ESBM: An Entity Summarization Benchmark

Qingxia Liu[1], Gong Cheng[1], Kalpa Gunaratna[2], and Yuzhong Qu[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University, China

[2] Samsung Research America, Mountain View CA, USA

Outline

- Introduction
- Creating ESBM
- Analyzing ESBM
- Evaluating with ESBM
- Conclusion

Entity Summarization

Description of Tim Berners-Lee:

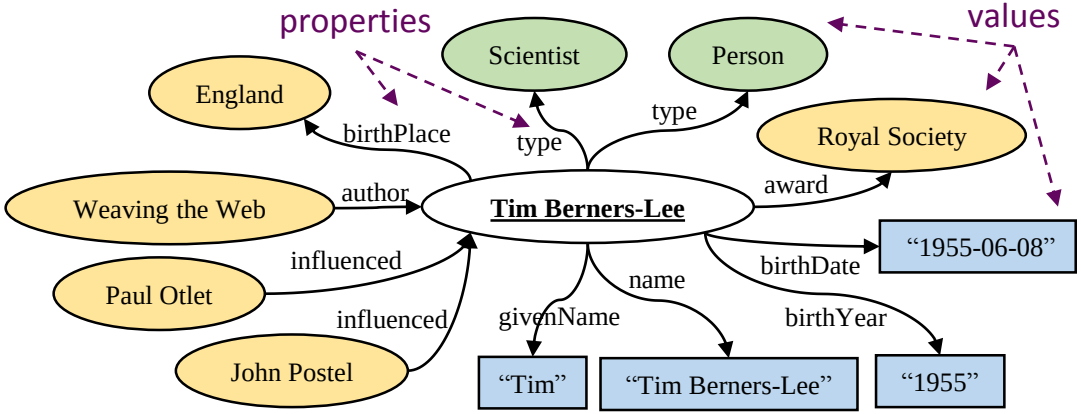
<Tim Berners Lee, alias, “TimBL”>
<Tim Berners Lee, name, “Tim Berners-Lee”>
<Tim Berners Lee, givenName, “Tim”>
<Tim Berners Lee, birthYear, “1955”>
<Tim Berners Lee, birthDate, “1955-06-08”>
<Tim Berners Lee, birthPlace, England>
<Tim Berners Lee, birthPlace, London>
<Tim Berners Lee, type, People Educated At Emanuel School>
<Tim Berners Lee, type, Scientist>
<Tim Berners Lee, type, Living People>
<Tim Berners Lee, type, Person>
<Tim Berners Lee, type, Agent>
<Tim Berners-Lee, award, Royal Society>
<Tim Berners-Lee, award, Royal Academy of Engineering >
<Tim Berners-Lee, award, Order of Merit>
<Tim Berners-Lee, award, Royal Order of the British Empire>
<Tim Berners-Lee, spouse, Rosemary Leith>
<Tim Berners-Lee, child, Ben Berners-Lee>
<Tim Berners-Lee, child, Alice Berners-Lee>
<Conway Berners-Lee, child, Tim Berners-Lee>
<Weaving the Web, author, Tim Berners-Lee>
<Tabulator, author, Tim Berners-Lee>
<Paul Otlet, influenced, Tim Berners-Lee>
<John Postel, influenced, Tim Berners-Lee>
<World Wide Web, developer, Tim Berners-Lee>
<World Wide Web Foundation, foundedBy, Tim Berners-Lee>
<World Wide Web Foundation, keyPerson, Tim Berners-Lee>

