Robust Information Retrieval



SIGIR 2024 tutorial

Yu-An Liu[a,b], Ruqing Zhang[a,b], Jiafeng Guo[a,b] and **Maarten de Rijke**[c]

https://sigir2024-robust-information-retrieval.github.io/

July 14, 2024

01:30 – 05:00 PM

[a] Institute of Computing Technology, Chinese Academy of Sciences
[b] University of Chinese Academy of Sciences
[c] University of Amsterdam

Section 6:
Conclusions and future directions

Tutorial summary

- **Introduction**
- **Preliminaries**
  - Definition of robustness in IR
  - Taxonomy of robustness in IR
- **Adversarial robustness**
  - Benchmark, settings, task definition and evaluations
  - Adversarial attacks: steal black-box knowledge $\rightarrow$ identify vulnerable positions $\rightarrow$ add adversarial perturbations
  - Adversarial defenses: empirical defense, certified defense and attack detection
- **Out-of-distribution robustness**
  - OOD generalizability on unseen documents: new corpus and incrementation of original corpus
  - OOD generalizability on unseen queries: query variation and unseen query type
- **Robust IR in the age of LLMs**

Robustness: The Achilles' heel of neural IR models

Thetis Dipping Achilles into the River Styx - Antoine Borel (1743-1810)

If robustness is so hard, what can we do with our neural IR systems today?

Mitigating robustness

- **Before going-to-production:** Optimizing training objectives, introducing perturbations in advance
- **While in production:** Customizing analysis tools, monitoring of operational status regularly
- **Post-hoc correction:** Improving system interpretability, optimizing for weaknesses

What about tomorrow?

What about tomorrow?

Much done, much left to do

Future directions for adversarial robustness

There are currently some dilemmas of adversarial robustness in IR that are worthy of future attention in the endeavor:

Future directions for adversarial robustness

There are currently some dilemmas of adversarial robustness in IR that are worthy of future attention in the endeavor:

- **Game theory:** Modeling the market behavior of real SEO
- **Toolkit:** A systematic platform for integrating attack and defense methods
- **Industrial practice:** Considering the deployment in specific operations

Future directions for adversarial robustness: Game theory

Background: In real search engine SEO scenarios, there are multiple attackers, working individually or in groups, with consistent or not-exactly-consistent goals.

Future directions for adversarial robustness: Game theory

Background: In real search engine SEO scenarios, there are multiple attackers, working individually or in groups, with consistent or not-exactly-consistent goals.

Dilemma: It is difficult to analyze the impact of this scaled SEO behavior on search engines, let alone counter them.

Future directions for adversarial robustness: Game theory

Background: In real search engine SEO scenarios, there are multiple attackers, working individually or in groups, with consistent or not-exactly-consistent goals.

Dilemma: It is difficult to analyze the impact of this scaled SEO behavior on search engines, let alone counter them.

Promising way: Game theory

- Multiple attackers seeking to profit is essentially a gaming problem
- Game theory can be used to find an equilibrium in this scenario
- SEO can be curbed by adjusting search engine rules to tilt the balance in favor of the user

Future directions for adversarial robustness: Toolkit

Background: With the development of adversarial robustness, various attacks, defense methods and experimental datasets have emerged.

Future directions for adversarial robustness: Toolkit

Background: With the development of adversarial robustness, various attacks, defense methods and experimental datasets have emerged.

Dilemma: The lack of a unified specification leads to poor direct comparability of methods, which in turn affects the accurate understanding of model robustness.

Future directions for adversarial robustness: Toolkit

Background: With the development of adversarial robustness, various attacks, defense methods and experimental datasets have emerged.

Dilemma: The lack of a unified specification leads to poor direct comparability of methods, which in turn affects the accurate understanding of model robustness.

Promising way: Toolkit

- A high-quality codebase for robust IR research
- A unified data processing pipeline, simplified model configuration and automatic hyper-parameters tuning features equipped

Future directions for adversarial robustness: Industrial practice

Background: Current adversarial attacks and defenses are mainly studied in relatively plain and contained experimental scenarios

Future directions for adversarial robustness: Industrial practice

Background: Current adversarial attacks and defenses are mainly studied in relatively plain and contained experimental scenarios

Dilemma: In real search engines, the situation that may be faced is much more complex, which may make it difficult to apply existing methods on the ground

Future directions for adversarial robustness: Industrial practice

Background: Current adversarial attacks and defenses are mainly studied in relatively plain and contained experimental scenarios

