# Reinforcement Learning Driven Translation Model for Search-oriented Conversational Systems

Wednesday 31st October, 2018

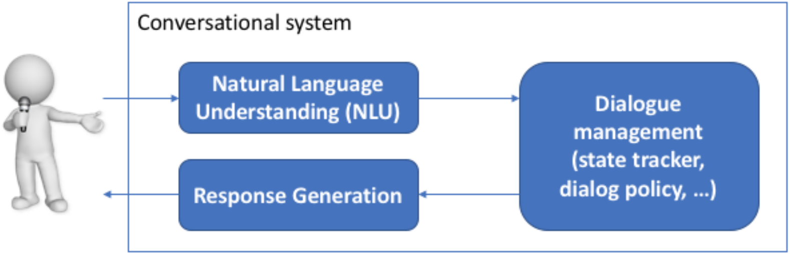
Wafa Aissa, Laure Soulier, and Ludovic Denoyer

We thank EMNLP for the student scholarship award.

Machine Learning & Deep Learning for Information Access

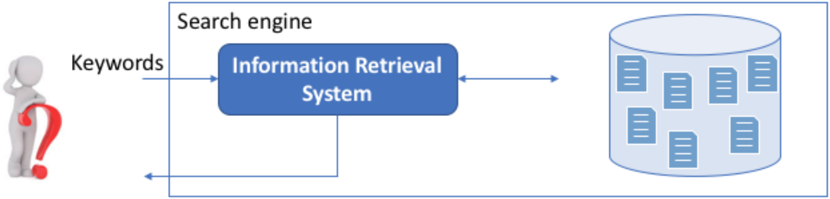
# Context and motivations

# Towards search-oriented conversational systems



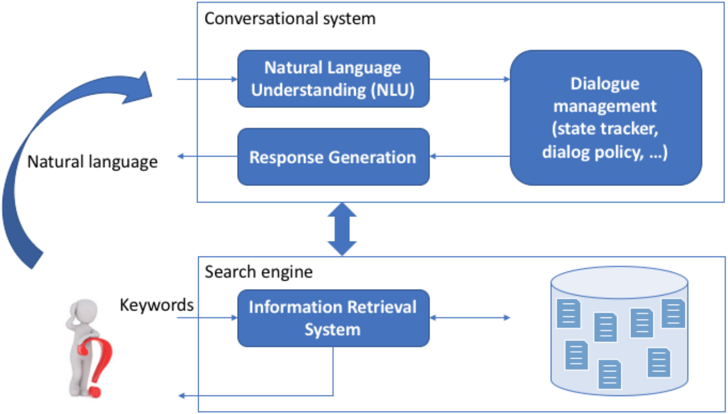

- Chit-chat conversational systems: simple conversations (Li et al. 2016; Ritter, Cherry, and Dolan 2011)
- Task-oriented conversational systems: closed world (slot-filling patterns, KB extractions, ...) (Bordes and Weston 2016; Dhingra et al. 2017; Wang and Lemon 2013)
- Users interact in natural language

# Towards search-oriented conversational systems



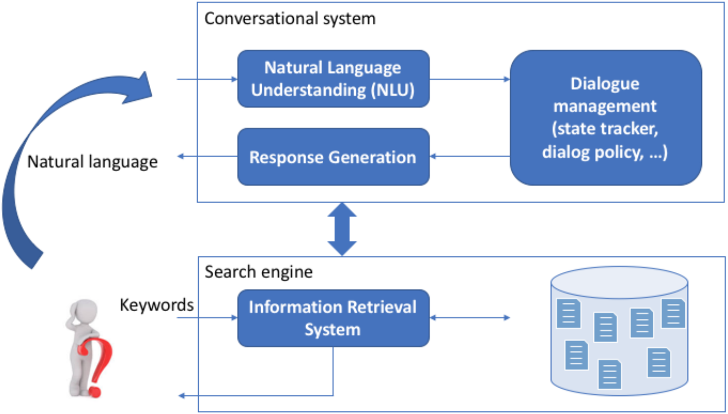

- "Open world"
  - Information needs are often ambiguous, vague, exploratory (Agichtein, Brill, and Dumais 2006; Joachims 2002)
  - Large document datasets, structured and unstructured information sources
- Information needs expressed through keywords