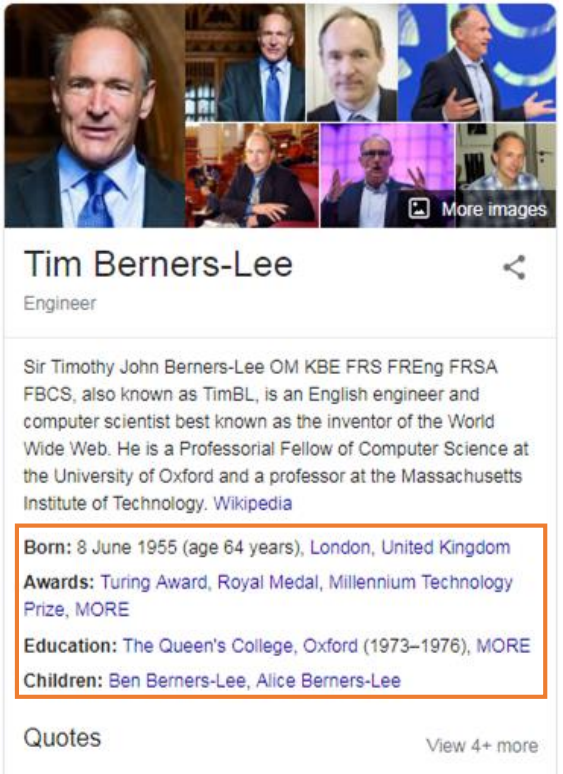
Summary:

<Tim Berners Lee, birthDate, “1955-06-08”>
<Tim Berners Lee, birthPlace, England>
<Tim Berners Lee, type, Scientist>
<Tim Berners-Lee, award, Royal Society>
<World Wide Web, developer, Tim Berners-Lee>

Entity Summarization

- RDF Data: T
  - triple t∈T: <subj, pred, obj>
- Entity Description: Desc(e)
  - Desc(e) ={t∈T: subj(t)=e or obj(t)=e}
  - triple t∈Desc(e): <e, property, value>
  - values: class, entity, literal
- Entity Summarization (ES): S(e, k)
  - S⊆Desc(e) , |S|≤k





Existing Benchmarks

- Limitations
  - Task specificness
  - Single dataset
  - Small size
  - Triple incomprehensiveness

Table 1: Existing benchmarks for evaluating entity summarization.

	Dataset	Number of entities	Availability
WhoKnows?Movies! [22]	Freebase	60	**Available**[1]
Langer et al. [13]	DBpedia	14	Unavailable
FRanCo [1]	DBpedia	265	Unavailable
Benchmark for evaluating RELIN [2]	DBpedia	149	Unavailable
Benchmark for evaluating DIVERSUM [20]	IMDb	20	Unavailable
Benchmark for evaluating FACES [7]	DBpedia	50	**Available**[2]
Benchmark for evaluating FACES-E [8]	DBpedia	80	**Available**[2]

[1] http://yovisto.com/labs/iswc2012
[2] http://wiki.knoesis.org/index.php/FACES

Our Work

- Motivation

  Research Challenges for Entity Summarization:
  - Lack of good benchmarks
  - Lack of evaluation efforts

- Contributions
  - Created an Entity Sumarization Benchmark (**ESBM** v1.2)
    - overcoming the limitations of existing benchmarks
    - meeting the desiderata for a successful benchmark
  - Evaluated entity summarizers with ESBM
    - made the most extensive evaluation effort to date
    - evaluated 9 existing general-purpose entity summarizers
    - evaluated 1 supervised learning-based entity summarizer for reference

Creating ESBM

Design Goals

- To satisfy seven desiderata for a successful benchmark[18]
  - accessibility, affordability, clarity, relevance, solvability, portability, scalability

- To overcome limitations of available benchmarks
  - General-purpose summaries
  - Including class-, entity-, literal-valued triples
  - Multiple datasets
  - Currently largest available benchmark

[18] Sim, S.E., Easterbrook, S.M., Holt, R.C.: Using benchmarking to advance research: A challenge to software engineering. In: ICSE 2003. pp. 74{83 (2003).