Dilemma: In real search engines, the situation that may be faced is much more complex, which may make it difficult to apply existing methods on the ground

Promising way: Industrial practice

- Foster academic-industrial collaborations on the topic
- Designing appropriate defense mechanisms for realistic and specific SEO scenarios

Future directions for OOD robustness

There are also currently some dilemmas of OOD robustness in IR that are worthy of future attention in the endeavor:

Future directions for OOD robustness

There are also currently some dilemmas of OOD robustness in IR that are worthy of future attention in the endeavor:

- **Causality modeling:** Identifying spurious correlation factors between documents and queries
- **Toolkit:** A systematic platform for integrating OOD documents and queries
- **Industrial practice:** Considering the deployment in specific operations

Future directions for OOD robustness: Causality modeling

Background: Some neural IR models focus on spurious correlations within the domain, leading to poor out-of-distribution performance

Future directions for OOD robustness: Causality modeling

Background: Some neural IR models focus on spurious correlations within the domain, leading to poor out-of-distribution performance

Dilemma: To address this problem, we currently rely on constructing large amounts of new domain data, which has significant overhead.

Future directions for OOD robustness: Causality modeling

Background: Some neural IR models focus on spurious correlations within the domain, leading to poor out-of-distribution performance

Dilemma: To address this problem, we currently rely on constructing large amounts of new domain data, which has significant overhead.

Promising way: Causality modeling

- Causal modeling can effectively identify the key factors in a document that determine the relevance of a query
- When the domain changes, these key factors remain the same

Future directions for OOD robustness: Toolkit

Background: In current approaches, OOD solutions for queries and documents are relatively separate, yet in search engines, these two problems often arise simultaneously

Future directions for OOD robustness: Toolkit

Background: In current approaches, OOD solutions for queries and documents are relatively separate, yet in search engines, these two problems often arise simultaneously

Dilemma: It is currently difficult to analyze the full performance of specific methods under various OOD issues

Future directions for OOD robustness: Toolkit

Background: In current approaches, OOD solutions for queries and documents are relatively separate, yet in search engines, these two problems often arise simultaneously

Dilemma: It is currently difficult to analyze the full performance of specific methods under various OOD issues

Promising way: Toolkit

- A unified experimental platform is needed to accommodate possible OOD problems
- A good solution should perform consistently in a variety of OOD scenarios

Future directions for OOD robustness: Industrial practice

Background: In real search engines, there are more specific requirements for the fitness of neural IR models.

Future directions for OOD robustness: Industrial practice

Background: In real search engines, there are more specific requirements for the fitness of neural IR models.

Dilemma: Current research on OOD is still based on a combination of existing experimental datasets.

Future directions for OOD robustness: Industrial practice

Background: In real search engines, there are more specific requirements for the fitness of neural IR models.

Dilemma: Current research on OOD is still based on a combination of existing experimental datasets.

Promising way: Industrial practice

- Conduct experiments on real data from industrial scenarios, such as corpus increments over time
- Designing the appropriate OOD solutions for realistic and specific search engine scenarios

So much to do . . .

- Developing new techniques for early detection and mitigation of adversarial attacks

So much to do . . .

- Developing new techniques for early detection and mitigation of adversarial attacks
- Exploring synergies between different aspects of robustness, such as adversarial and OOD

So much to do . . .

- Developing new techniques for early detection and mitigation of adversarial attacks
- Exploring synergies between different aspects of robustness, such as adversarial and OOD
- Enhancing model agility to quickly adapt to new data without extensive retraining

So much to do . . .

- Developing new techniques for early detection and mitigation of adversarial attacks
- Exploring synergies between different aspects of robustness, such as adversarial and OOD
- Enhancing model agility to quickly adapt to new data without extensive retraining
- Resources and sharing

There is still a long way to go . . .

What do we talk about when we talk about IR robustness?

What do we talk about when we talk about IR robustness?

"Oh, you mean adversarial robustness? OOD robustness? data distribution? model architecture?"

What do we talk about when we talk about IR robustness?

"Oh, you mean adversarial robustness? OOD robustness? data distribution? model architecture?"

"Actually, I mean this deployed model will not fail next month."

Ultimate goal for robust IR . . .

Built to withstand, designed to last!

Q & A

Thank you for joining us today!

All materials are available at

https://sigir2024-robust-information-retrieval.github.io/

References

References i