# Towards search-oriented conversational systems




- Challenge: Understand users' information needs expressed in natural language to identify relevant documents
  - Build keyword-based queries from natural language expressions
  - "End-to-end" approach directly dealing with the NL expression

# Towards search-oriented conversational systems




- Challenge: Understand users' information needs expressed in natural language to identify relevant documents
  - Build keyword-based queries from natural language expressions
  - "End-to-end" approach directly dealing with the NL expression

# Related Work

SCIENCES SORBONNE UNIVERSITÉ

- Neural Machine translation models
  - Principle: Encoding the input in a latent representation space and decoding its latent representation in the target language
  - In our context: (Song2017,Yin2017)

# Related Work

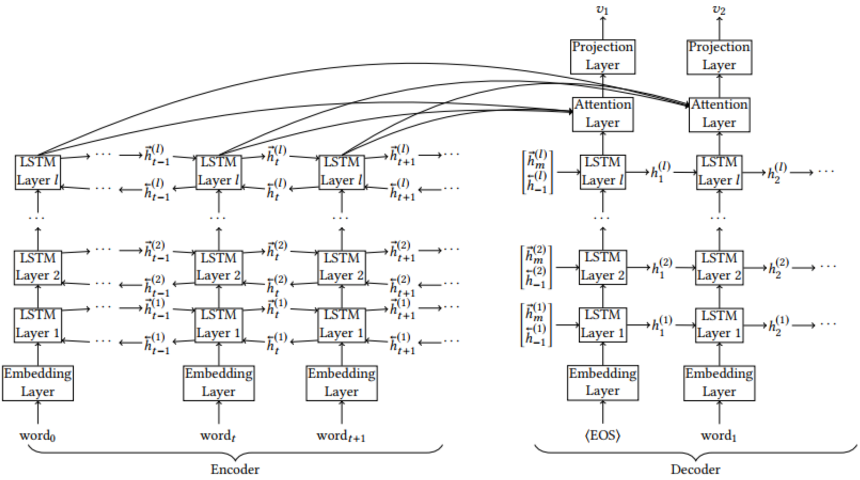
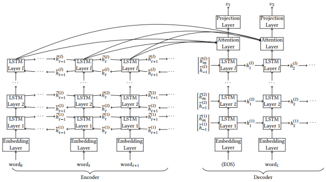



Figure 1: DeepProbe model (Yin, Chang, and Zhang 2017)

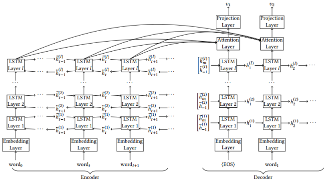
# Related Work

- Neural Machine translation models
  - Principle: Encoding the input in a latent representation space and decoding its latent representation in the target language
  - In our context: (Song2017,Yin2017)



# Related Work

- Neural Machine translation models
  - Principle: Encoding the input in a latent representation space and decoding its latent representation in the target language
  - In our context: (Song2017,Yin2017)

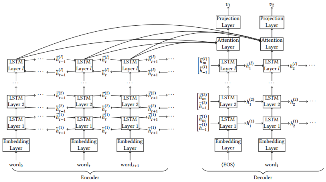



- Reinforcement learning approaches
  - Principle: Driving the approach by the task
  - In our context: (Nogueira2017)

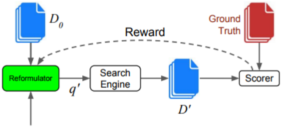
# Related Work

SCIENCES SORBONNE UNIVERSITÉ

- Neural Machine translation models
  - Principle: Encoding the input in a latent representation space and decoding its latent representation in the target language
  - In our context: (Song2017,Yin2017)