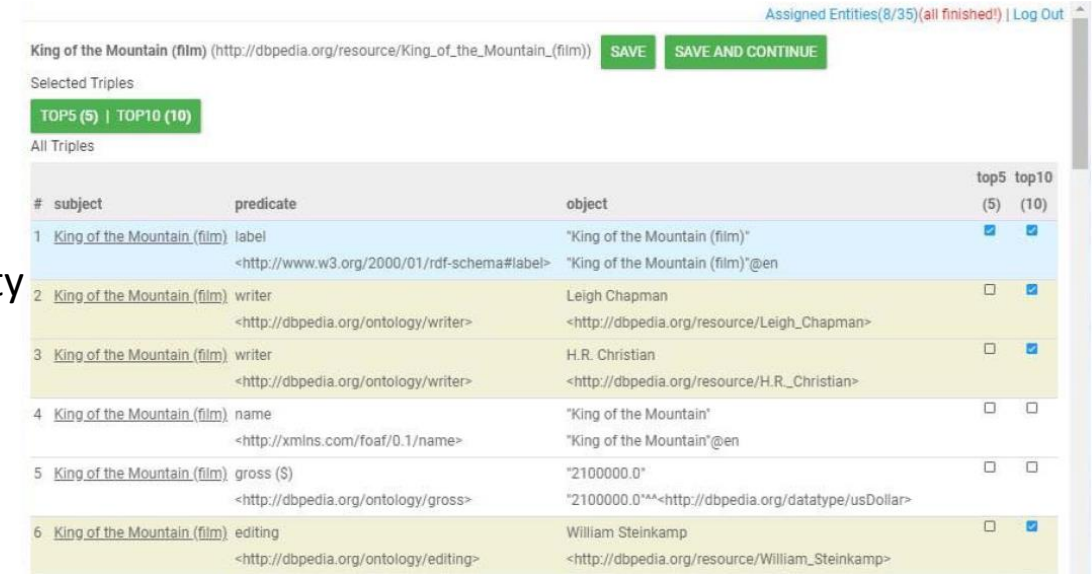
Entity Descriptions

- Datasets
  - DBpedia
    - imported dump files: *instance types, instance types transitive, YAGO types, mappingbased literals, mappingbased objects, labels, images, homepages, persondata, geo coordinates mappingbased*, and *article categories*
  - LinkedMDB
    - removed triples: *owl:sameAs*
- Entities

  sampled from seven large classes:
  - DBpedia: *Agent, Event, Location, Species, Work*
  - LinkedMDB: *Film, Person*
- Triples per entity
  - By class: 25.88-52.44 triples
  - Overall: 37.62 triples

Ground-Truth Summaries

- Task
  - 30 users
    - each assigned 35 entities
  - 175 entities
    - each assigned to 6 users
  - Each user created two summaries for each entity
    - for k=5 and k=10
- Total
  - 6 top-5 summaries and 6 top-10 summaries for each entity
  - 175*6*2=2100 ground-truth summaries



The ESBM Benchmark

- Usage
  - ESBM v1.2: specified training-validation-test splits for 5-fold cross validation
  - Early versions: EYRE 2018 workshop, EYRE 2019 workshop

