Improving NLU Training over Linked Data

Tobias Schmitt, **Cedric Kulbach,** York Sure-Vetter

Developments in Dialogue Systems

FZI

NLU as a Service[1]

Microsoft Bot Framework

IBM Watson

chatfuel

wit.ai

RASA

Universal Chat Platforms[1]

Skype

slack

Machine Learning[1]

Underlying Algorithms

Hand-crafted, Machine Learning, …

Labelled utterances to learn Alg.

Research Approaches

Objective: Train high performance NLU

Approach: Improve quality of Training Data

[1] Braun, D., Hernandez-Mendez, A., Prof. Dr. Matthes, F., Dr. Langen, M.: Evaluating natural language understanding services for conversational question answering systems (2017)

Developments in Dialogue Systems

NLU as a Service[1]

Microsoft Bot Framework

Machine Learning[1]

Underlying Algorithms

Hand-crafted, Machine Learning, …

Labelled utterances to learn Alg.

Research Approaches

Objective: Train high performance NLU

Approach: Improve quality of Training Data

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | @jjvincent | Mar 24, 2016, 6:43am EDT

Facebook and YouTube should have learned from Microsoft's racist chatbot

Facebook and YouTube have recently come under fire for offensive search suggestions

CNBC

Published: 11:15am, 19 Mar, 2018

AI Needs Content Strategy More Than Ever

Matthew Grocki

Aug 13, 2019

Research Questions

RQ 1	**Which type of entity values work best for training an entity recognition algorithm?**
RQ 2	**Which type of entity values work best for training different intent classifiers?**
RQ 3	**How can linked data improve NLU performances?**

Agenda

FZI

Related Work — 1

Conceptual training data design approach — 2

Experimental setup — 3

4 — Results

5 — Summary & Outlook

Related Work

Semantic Question Answering over Linked Data	Dialogue Systems
▪ DBpedia Bot (static, rule based) [Athreya et. Al.,2018] ▪ Frankenstein (linkage of static approaches, with static knowledge) [Singh et. Al,2018]	▪ Intent classification and Named Entity Recognition ▪ Crowdsourced generation of training data for chatbots [Bapat et. Al.,2018]

Overall Goal:

Bringing technology of dialogues systems towards semantic question answering systems or Integrating Semantic Question Answering Systems into an Dialogue System environment.

Intent and Entities of incoming messages

FZI

Component-based Dialogue System

“Where is the lecture Web Science taking place?”

Natural Language Understanding
Intent: location_of_lecture
Entity_Type: lecture
Entity_Value: “Web Science”

Dialogue Manager
Slots:
- lecture: Web Science
- semester: empty
- …

List of Actions:
- FindLocation
- FindLecturer

Natural Language Generation
“The lecture <lecture> takes place in the <location>”

“The lecture Web Science takes place in the audimax.”

