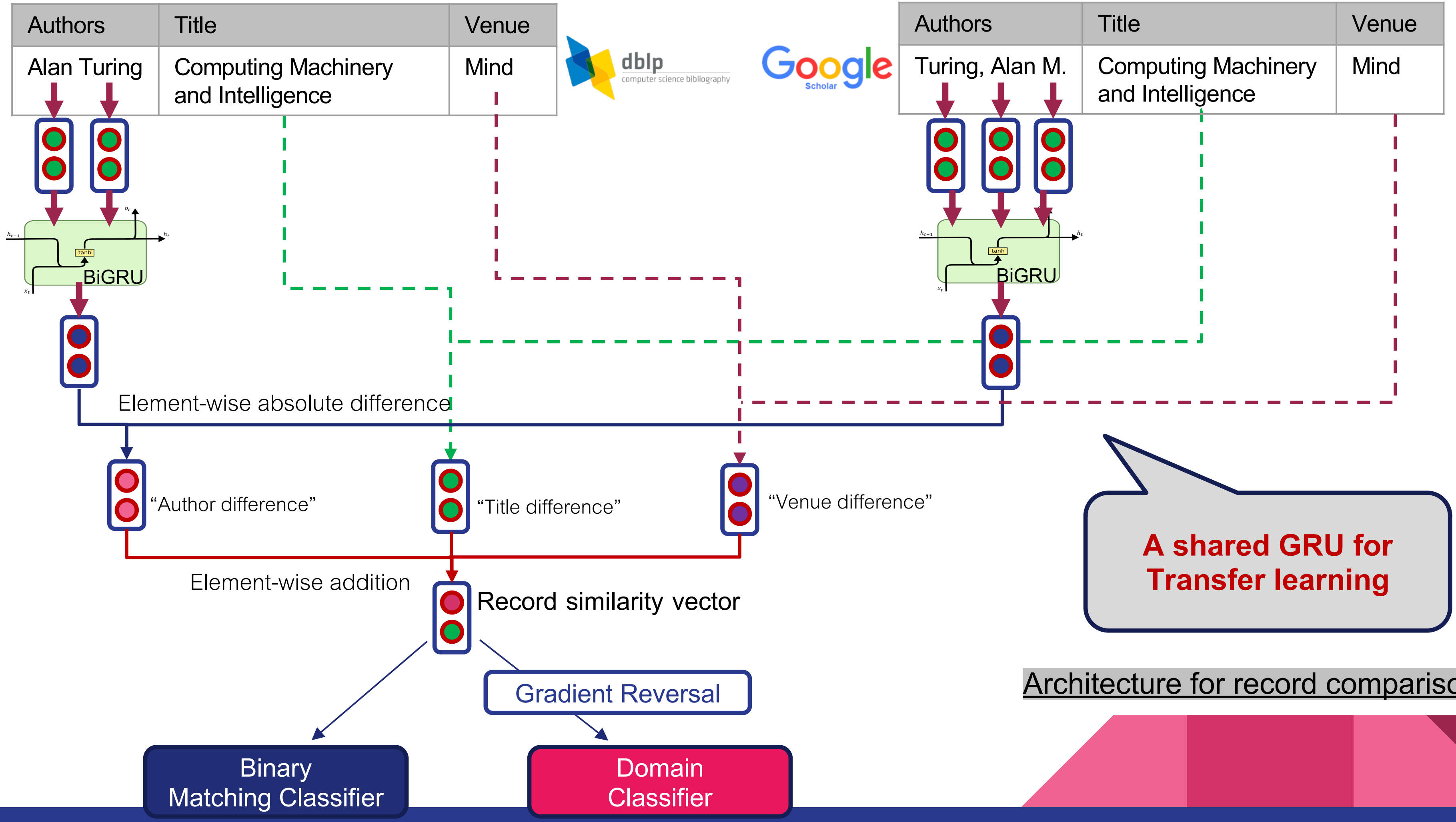
Architecture for record comparison



Architecture for record comparison



Architecture for record comparison



Architecture for record comparison

Transfer Learning Results

Transfer Learning Results

F1, Target	DBLP-ACM	DBLP-Scholar	Cora (**8 attributes!**)
Train on Source	92.32	41.03	38.30
+Adaptation	92.31	53.84	43.13
Train on Target			
Mudgal et al. 2018			

Transfer Learning Results

F1	DBLP-ACM	DBLP-Scholar	Cora (8 attributes!)
Train on Source	92.32	41.03	38.30
+Adaptation	**92.31**	53.84	43.13
Train on Target	**98.45**	**92.94**	**98.68**
Mudgal et al. 2018	98.4	93.3	--

- **Still a huge gap!!**

Active Learning

Active Learning

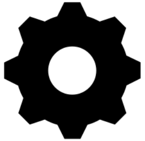
- Further adapt the transferred model by iteratively labeling "informative" examples from the target.
- 2 Major Problems
  - Problem 1: Need substantial positive/negative examples to tune the DL model.
  - Problem 2: Hard to improve **recall** in ER (Qian et al. 2017).
- We propose a principled way of sampling examples using entropy.

Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)

Our active learning strategy

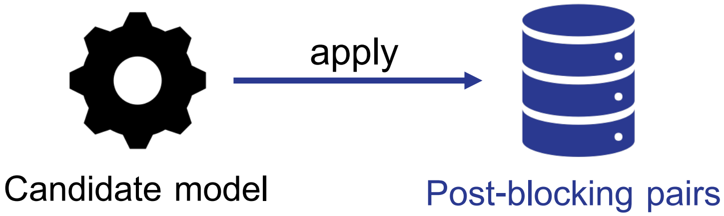
- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



Candidate model

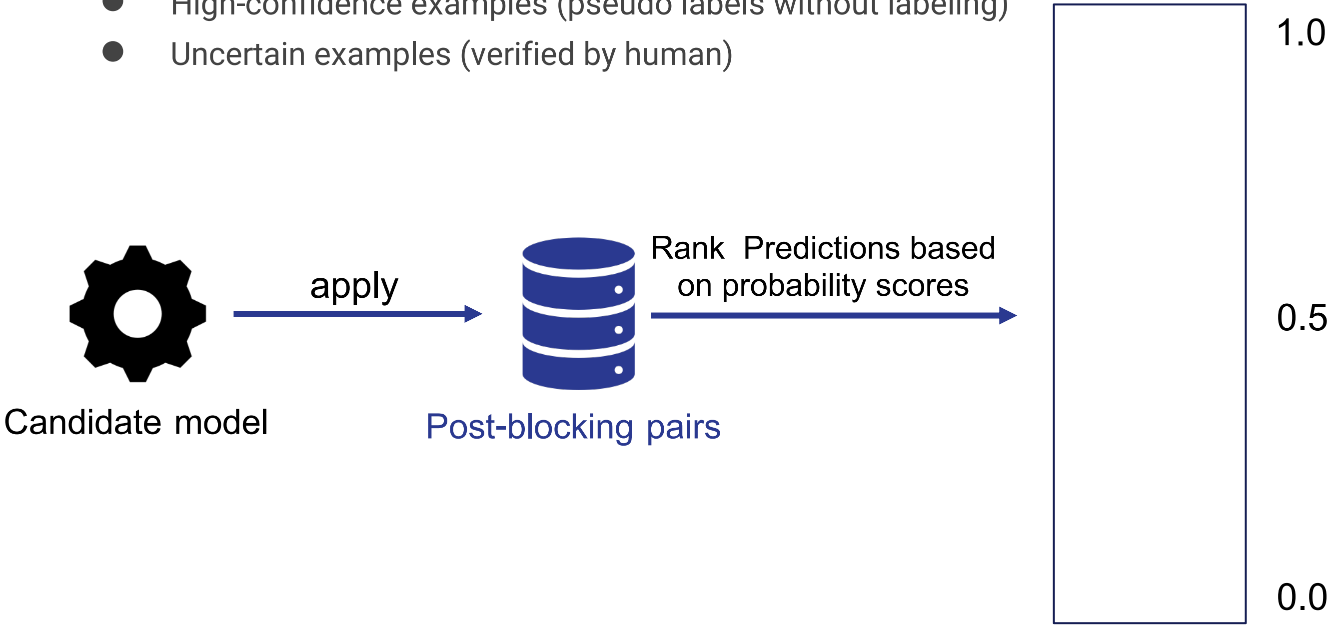
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