- Reinforcement learning approaches
  - Principle: Driving the approach by the task
  - In our context: (Nogueira2017)



# Related Work

SCIENCES SORBONNE UNIVERSITÉ

## Our objective

Bridging these two lines of work:

- Machine translation to learn the mapping between information needs expressed in natural language and information needs formulated using keywords (Song2017,Yin2017)
- Reinforcement learning to inject the task objectives within the machine translation model (Nogueira and Cho 2017)

# Contribution

# Overview

SCIENCES SORBONNE UNIVERSITÉ

- Notations
  - $x = x_1, ..., x_i, ..., x_n$ : NL user's information need.
  - $y \in \{0, 1\}^n$: Key-word query.
  - $y_i = 1$ if $x_i$ exists in query $y$ and 0 otherwise.

**example**

- NL = "Identify documents that discuss sick building syndrome or building related illnesses."
- Q = "sick building syndrome."
- $y = (0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0)$

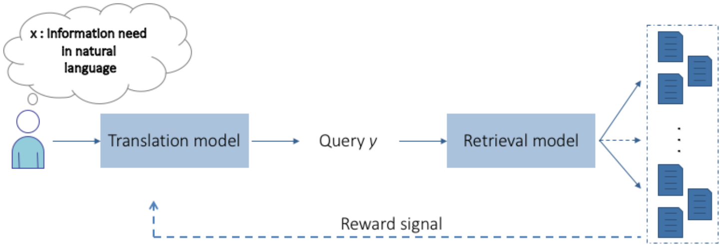
- Problem formulation

  $f_\theta$: estimate the probability $p(y|x)$ of generating the keywords $y$ given its natural language expression $x$, conditioned by $y_{<i}$.

$$p(y|x) = \prod_{y_i \in y} p(y_i | y_{<i}, x)$$

# Overview

SCIENCES SORBONNE UNIVERSITÉ




- The probability p(y|x) is first learned using a maximum likelihood estimation on the NL-query pairs → Translation model
- Then, it is refined using reinforcement learning techniques → IR-oriented reward signal

# Supervised Machine Translation Model



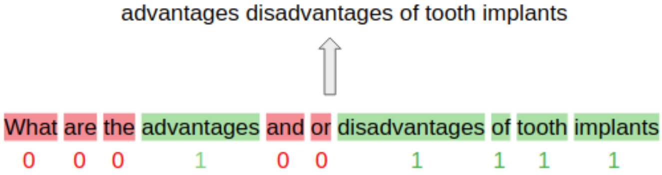



Figure 1: Translation process. 0: word rejected and 1: word selected

## From NL to Queries

NL-query pairs $D = \{(x^1, y^1), ..., (x^k, y^k), ..., (x^N, y^N)\}$

$$L_{SMT} = \sum_{(x^k, y^k) \in D} log(f(\theta, x^k))$$

$$\text{s.t. } f(\theta, x^k) = \sum_{y_i^k \in y^k} log(p(\hat{y}_i^k = y_i^k | \hat{y}_{<i}^k, x^k))$$

# Reinforcement learning

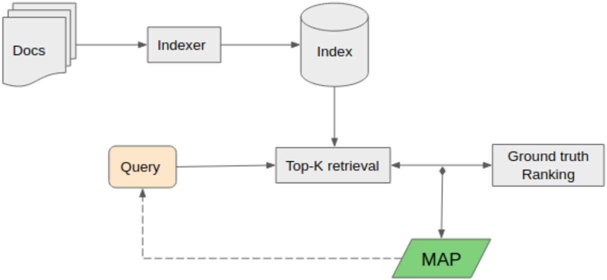


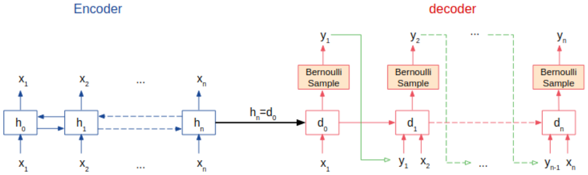

Figure 2: Reward signal

Reward: $R(\hat{y}) = MAP(\hat{y}, \mathcal{D}_x)$ based on documents ranking using $y$.