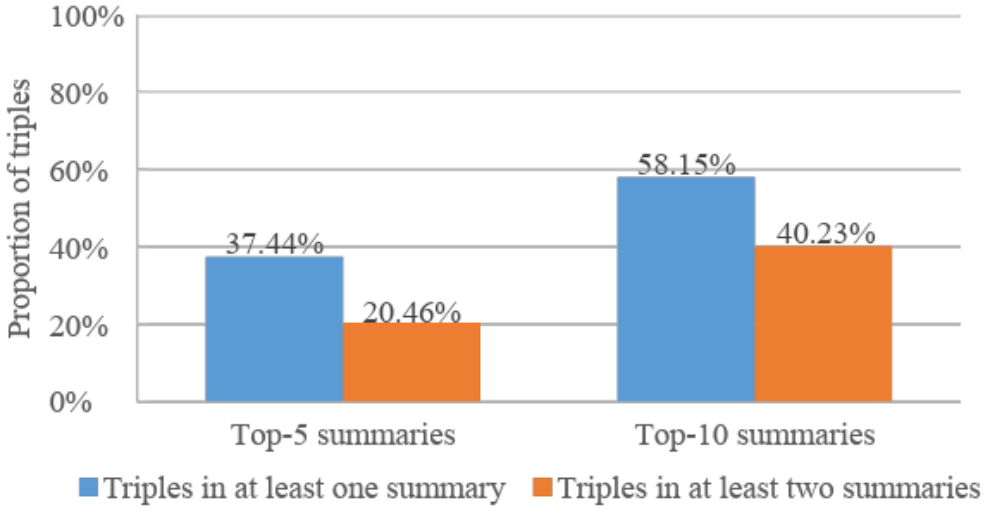
- Desiderata
  - Accessibility: permanent identifier on w3id.org
  - Affordability: open-source, example code for evaluation
  - Clarity: documented clearly and concisely
  - Relevance: entities sampled from real datasets
  - Solvability: not trivial and not too difficult
  - Portability: any general-purpose entity summarizer that can process RDF data
  - Scalability: reasonably large and diverse to evaluate mature entity summarizers

Analyzing ESBM

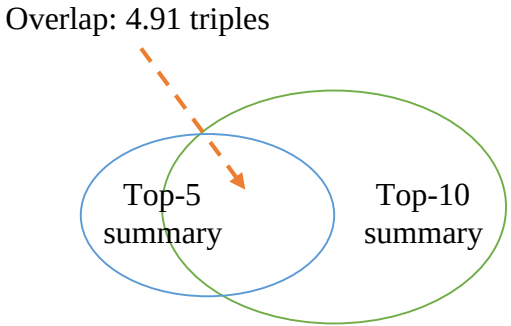
Basic Statistics

- 175 entities, 6584 triples, 2100 ground-truth summaries

Proportion of triples been selected into ground-truth summaries

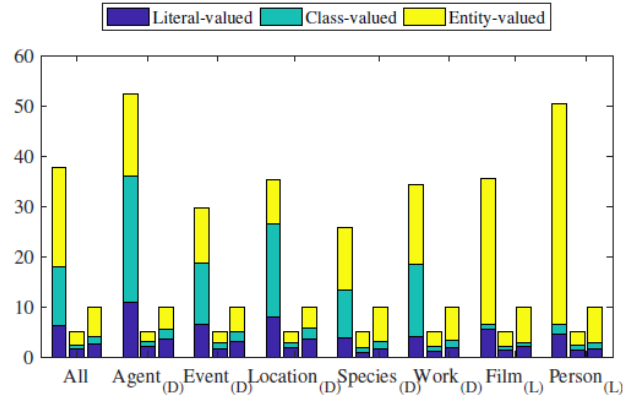


Overlap between top-5 and top-10 summaries

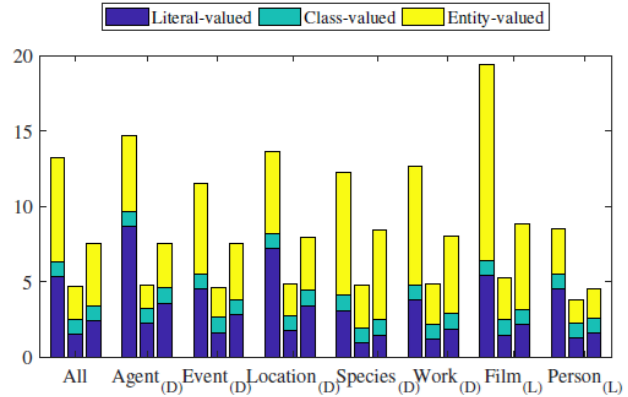


Triple Composition

- Literal-valued triples constitute a large proportion in ground-truth summaries.
  - 30% in top-5 ground-truth summaries and 25% in top-10 summaries
- Participants are not inclined to select multiple values of a property.
  - The average number of distinct properties in top-5 ground-truth summaries is 4.70 (very close to 5)



(a) Average number of triples describing an entity.