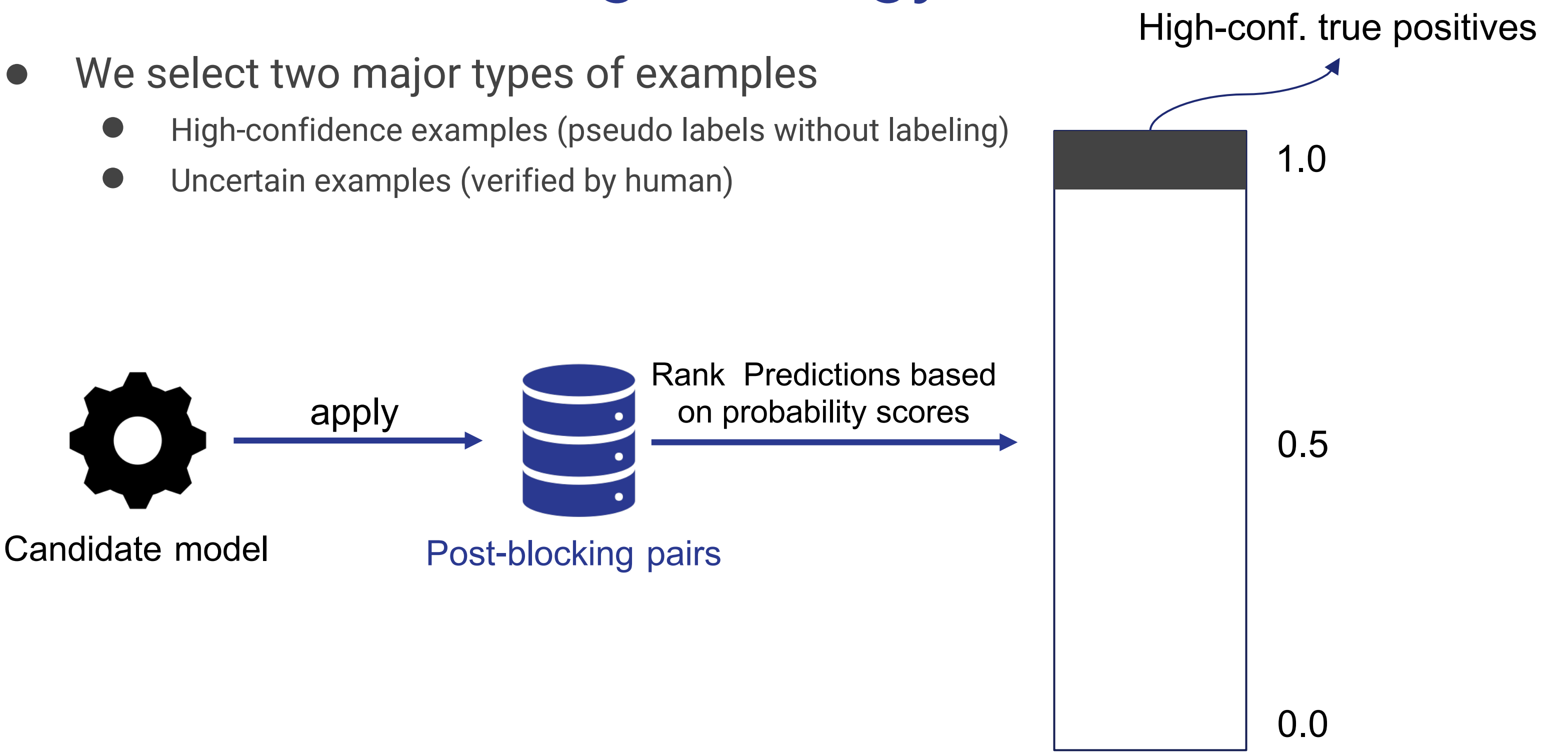
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



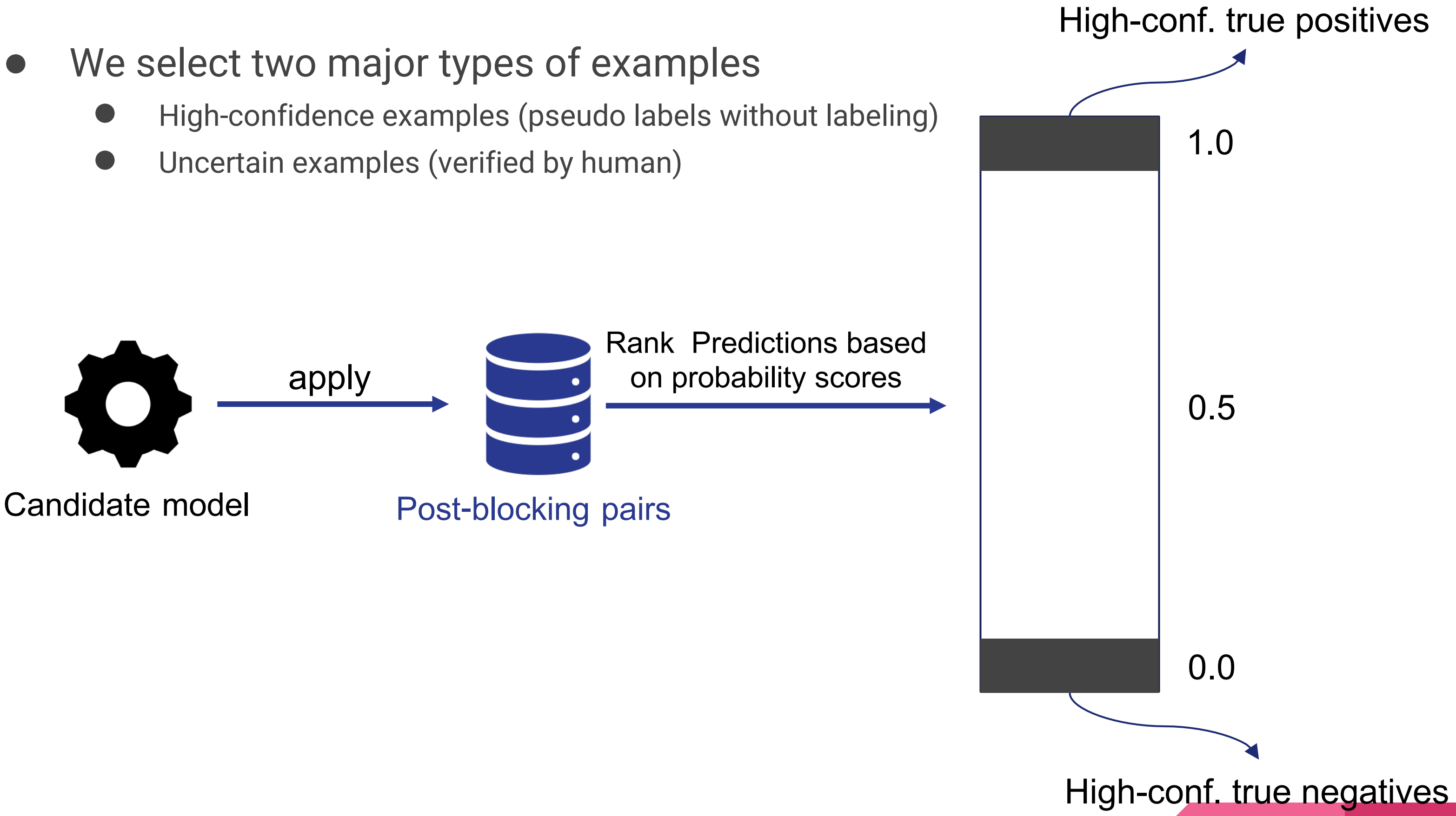
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