actions: {0, 1} or {select, discard}

$$L_{RL}(\theta) = arg\max_\theta \; \mathrm{E}_{\substack{(x;\mathcal{D}_x)\in GT \\ \hat{y}\sim f_\theta(x)}} [R(\hat{y}) - \bar{R}]$$

# Neural architecture




- Each element $x_i$ of $x$ is modeled through word embeddings $w_{x_i}$
- Encoder: bi-directional LSTM (hidden state $h_n$)
- Decoder: LSTM
  - Input: hidden vector $h_n$ learned in the encoder network, the current word $x_i$, and a binary indicator $y_{i-1}$ expressing whether previous word $x_{i-1}$ has been selected or not
  - Output: the word selection probability $p(y_i|y_{<i}, x)$

# Experimental evaluation

# Protocol

SCIENCES SORBONNE UNIVERSITÉ

- Evaluation objective: measuring the effectiveness of predicted queries
- Datasets
  - TREC Robust 2004: 250 NL-query pairs, 15.333 documents
  - TREC Web 2000, 2001: 100 NL-query pairs, 11.47 documents

| | |
|---|---|
| Title | Lewis and Clark expedition |
| Description | What are some useful sites containing information about the historic Lewis and Clark expedition? |

Table 1: An example of a NL-query pair

# Protocol

- Evaluation methodology: Run BM25 model on predicted queries (PyLucene)
- Metric: MAP
- Baselines:
  - NL: the NL information need, TREC descriptions
  - Q: original TREC titles
  - Qbin: the binary formulated queries
  - Random: randomly select 3 words from NL
  - SMT: standalone statistical machine translation
  - RL: standalone reinforcement learning

# Implementation details

SCIENCES SORBONNE UNIVERSITÉ

- Word embeddings: FastText pretrained .
- 10-fold cross-validation.
- Encoder: Bi-LSTM with 100 hidden units.
- Decoder: LSTM layer with 100 hidden units.
- Pre-train the supervised translation model for 20 iterations.
- Continue training with RL for 1000 iterations.
- 12 sentences mini-batch Adam algorithm to pre-train the model and SGD for the reinforcement learning part.

# Retrieval effectiveness of our approach

SCIENCES SORBONNE UNIVERSITÉ

| Baseline | TREC Robust(2004) | | TREC Web (2000-2001) | |
|---|---|---|---|---|
| | MAP | *%Chg* | MAP | *%Chg* |
| NL | 0.08925 | +15.25% *** | 0.15913 | +12.88% * |
| Q | 0.09804 | +4.92% | 0.16543 | +8.58% |
| Q bin | 0.08847 | +16.26% * | 0.17402 | +3.22% |
| Random | 0.01808 | +468.91% *** | 0.04060 | +342.44% *** |
| SMT | 0.06845 | +50.27% *** | 0.08891 | +102.04% *** |
| RL | 0.08983 | +14.51% *** | 0.16474 | +9.04% |
| **SMT+RL** | **0.10286** | | **0.17963** | |

Table 2: Comparative effectiveness analysis of our approach. *%Chg*: improvement of **SMT+RL** over corresponding baselines. Paired t-test significance *: $0.01 < t \leq 0.05$ ; **: $0.001 < t \leq 0.01$ ; ***: $t \leq 0.001$.

- Low results for **NL**: benefit of using keywords.
- **SMT+RL** overpasses **SMT**: benefit of reinforcement learning.
- **RL** baseline achieves relatively good retrieval performances.
- Reinforcement learning techniques are more effective with pre-training.

# Qualitative results

SCIENCES SORBONNE UNIVERSITÉ

| NL | Q | Q bin | SMT+RL |
|---|---|---|---|
| what are new methods of producing steel | steel producing | producing steel | new methods of producing steel |
| what are the advantages and or disadvantages of tooth implant | implant dentistry | implant | advantages disadvantages tooth implant |
| find documents that discuss the toronto film festival awards | toronto film awards | toronto film awards | the toronto film festival awards |

Table 3: Examples of query formulation for **NL** queries, the original query **Q**, the binary version **Q bin** of the original query, and our model **SMT+RL**.