(b) Average number of distinct properties describing an entity.

Three bars in each group: Entity descriptions, Top-5 ground-truth summaries, Top-10 ground-truth summaries

Entity Heterogeneity

- Entity Description
  - Jaccard similarity between property sets from each pair of classes is very low.

DBpedia

	Work	Species	Location	Event
Agent	0.088	0.065	0.066	0.081
Event	0.089	0.090	0.102	
Location	0.090	0.077		
Species	0.087			

	Film
Person	0.085

LinkedMDB

Entity Heterogeneity

- Ground-truth Summaries
  - Popular properties:
    - properties that appear in >50% ground truth summaries for each class
  - Only 1~2/13.24 properties are popular in top-5 ground-truth summaries
    - The importance of properties is generally contextualized by concrete entities.

Table 2: Popular properties in ground-truth summaries.

In top-5 summaries							In top-10 summaries						
Agent	Event	Location	Species	Work	Film	Person	Agent	Event	Location	Species	Work	Film	Person
type	type	type	type	type	director	type	type	type	type	family	type	director	type
birthDate	date	country	family		type	actor	subject	subject	country	type	subject	actor	actor
							birthDate	date	subject	order	genre	type	label
								label		class		writer	page
										genus		producer	
										subject		date	
										kingdom		language	

Inter-Rater Agreement

- Average overlap between 6 ground-truth summaries
  - Moderate degree of agreement
  - Comparable with those reported for other benchmarks

	ESBM	[2]	[7]	[8]
Overlap between top-5 summaries	1.99 (39.8%)	2.91 (58.2%)	1.92 (38.4%)	2.12 (42.4%)
Overlap between top-10 summaries	5.42 (54.2%)	7.86 (78.6%)	4.64 (46.4%)	5.44 (54.4%)
Ground-truth summaries per entity	6	4.43	≥ 7	≥ 4

[2] Cheng, G., Tran, T., Qu, Y.: RELIN: relatedness and informativeness-based centrality for entity summarization. In: ISWC 2011, Part I. pp. 114-129 (2011).
[7] Gunaratna, K., Thirunarayan, K., Sheth, A.P.: FACES: diversity-aware entity summarization using incremental hierarchical conceptual clustering. In: AAAI 2015. pp. 116-122 (2015).
[8] Gunaratna, K., Thirunarayan, K., Sheth, A.P., Cheng, G.: Gleaning types for literals in RDF triples with application to entity summarization. In: ESWC 2016. pp. 85-100 (2016).

Evaluating with ESBM

Participating Entity Summarizers

- Existing Entity Summarizers
  - RELIN, DIVERSUM, LinkSUM, FACES, FACES-E, CD
  - MPSUM, BAFREC, KAFCA
- ORACLE Entity Summarizer
  - k triples that are selected by the most participants into ground-truth summaries
- Supervised Learning-Based Entity Summarizer
  - 6 models:
    - SMOreg, LinearRegression, MultilayerPerceptron, AdditiveRegression, REPTree, RandomForest
  - 7 features:
    - gf_T(global frequency of property), lf(local frequency of property), vf_T(frequency of value), si(self-information of triple)
    - isC(value is class), isE(value is entity), isL(value is literal)

Settings

- Evaluation Criteria

$$\mathrm{P} = \frac{|S_m \cap S_h|}{|S_m|}, \quad \mathrm{R} = \frac{|S_m \cap S_h|}{|S_h|}, \quad \mathrm{F1} = \frac{2 \cdot \mathrm{P} \cdot \mathrm{R}}{\mathrm{P} + \mathrm{R}}.$$