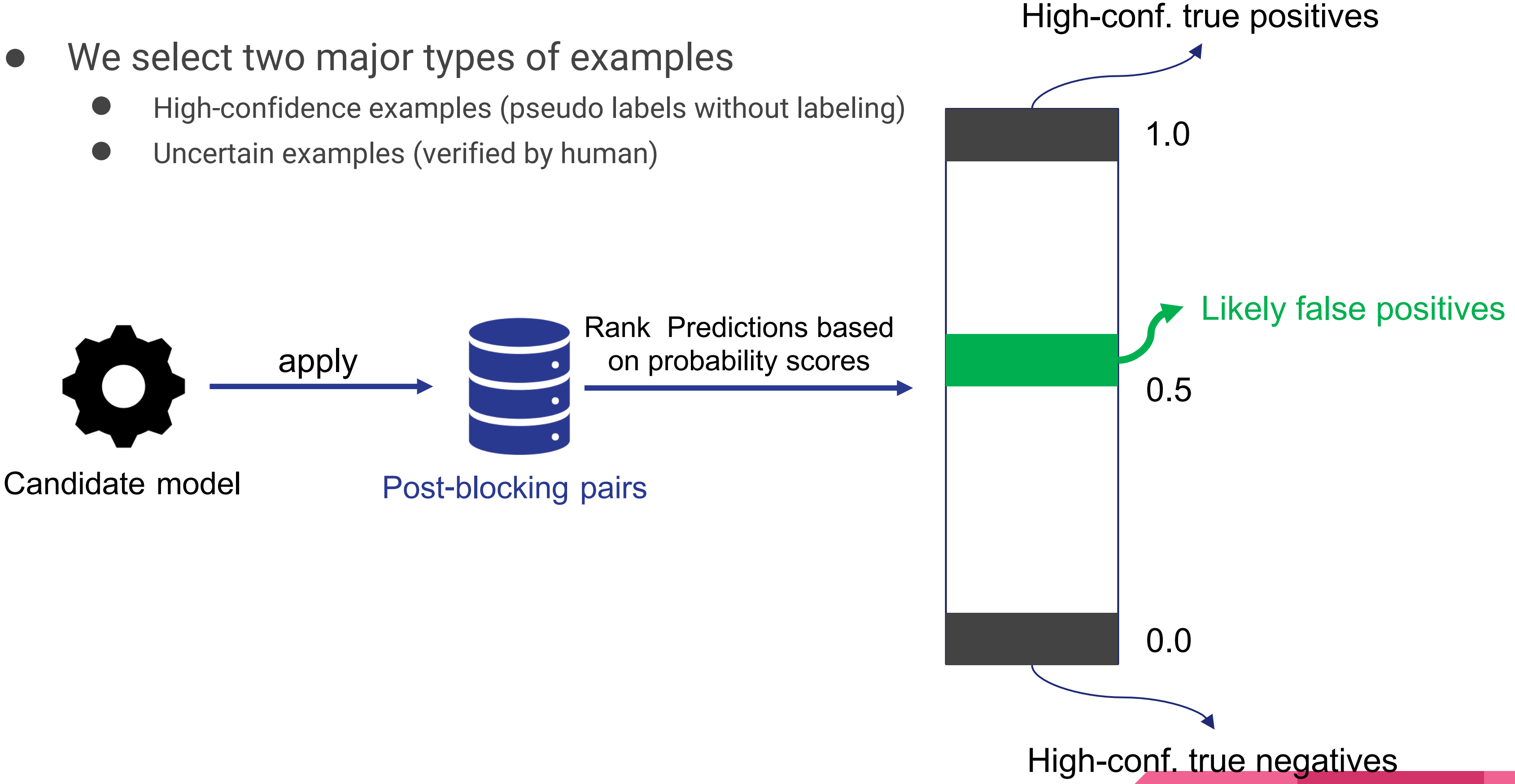
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



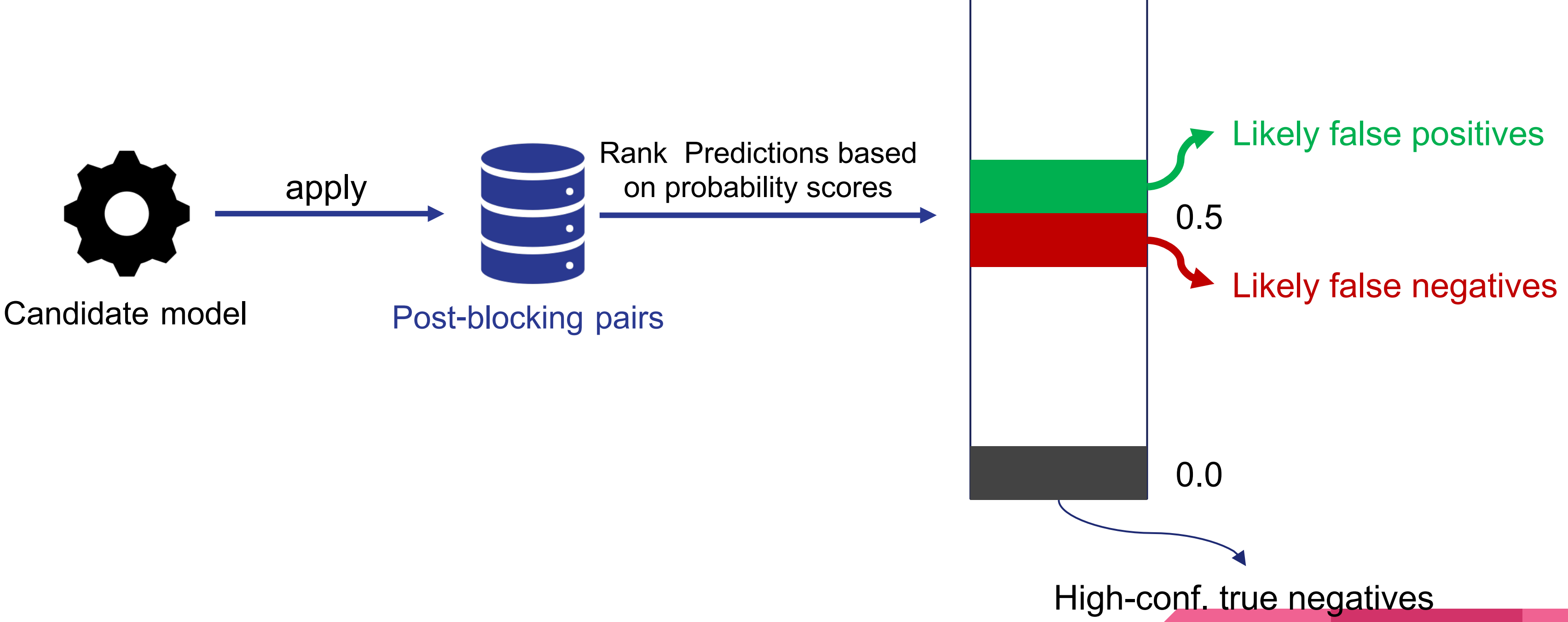
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