NLU Training Process

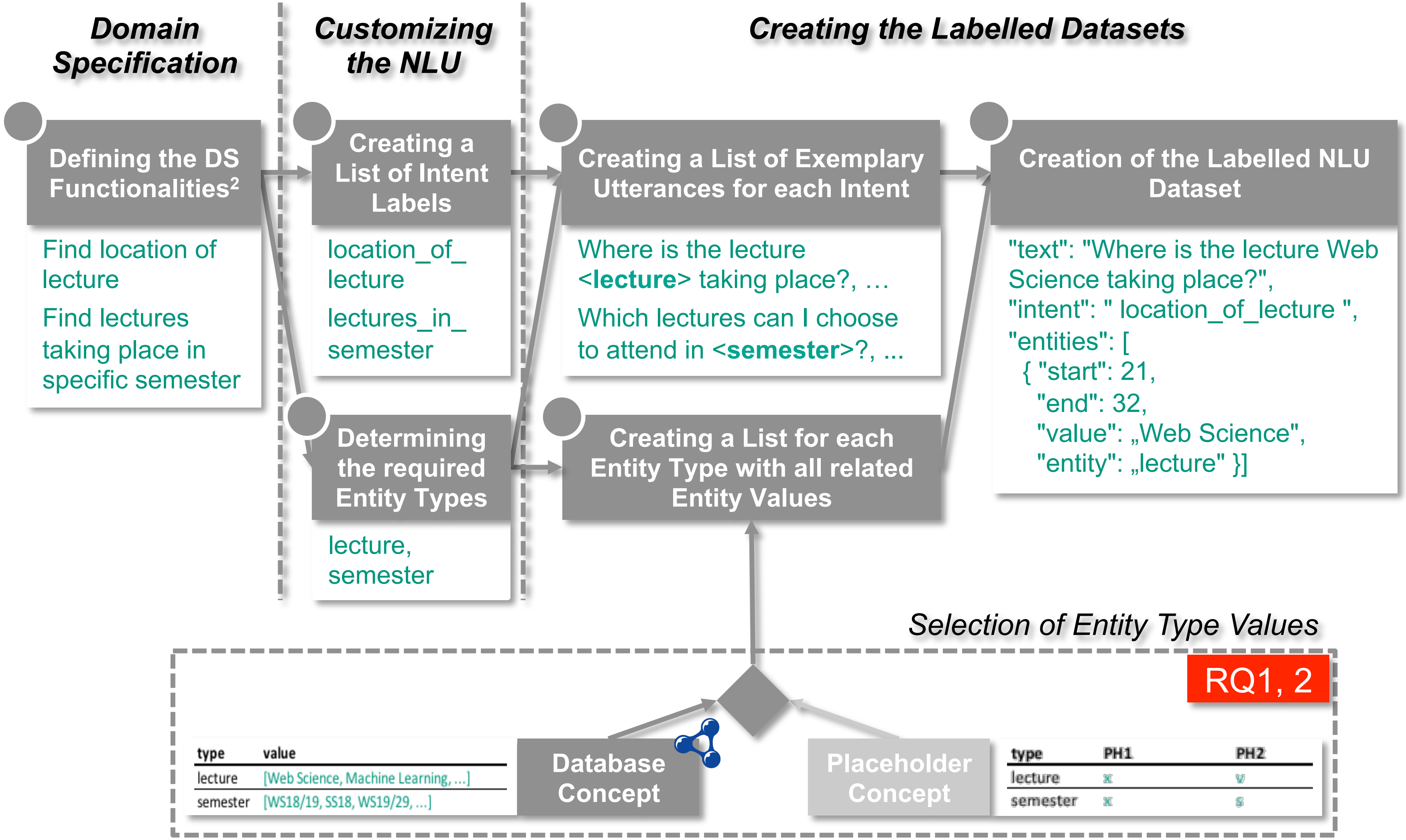
Intent → NLU
Utterance → NLU
Entity (Type + Value) → NLU

Utterance → NLU → Intent recognition
Utterance → NLU → Entity extraction

Format of the Dataset (JSON)

```
"text": “Where is the lecture Web Science taking
place?",
"intent": “location_of_lecture",
“entities": [
{
  "start": 21,
  "end": 32,
  "value": “Web Science",
  "entity": “lecture"
}]
```

Concepts NLU training dataset creation



[2] Grötz, 2018

Concepts NLU training dataset creation

Concept	
Entity value generation	
List of entity values	
Utterance	<beginning of an utterance> <entity type x> <…> <entity type y> <end of utterance>
Example	Where is the lecture <lecture> taking place?

Concepts NLU training dataset creation

<table>
<tr><th>Concept</th><th>Database Concept</th></tr>
<tr><td>Entity value generation</td><td>Database queries</td></tr>
<tr><td>List of entity values</td><td><table><tr><th>Entity type</th><th>Entity values</th></tr><tr><td>Type 1</td><td>[List of values]</td></tr><tr><td>Type 2</td><td>[List of values]</td></tr><tr><td>…</td><td>…</td></tr></table></td></tr>
<tr><td>Utterance</td><td><beginning of an utterance> <entity type x> <…> <entity type y> <end of utterance></td></tr>
<tr><td>Example</td><td>Entity Type: lecture
Entity Values: [Web Science, …]
Where is the lecture <lecture> taking place?</td></tr>
</table>