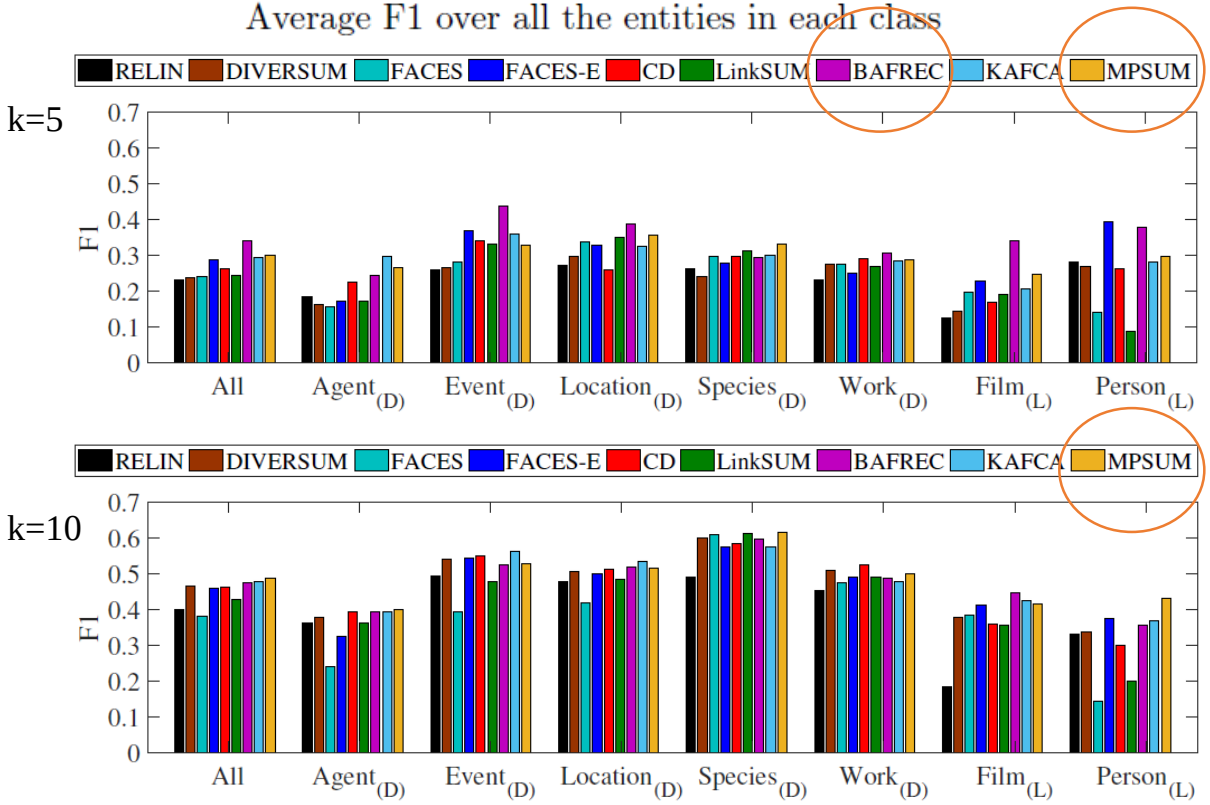
  - S_m: machine-generated entity summary
  - S_h : human-made ground-truth summary
  - $P \neq R$ if $|S_m| < |S_h| = k$

Overall Results

Average F1 over all the entities in a dataset.