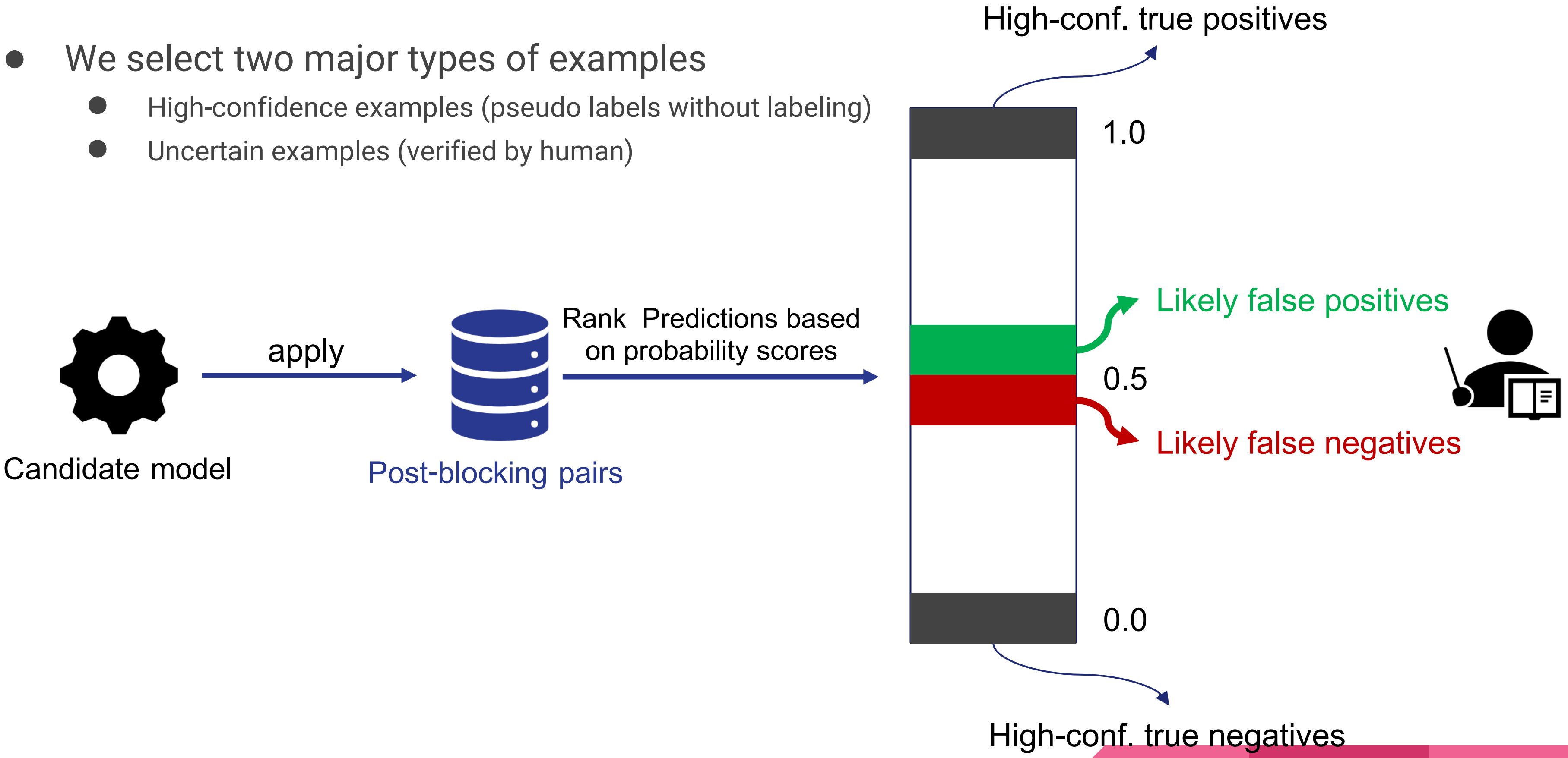
Our active learning strategy

- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



Our active learning strategy

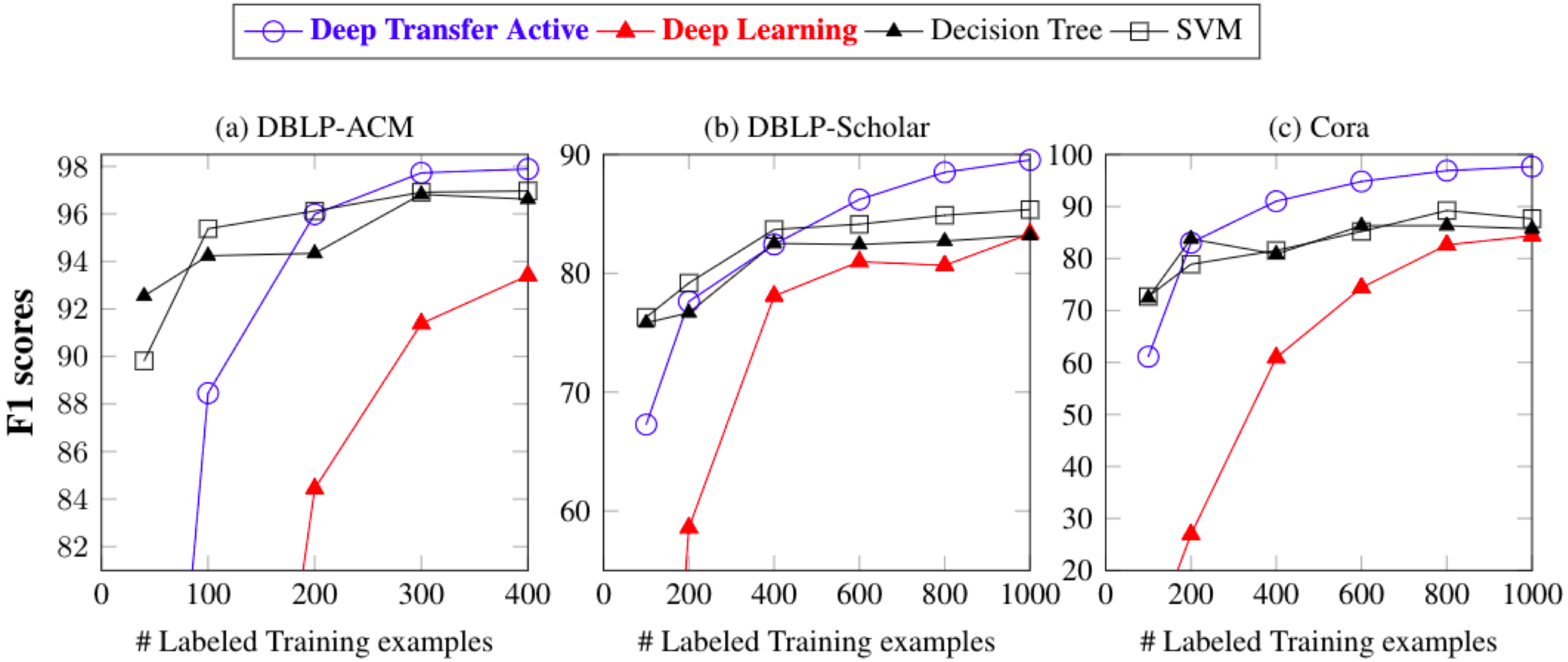
- We select two major types of examples
  - High-confidence examples (pseudo labels without labeling)
  - Uncertain examples (verified by human)