Concepts NLU training dataset creation

<table>
<tr><th rowspan="2">Concept</th><th rowspan="2">Database Concept</th><th colspan="2">Placeholder Value Concepts</th></tr>
<tr><th>Identical Pl. Values
(PH Type 1)</th><th>Different Pl. Value
(PH Type 2)</th></tr>
<tr><td>Entity value generation</td><td>Database queries</td><td>One random character sequence for all entity types</td><td>Random character sequence for each entity type</td></tr>
<tr><td>List of entity values</td>
<td><table><tr><th>Entity type</th><th>Entity values</th></tr><tr><td>Type 1</td><td>[List of values]</td></tr><tr><td>Type 2</td><td>[List of values]</td></tr><tr><td>…</td><td>…</td></tr></table></td>
<td><table><tr><th>Entity type</th><th>Entity values</th></tr><tr><td>Type 1</td><td rowspan="3">Random Value</td></tr><tr><td>Type 2</td></tr><tr><td>…</td></tr></table></td>
<td><table><tr><th>Entity type</th><th>Entity values</th></tr><tr><td>Type 1</td><td>Random Value 1</td></tr><tr><td>Type 2</td><td>Random Value 2</td></tr><tr><td>…</td><td>…</td></tr></table></td></tr>
<tr><td>Utterance</td><td colspan="3"><beginning of an utterance> <entity type x> <…> <entity type y> <end of utterance></td></tr>
<tr><td rowspan="2">Example</td><td>Entity Type: lecture
Entity Values: [Web Science, …]</td><td>Entity Type: lecture
Entity Value: x</td><td>Entity Type: lecture, …
Entity Value: v</td></tr>
<tr><td colspan="3">Where is the lecture <lecture> taking place?</td></tr>
</table>

Experimental setup

NLU dataset creation

- Domain:
  - Study program for KIT students
- Intents:
  - lecturer_of_lecture, semester_of_lecture, location_of_room, …
- Entities:
  - lecture, lecturer, semester, building, …
- 299 different utterances
  - 80%- 20% train, test split

- 5 experiments:
  - EX 1: Train domain entities
  - EX 2: All domain entities
  - EX 3: PH type 1
  - EX 4: PH type 2
  - EX 5: Train domain entities + DBPedia Entity values
- Test dataset:
  - Once with Database Entities
  - Once with DBpedia Entities

- RASA[3] as NLU component
- Intent Classifier
  - Embedding Intent Classifier
- NER:
  - Conditional Random Fields
- Standard configurations

[3] https://rasa.com/docs/rasa/nlu/about/

Experiments

	Entities for training	Entities for testing	
EX 1	80% Database	20% Database (Domain Test)	DBpedia Entities (Dbpedia Test)
EX 2	100% Database		
EX 3	Placeholder Type 1		
EX 4	Placeholder Type 2		
EX 5	80% Database + DBpedia		

Results on Domain Test

	Named Entity Recognition			Embedding Classifier		
	Precision	Recall	F1-Score	Precision	Recall	F1 Score
			Domain Test			
EX 1	0.995	0.895	0.940	0.863	0.846	0.845
EX 2	1.000	0.967	0.979	0.850	0.815	0.810

- Increased F1 score in NER task when more entities are used for training
- Decreased F1 score in intent classification task when more unique entities are used for training

Results on Domain Test

	Named Entity Recognition			**Embedding Classifier**		
	Precision	Recall	F1-Score	Precision	Recall	F1 Score
			Domain Test			
EX 1	0.995	0.895	0.940	0.863	0.846	0.845
EX 2	1.000	0.967	0.979	0.850	0.815	0.810
EX 3	0.893	0.286	0.428	0.866	0.831	0.829
EX 4	0.928	0.220	0.351	0.831	0.815	0.803

FZI

- Placeholder concepts do not improve NER or intent classification performance on domain tests

Results on Domain Test

	Named Entity Recognition			Embedding Classifier		
	Precision	Recall	F1-Score	Precision	Recall	F1 Score
			Domain Test			
EX 1	0.995	0.895	0.940	0.863	0.846	0.845
EX 2	1.000	0.967	0.979	0.850	0.815	0.810
EX 3	0.893	0.286	0.428	0.866	0.831	0.829
EX 4	0.928	0.220	0.351	0.831	0.815	0.803
EX 5	0.991	0.856	0.918	0.866	0.831	0.827

- Training with domain and DBpedia entities leads to slightly weaker performance results.