	DBpedia		LinkedMDB	
	$k = 5$	$k = 10$	$k = 5$	$k = 10$
RELIN	0.242 -○○▼▼▼▼▼▼	0.455 -▼○○▼○▼▼▼	0.203 -○○▼○▲▼○▼	0.258 -▼○▼▼○▼▼▼
DIVERSUM	0.249 ○-○○▼▼▼▼▼	0.507 ▲-▲○○○○○○	0.207 ○-○▼○▲▼○▼	0.358 ▲-▲○○▲▼○▼
FACES	0.270 ○○-○○○▼▼▼	0.428 ○▼-▼▼▼▼▼▼	0.169 ○○-▼▼○▼▼▼	0.263 ○▼-▼▼○▼▼▼
FACES-E	0.280 ▲○○-○○▼▼▼	0.488 ○○▲-○○○○○	0.313 ▲▲▲-▲▲▼▲○	0.393 ▲○▲-▲▲○○○
CD	0.283 ▲▲○○-○▼○○	0.513 ▲○▲○-○○○○	0.217 ○○▲▼-▲▼○▼	0.331 ▲○▲▼-▲▼▼▼
LinkSUM	0.287 ▲▲○○○-▼○○	0.486 ○○▲○○-○○○	0.140 ▼▼○▼▼-▼▼▼	0.279 ○▼○▼▼-▼▼▼
BAFREC	0.335 ▲▲▲▲▲▲-○○	0.503 ▲○▲○○○-○○	0.360 ▲▲▲▲▲▲-▲▲	0.402 ▲▲▲○▲▲-○○
KAFCA	0.314 ▲▲▲▲○○○-○	0.509 ▲○▲○○○○-○	0.244 ○○▲▼○▲▼-○	0.397 ▲○▲○▲▲○-○
MPSUM	0.314 ▲▲▲▲○○○○-	0.512 ▲○▲○○○○○-	0.272 ▲▲▲○▲▲▼○-	0.423 ▲▲▲○▲▲○○-
ORACLE	0.595	0.713	0.619	0.678

Results on Different Entity Types



Results of Supervised Learning

- F1 results
  - RandomForest, REPTree achieve the highest F1.
  - Four methods outperform all the existing entity summarizers.
  - Two methods only fail to outperform existing entity summarizers in one setting.

	DBpedia		LinkedMDB	
	$k = 5$	$k = 10$	$k = 5$	$k = 10$
RELIN	0.242 -○○▼▼▼▼▼▼	0.455 -▼○○▼○▼▼▼	0.203 -○○▼○▲▼○▼	0.258 -▼○▼▼○▼▼▼
DIVERSUM	0.249 ○-○○▼▼▼▼▼	0.507 ▲-▲○○○○○○	0.207 ○-○▼○▲▼○▼	0.358 ▲-▲○○▲▼○▼
FACES	0.270 ○○-○○○▼▼▼	0.428 ○▼-▼▼▼▼▼▼	0.169 ○○-▼▼○▼▼▼	0.263 ○▼-▼▼○▼▼▼
FACES-E	0.280 ▲○○-○○▼▼▼	0.488 ○○▲-○○○○○	0.313 ▲▲▲-▲▲▼▲○	0.393 ▲○▲-▲▲○○○
CD	0.283 ▲▲○○-○▼○○	0.513 ▲○▲○-○○○○	0.217 ○○▲▼-▲▼○▼	0.331 ▲○▲▼-▲▼▼▼
LinkSUM	0.287 ▲▲○○○-▼○○	0.486 ○○▲○○-○○○	0.140 ▼▼○▼▼-▼▼▼	0.279 ○▼○▼▼-▼▼▼
BAFREC	0.335 ▲▲▲▲▲▲-○○	0.503 ▲○▲○○○-○○	0.360 ▲▲▲▲▲▲-▲▲	0.402 ▲▲▲○▲▲-○○
KAFCA	0.314 ▲▲▲▲○○○-○	0.509 ▲○▲○○○○-○	0.244 ○○▲▼○▲▼-○	0.397 ▲○▲○▲▲○-○
MPSUM	0.314 ▲▲▲▲○○○○-	0.512 ▲○▲○○○○○-	0.272 ▲▲▲○▲▲▼○-	0.423 ▲▲▲○▲▲○○-
ORACLE	0.595	0.713	0.619	0.678
SMOreg	0.279	0.543	0.403	0.472
LinearRegression	0.319	0.556	0.401	0.471
MultilayerPerceptron	0.340	0.560	0.390	0.477
AdditiveRegression	0.345	0.558	0.415	0.510
REPTree	0.392	0.570	0.455	0.538
RandomForest	0.399	0.576	0.449	0.506

Demonstrated the powerfulness of supervised learning for entity summarization.