Selected Results

More results in ACL-19 paper

Transfer and Active Learning Results, Citations



Transfer and Active Learning Results, Citations

DBLP-Scholar, Method	# Labeled	F1
SVM	1000	85.36+−0.32
DL		83.33+−1.26
DL + Active		
DL + Transfer + Active		
SVM	Full, 17,223	
DL		

- Transfer from DBLP-ACM+Cora

Transfer and Active Learning Results, Citations

DBLP-Scholar, Method	# Labeled	F1
SVM	1000	85.36+−0.32
DL		83.33+−1.26
DL + Active		88.76+−0.76
DL + Transfer + Active		**89.54+−0.39**
SVM	Full, 17,223	
DL		

- Transfer from DBLP-ACM+Cora

Transfer and Active Learning Results, Citations

DBLP-Scholar, Method	# Labeled	F1
SVM	1000	85.36+−0.32
DL		83.33+−1.26
DL + Active		88.76+−0.76
DL + Transfer + Active		**89.54+−0.39**
SVM	Full, 17,223	88.56+−0.46
DL		92.94+−0.47

- Transfer from DBLP-ACM+Cora

Transfer and Active Learning Results, Citations

Cora, Method	# Labeled	F1
SVM	1000	87.66+−3.15
DL		84.35+−4.25
DL + Active		97.05+−0.64
DL + Transfer + Active		**97.68+−0.39**
SVM	Full, 30,000	95.39+−0.31
DL		98.68+−0.26

- Transfer from DBLP-Scholar+DBLP-ACM

More Genres, Restaurants

Fodors-Zagats	Target Train Size	F1
Train on Source	0	11.8
+Adaptation	0	70.1
+Active Labels	**100**	100.0
Train on Target	946	100.0
Mudgal et al. 2018	946	100.0

- Transfer from Zomato-Yelp

Conclusion and Future Work

- Our deep ER models yield competitive performance to SOTA learning-based methods with an order of magnitude less labels.
- Transfer learning alone doesn't suffice but yields stable performance when combined with active learning.
- DL can provide a unified data integration method.
- Our other recent work related to Entity Resolution
  - *Active Learning for Large-Scale Entity Resolution*. Qian, Popa, Sen. **CIKM 2017**.
  - *LUSTRE: An Interactive System for Entity Structured Representation and Variant Generation*. Qian, Bhutani, Li, Jagadish, Hernandez. **ICDE 2018**.
  - *Exploiting Structured Representation of Named Entities using Active Learning*. Bhutani, Qian, Li, Jagadish, Hernandez, Vasa. **COLING 2018**
  - *SystemER: A Human-in-the-loop System for Explainable Entity Resolution*. Qian, Popa, Sen. **VLDB 2019**
  - **Tutorial:** Learning-base Methods with Human-in-the-loop for Entity Resolution. **CIKM. 2019**

Conclusion and Future Work

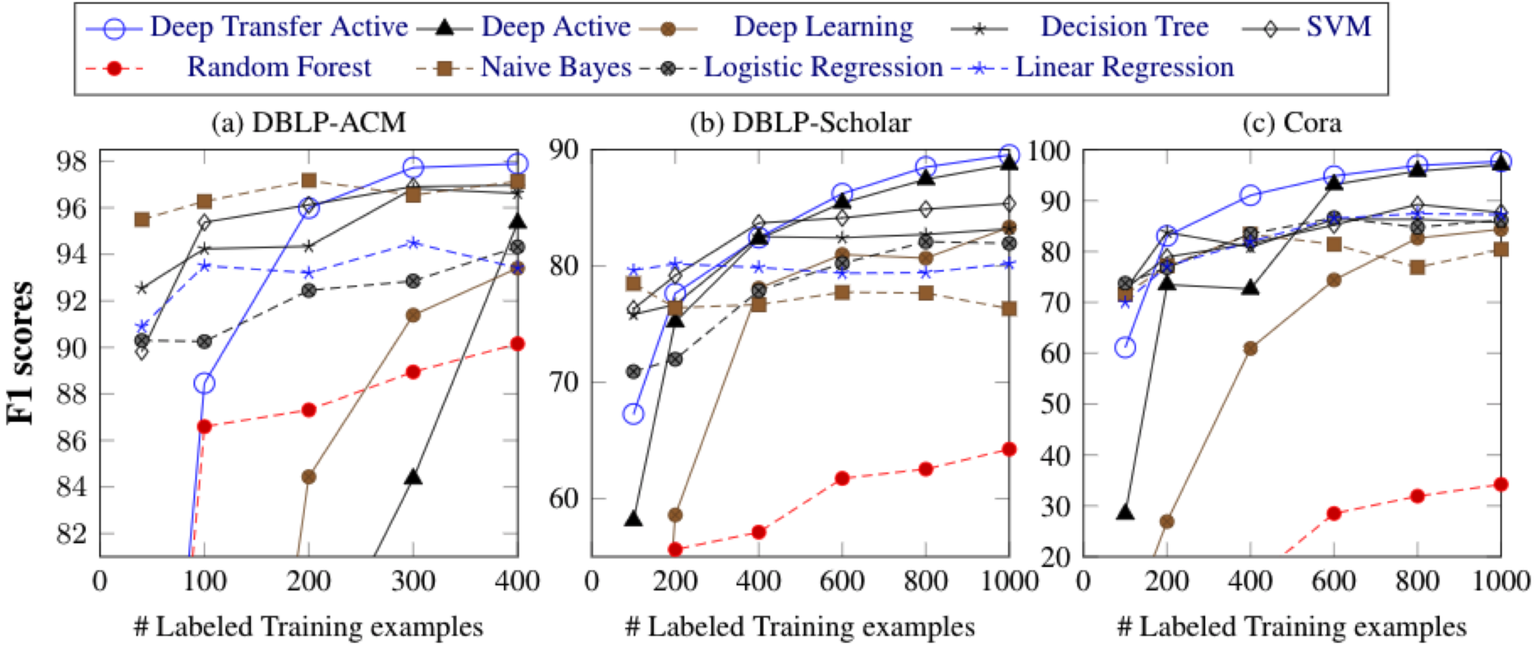
- Our deep ER models yield competitive performance to SOTA learning-based methods with an order of magnitude less labels.
- Transfer learning alone doesn’t suffice but yields stable performance when combined with active learning.
- DL can provide a unified data integration method.
- More scenarios, genres, and languages?