Results on DBpedia Test

	Named Entity Recognition			Embedding Classifier		
	Precision	Recall	F1-Score	Precision	Recall	F1 Score
			Domain Test			
EX 1	0.995	0.895	0.940	0.863	0.846	0.845
EX 2	1.000	0.967	0.979	0.850	0.815	0.810
EX 3	0.893	0.286	0.428	0.866	0.831	0.829
EX 4	0.928	0.220	0.351	0.831	0.815	0.803
EX 5	0.991	0.856	0.918	0.866	0.831	0.827
			DBpedia Test			
EX 1	0.878	0.770	0.812	0.845	0.815	0.810
EX 2	1.000	0.806	0.851	0.852	0.815	0.817
EX 3	0.835	0.312	0.451	0.891	0.877	0.872
EX 4	0.848	0.220	0.351	0.832	0.815	0.802

- Placeholder concepts lead to poor results in NER task but increases the intent classification performance.

Results on DBpedia Test

	Named Entity Recognition			Embedding Classifier		
	Precision	Recall	F1-Score	Precision	Recall	F1 Score
			Domain Test			
EX 1	0.995	0.895	0.940	0.863	0.846	0.845
EX 2	1.000	0.967	0.979	0.850	0.815	0.810
EX 3	0.893	0.286	0.428	0.866	0.831	0.829
EX 4	0.928	0.220	0.351	0.831	0.815	0.803
EX 5	0.991	0.856	0.918	0.866	0.831	0.827
			DBpedia Test			
EX 1	0.878	0.770	0.812	0.845	0.815	0.810
EX 2	1.000	0.806	0.851	0.852	0.815	0.817
EX 3	0.835	0.312	0.451	0.891	0.877	0.872
EX 4	0.848	0.220	0.351	0.832	0.815	0.802
EX 5	0.992	0.950	0.968	0.848	0.831	0.819

- Using DBpedia entity values increases the NER task performance when evaluating on a different domain.

Summary

FZI

RQ 1 **Which type of entity values work best for training an entity recognition algorithm?**	Depending on the field of application of the DS and the database size: **Closed Domain:** Leave Placeholders and Linked Data out of Training **Open Domain:** Integrate Entity Values from other Domains
RQ 2 **Which type of entity values work best for training different intent classifiers?**	Depending on the field of application of the DS and the database: **Closed Domain:** Few utterances are sufficient **Open Domain:** Integrate generated Entity Values (Type 1)
RQ 3 **How can linked data improve NLU performances?**	Training with Entity values from different Domains can improve NER of a NLU.

Outlook

Scalability

Evaluated only on a small, domain-specific dataset.
The question of the generalizability of the approach to larger datasets arises.

Robustness

We defined the robustness of a NLU through the performance metrics on not yet seen entity values.

Integration

Integration of placeholder holder concepts or domain knowledge into an automated dataset creation process.

References

- [Braun et. Al.,2017] - Braun, D., Hernandez-Mendez, A., Prof. Dr. Matthes, F., Dr. Langen, M.: Evaluating natural language understanding services for conversational question answering systems (2017)
- [Athreya et. Al.,2018] - Athreya, R.G., Ngomo, A.N., Usbeck, R.: Enhancing community interactions with data-driven chatbots-the dbpedia chatbot (2018)
- [Singh et. Al,2018] - Singh, K., Lange, C., Vidal, M.E., Lehmann, J., Auer, S., Radhakrishna, A.S., Both, A., Shekarpour, S., Lytra, I., Usbeck, R., Vyas, A., Khikmatullaev, A., Punjani, D.: Why Reinvent the Wheel: Let's Build Question Answering Systems Together. Lyon, France (2018)
- [Grötz, 2018] - Grötz, R.: Sprich mit mir! iX - Magazin für Professionelle Informationstechnik 6, 50 (2018)
- [Bapat et. Al.,2018] - Bapat, R., Kucherbaev, P., Bozzon, A.: Effective crowdsourced generation of training data for chatbots natural language understanding (2018).

Thank you