- **Q** and **Qbin**: could be considered as oracle
- **SMT+RL** provides higher MAP results than the **Q** and **Q bin** with improvements from +3.22% to +16.26%.
- Queries in **Q** identify the most important words leading to an exploratory query (e.g. "steel productions"), **Our model** provides additional words that precise which facet of the query is concerned (e.g., "new methods of...")

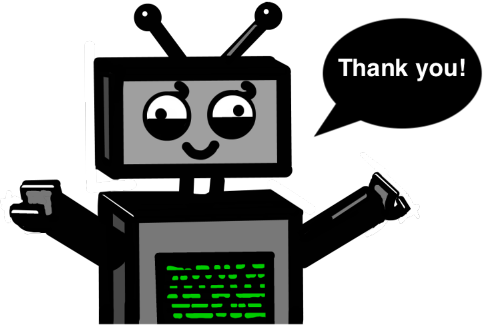
## Conclusion and Perspectives

SCIENCES SORBONNE UNIVERSITÉ

- A selection model to transform the user's information need in NL into a keyword query to increase the retrieval effectiveness in a SOCS context.
- Our model bridges two lines of work dealing with supervised machine translation and reinforcement learning.
- Evaluation on two TREC datasets and promising results in terms of effectiveness.
- For future work:
  (a) Augment the dataset as in Song, Kim, and Park 2017.
  (b) Adapt our model by totally skipping the query formulation step and designing retrieval models dealing with NL expressions.

# Thank you for your attention!

SCIENCES SORBONNE UNIVERSITÉ




Join us for the poster session.

# References

# References I

SCIENCES SORBONNE UNIVERSITÉ


Agichtein, Eugene, Eric Brill, and Susan Dumais (2006). "Improving Web Search Ranking by Incorporating User Behavior Information". In: *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '06, pp. 19–26.

Bordes, Antoine and Jason Weston (2016). "Learning End-to-End Goal-Oriented Dialog". In: *CoRR* abs/1605.07683. url: `http://arxiv.org/abs/1605.07683`.

Dhingra, Bhuwan et al. (2017). "Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access". In: *Association for Computational Linguistics (ACL)*, pp. 484–495.

Joachims, Thorsten (2002). "Optimizing Search Engines Using Clickthrough Data". In: *Proceedings of the SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '02. ACM, pp. 133–142.

Li, Xiujun et al. (2016). "A User Simulator for Task-Completion Dialogues". In: *CoRR* abs/1612.05688. url: `http://arxiv.org/abs/1612.05688`.

Nogueira, Rodrigo and Kyunghyun Cho (2017). "Task-Oriented Query Reformulation with Reinforcement Learning". In: *Neu'IR Workshop - SIGIR*.

# References II

SCIENCES SORBONNE UNIVERSITÉ

Ritter, Alan, Colin Cherry, and William B. Dolan (2011). "Data-driven Response Generation in Social Media". In: *Conference on Empirical Methods in Natural Language Processing*. EMNLP '11.

Song, Hyun-Je, A-Yeong Kim, and Seong-Bae Park (2017). "Translation of Natural Language Query Into Keyword Query Using a RNN Encoder-Decoder". In: *International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17, pp. 965–968.

Wang, Zhuoran and Oliver Lemon (2013). "A Simple and Generic Belief Tracking Mechanism for the Dialog State Tracking Challenge: On the believability of observed information". In: *SIGDIAL 2013 Conference*, 423–432.

Yin, Zi, Keng-hao Chang, and Ruofei Zhang (2017). "DeepProbe: Information Directed Sequence Understanding and Chatbot Design via Recurrent Neural Networks". In: *International Conference on Knowledge Discovery and Data Mining*, pp. 2131–2139.