Acknowledgements

Sidharth Mudgal (code, data setups), Vamsi Meduri, and Phoebe Mulcaire

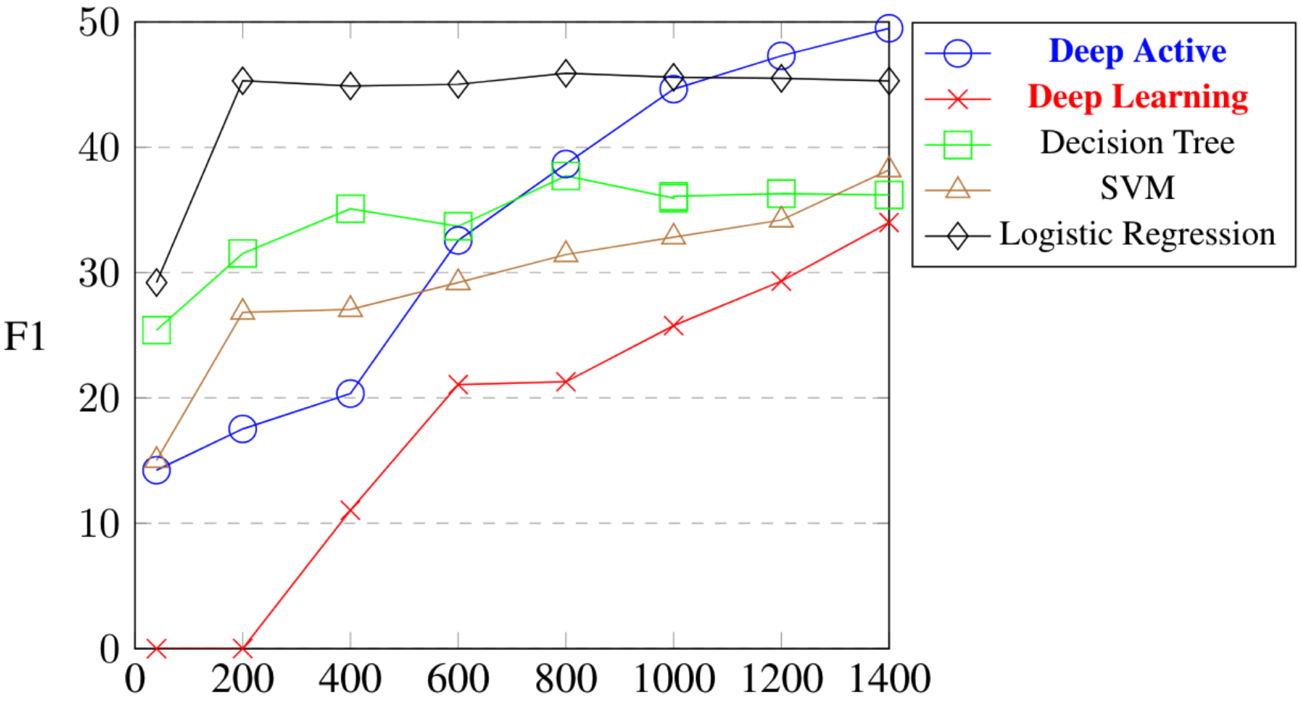
Questions?

Backup



Deep Transfer Active
Deep Active
Deep Learning
Decision Tree
SVM
Random Forest
Naive Bayes
Logistic Regression
Linear Regression
(a) DBLP-ACM
(b) DBLP-Scholar
(c) Cora
F1 scores
Labeled Training examples
Labeled Training examples
Labeled Training examples

More Genres, Software (Amazon-Google)



Effect of Sampling Strategies

	Precision	Recall	F1
Uncertain Examples	96.16	89.64	92.07+−9.73
+ Partition	96.14	97.12	96.61+−0.57
+ High Confidence	93.32	97.21	95.19+−2.21
High-Confidence + Partition			

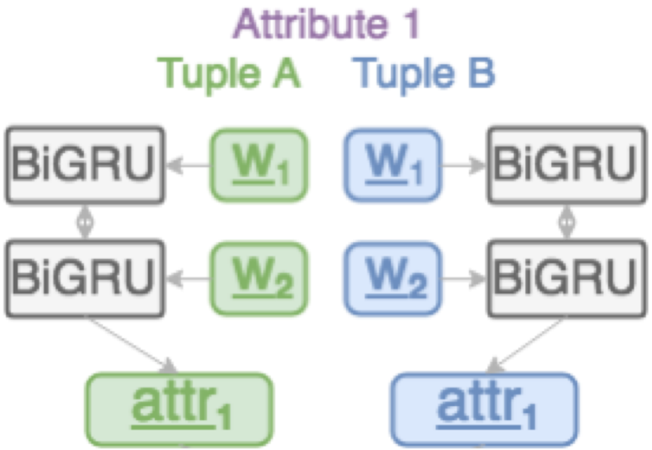
- 300 labeled examples of different sampling strategies (DBLP-ACM)

Attribute Representations (Mudgal et al. 2018)