Results of Supervised Learning

- Features

  for each t=<e, p,v> in Desc(e):
  - gf_T: # triples in the dataset where p appears
  - lf: # triples in Desc(e) where p appears
  - vf_T: # triples in dataset where v appears
  - si: self-information of triple t
  - isC: whether v is a class
  - isE: whether v is an entity
  - isL: whether v is a literal
- Results
  - significantly effective: gf_T, lf
    - for LinkedMDB: vf_T, si
  - not significant: isC, isE, isL

Table 5: F1 of RandomForest after removing each individual feature, its difference from using all features ($\Delta\%$), and the significance level for the difference (p).

DBpedia								LinkedMDB							
$k=5$				$k=10$				$k=5$				$k=10$			
	F1	$\Delta\%$	p		F1	$\Delta\%$	p		F1	$\Delta\%$	p		F1	$\Delta\%$	p
All	0.399	—	—	All	0.576	—	—	All	0.449	—	—	All	0.506	—	—
-gf_T	0.346	−5.360	0.000	-lf	0.546	−0.030	0.000	-gf_T	0.383	−0.066	0.000	-lf	0.473	−0.033	0.008
-lf	0.366	−3.307	0.000	-gf_T	0.551	−0.025	0.000	-lf	0.413	−0.036	0.025	-vf_T	0.477	−0.029	0.010
-isC	0.392	−0.720	0.261	-vf_T	0.569	−0.007	0.198	-vf_T	0.414	−0.035	0.022	-gf_T	0.479	−0.027	0.007
-isE	0.397	−0.267	0.720	-isE	0.570	−0.006	0.262	-si	0.442	−0.007	0.574	-si	0.486	−0.020	0.009
-si	0.400	+0.027	0.973	-isC	0.571	−0.005	0.303	-isE	0.455	+0.005	0.651	-isL	0.491	−0.015	0.079
-isL	0.401	+0.160	0.816	-si	0.572	−0.004	0.402	-isL	0.456	+0.007	0.504	-isE	0.492	−0.014	0.148
-vf_T	0.407	+0.720	0.346	-isL	0.578	+0.002	0.683	-isC	0.463	+0.013	0.281	-isC	0.514	+0.008	0.396

Summary of Evaluation Results

- Existing entity summarizers
  - Leading systems: BAFREC, MPSUM
- Supervised Learning method
  - Outperforms existing entity summarizers
- Comparing with ORACLE
  - Still a large gap for improvement

Entity summarization on ESBM is a non-trivial task.

Conclusion

Limitations

- Evaluation Criteria
  - semantic overlap between triples
- Representativeness of Ground Truth
  - general-purpose VS. task-specific
- Form of Ground Truth
  - set-based VS. scoring-based

Take-home Message

- Contributions
  - Created an Entity Summarization Benchmark: ESBM
    - overcoming the limitations of existing benchmarks
  - Evaluated entity summarizers with ESBM
    - the most extensive evaluation effort to date

- ESBM
  - The currently largest available benchmark for entity summarization
  - Entity summarization on ESBM is a non-trivial task
  - Permanent link: https://w3id.org/esbm/
  - GitHub repository: nju-websoft/ESBM

- An Upcoming Paper
  - Junyou Li, Gong Cheng, Qingxia Liu, Wen Zhang, Evgeny Kharlamov, Kalpa Gunaratna, Huajun Chen. *Neural Entity Summarization with Joint Encoding and Weak Supervision.* IJCAI-PRICAI 2020
    - Deep learning based
    - Significantly outperformed all the existing systems on ESBM

Thank you !

Questions ?