Author

Alan

Turing

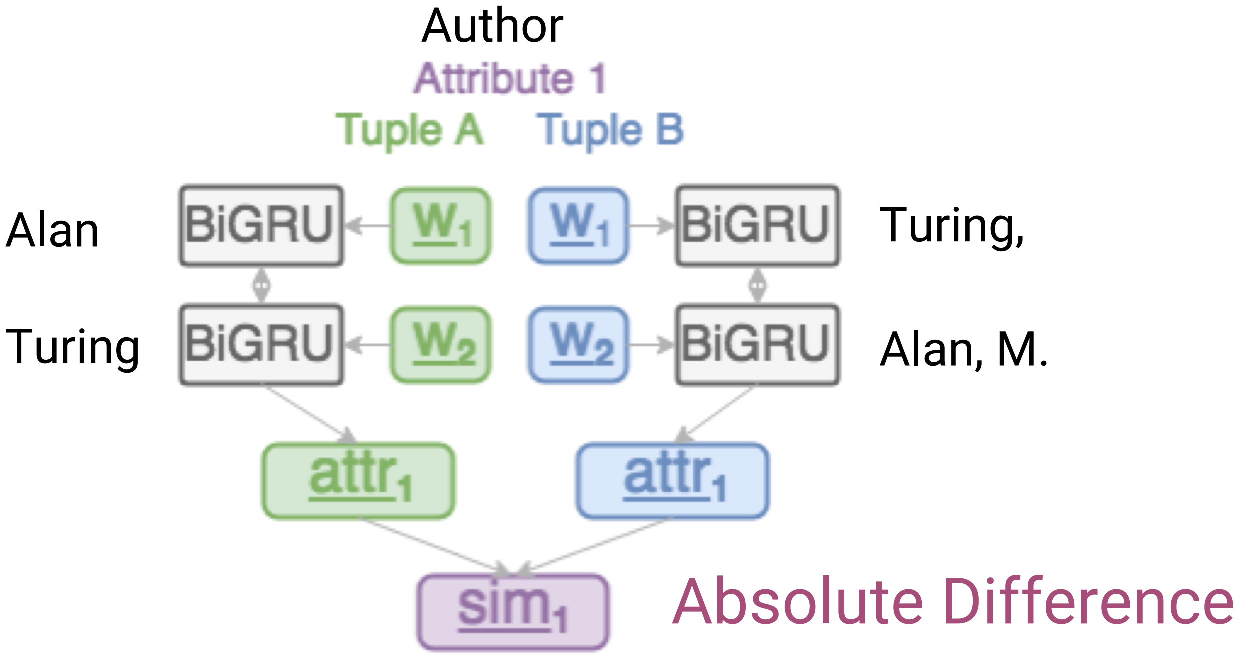


Turing,

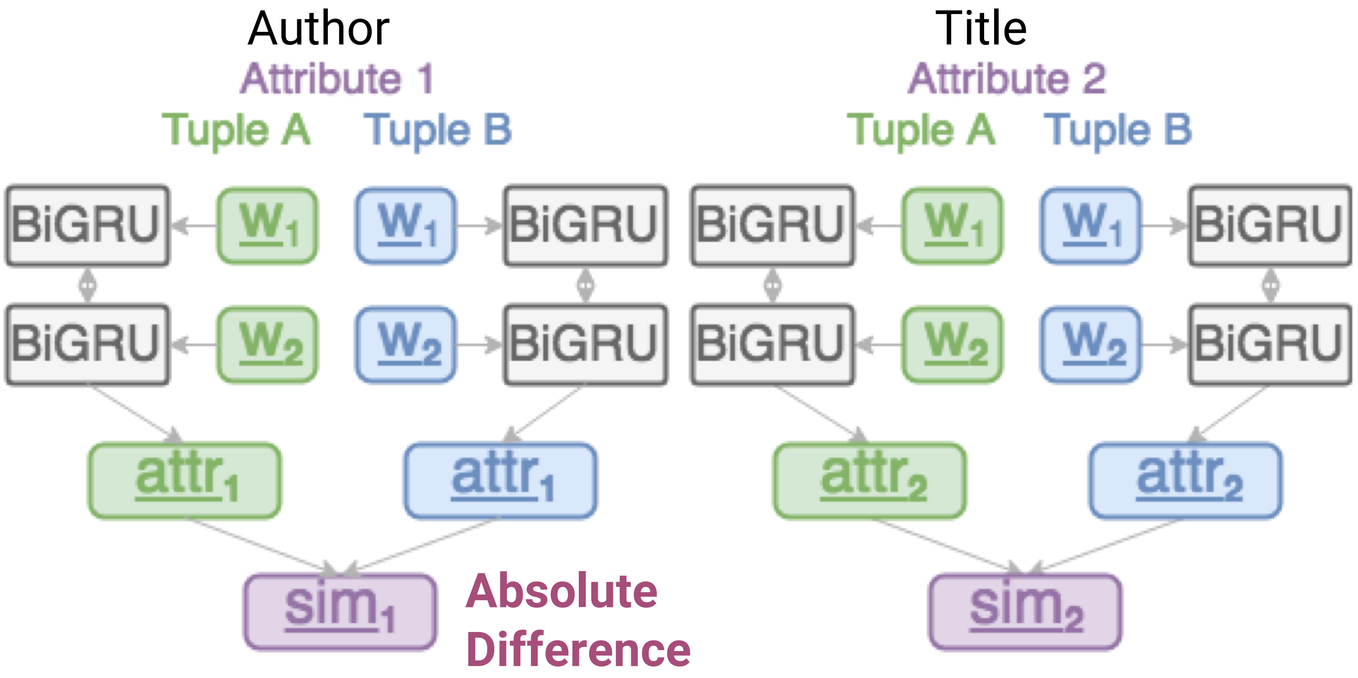
Alan M.

- Tuple A: Data Record from DBLP; Tuple B: Data record from Google Scholar
- Char-based

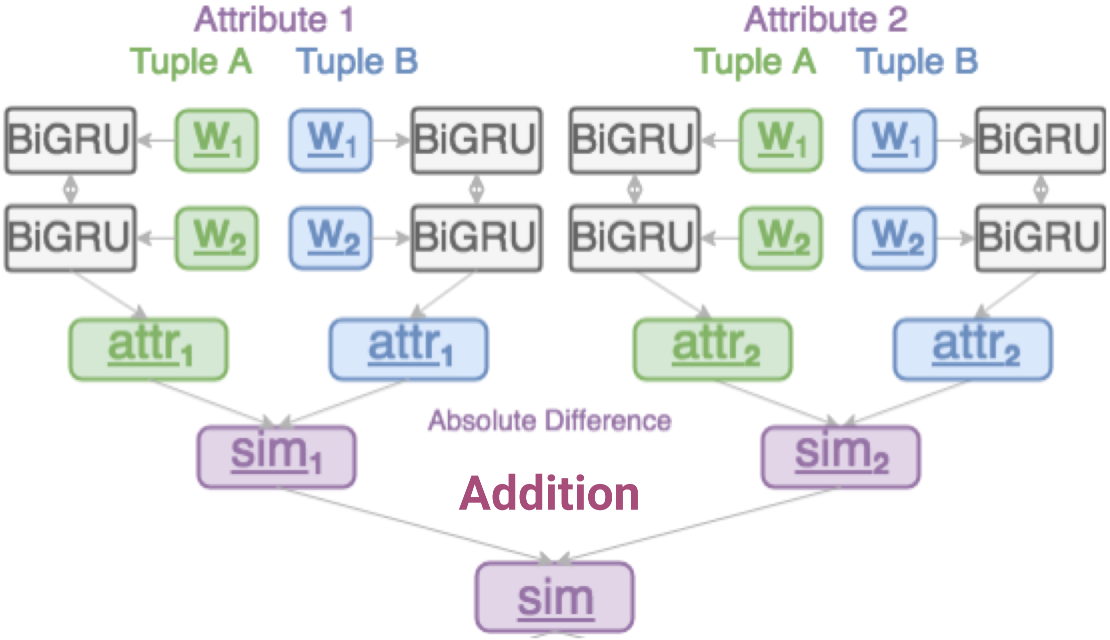
Attribute Similarity (Mudgal et al. 2018)



Attribute Similarity

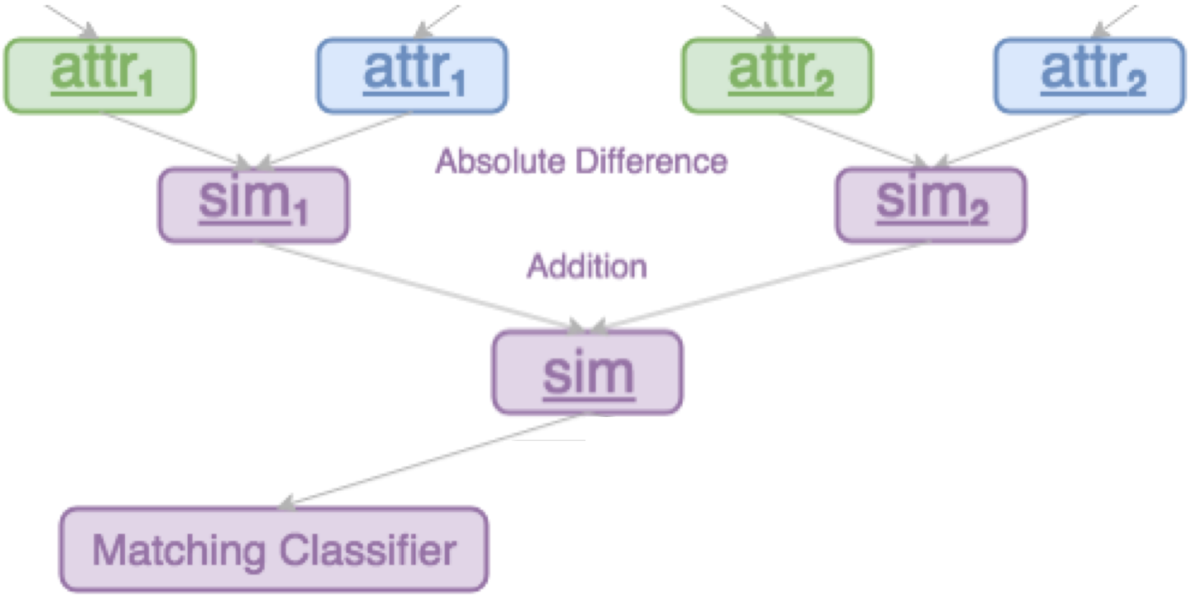


Transferable Record Similarity

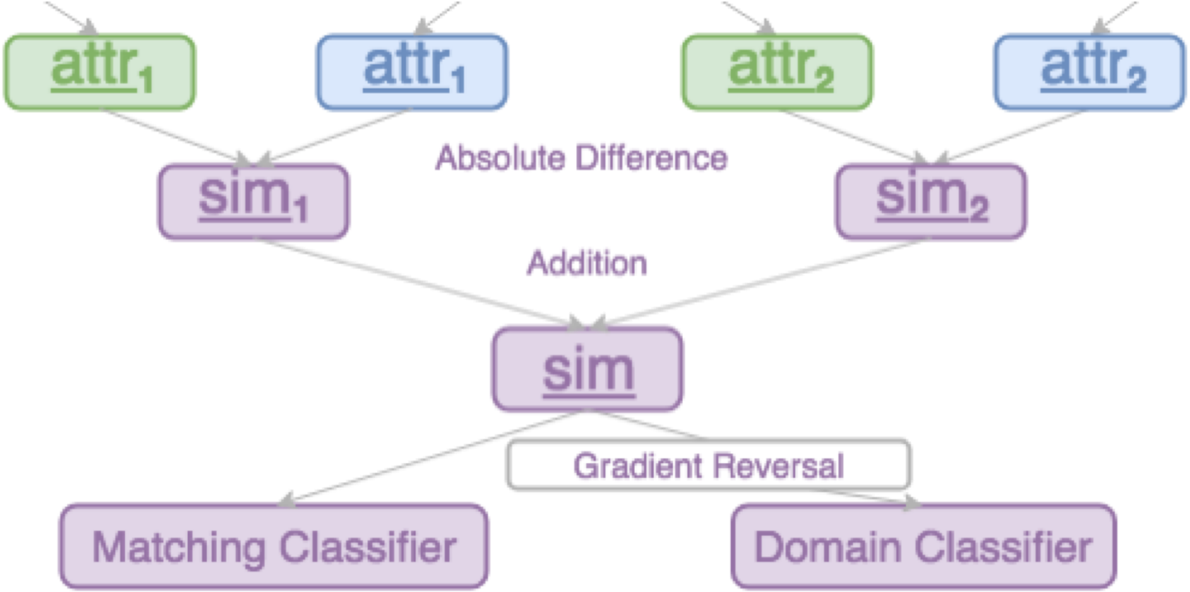


- DBLP-Scholar, DBLP-ACM: 4 attributes vs. Cora: 8 attributes (+ publisher etc)

Matching Classification



Domain Adaptation (Ganin and Lempitsky, 2015)



Active Learning for Deep ER

$H(x)$: Entropy given by the current model

High-Confidence: $\underset{D \subseteq D^U \, |D|=K}{\operatorname{argmin}} \sum_{x \in D} H(x)$ Uncertain: $\underset{D \subseteq D^U \, |D|=K}{\operatorname{argmax}} \sum_{x \in D} H(x)$

- Manually label uncertain examples and automatically label high-confidence examples by model predictions. Tune the model.

Active Learning for Deep ER

$H(x)$: Entropy given by the current model

High-Confidence: $\underset{D \subseteq D^U |D|=K}{\operatorname{argmin}} \sum_{x \in D} H(x)$ Uncertain: $\underset{D \subseteq D^U |D|=K}{\operatorname{argmax}} \sum_{x \in D} H(x)$

- Manually label uncertain examples and automatically label high-confidence examples by model predictions. Tune the model.
- Solves Problem 1: Need substantial positive/negative examples to tune the DL model.

Active Learning for Deep ER

$H(x)$: Entropy given by the current model

High-Confidence: $\underset{D \subseteq D^U |D|=K}{\operatorname{argmin}} \sum_{x \in D} H(x)$ Uncertain: $\underset{D \subseteq D^U |D|=K}{\operatorname{argmax}} \sum_{x \in D} H(x)$

- Manually label uncertain examples and automatically label high-confidence examples by model predictions.
- Solves Problem 1: Need substantial positive/negative examples to tune the DL model.
- Problem 2: Low recall.

Partition for ER Active Learning

Further subdivide each by model prediction. Positive and negative.

$$\overline{D}^U = \{x \in D^U | p(x) \geq 0.5\}, \underline{D}^U = \{x \in D^U | p(x) < 0.5\}$$

High-Confidence Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\operatorname{argmin}} \sum_{x \in D} H(x), \underset{D \subseteq \underline{D}^U |D|=k}{\operatorname{argmin}} \sum_{x \in D} H(x)$

Uncertain Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\operatorname{argmax}} \sum_{x \in D} H(x), \underset{D \subseteq \underline{D}^U |D|=k}{\operatorname{argmax}} \sum_{x \in D} H(x)$

Partition for ER Active Learning

Further subdivide each by model prediction. Positive and negative.

$$\overline{D}^U = \{x \in D^U | p(x) \geq 0.5\}, \underline{D}^U = \{x \in D^U | p(x) < 0.5\}$$

High-Confidence Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\operatorname{argmin}} \sum_{x \in D} H(x), \underset{D \subseteq \underline{D}^U |D|=k}{\operatorname{argmin}} \sum_{x \in D} H(x)$

Uncertain Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\operatorname{argmax}} \sum_{x \in D} H(x), \underset{D \subseteq \underline{D}^U |D|=k}{\operatorname{argmax}} \sum_{x \in D} H(x)$

- **likely** true positives/negatives and **likely** false positives/negatives

Partition for ER Active Learning

Further subdivide each by model prediction.

$$\overline{D}^U = \{x \in D^U | p(x) \geq 0.5\}, \underline{D}^U = \{x \in D^U | p(x) < 0.5\}$$

High-Confidence Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\text{argmin}} \sum_{x \in D} H(x), \quad \underset{D \subseteq \underline{D}^U |D|=k}{\text{argmin}} \sum_{x \in D} H(x)$

Uncertain Positives/Negatives: $\underset{D \subseteq \overline{D}^U |D|=k}{\text{argmax}} \sum_{x \in D} H(x), \quad \underset{D \subseteq \underline{D}^U |D|=k}{\text{argmax}} \sum_{x \in D} H(x)$

- **likely** true positives/negatives and **likely** false positives/negatives
- Empirically this equal partition helps (discussed later).

Effect of Sampling Strategies

	Precision	Recall	F1
Uncertain Examples	96.16	89.64	92.07+−9.73
+ Partition			
+ Partition + High-confidence			

- 300 labeled examples of different sampling strategies (DBLP-ACM)

Effect of Sampling Strategies

	Precision	Recall	F1
Uncertain Examples	96.16	89.64	92.07+−9.73
+ Partition	96.14	97.12	96.61+−0.57
+ Partition + High-confidence			

- 300 labeled examples of different sampling strategies (DBLP-ACM)

Effect of Sampling Strategies

	Precision	Recall	F1
Uncertain Examples	96.16	89.64	92.07+−9.73
+ Partition	96.14	97.12	96.61+−0.57
+ Partition + High-confidence	**97.63**	**97.84**	**97.73+−0.43**

- 300 labeled examples of different sampling strategies (DBLP-ACM)

Breakdown of Labeled Samples

	False Pos.	False Neg.	True Pos.	True Neg.
No Partition	33.9%	13.9%	19.1%	33.1%
Partition	26.5%	19.7%	23.5%	30.3%

- 300 labeled examples of different sampling strategies (DBLP-ACM)