Being diverse is not enough: Rethinking Diversity Evaluation to Meet Challenges of News Recommender Systems (NRS)

Célina Treuillier - Sylvain Castagnos - Evan Dufraisse - Armelle Brun
Introduction
- Current challenges of NRS
- Diversity and NRS
- Evaluation of NRS

Experimental analysis
- Dataset selection
- Holistic analysis
- Temporal analysis

Take-home messages
Introduction
News Recommender Systems

Accessed

Recommended

Filter bubble

Limited range of opinions

Introduction
Experimental analysis
Take-home messages
News Recommender Systems

Introduction

Experimental analysis

Take-home messages
Diversity & NRS

- Provide a larger spectrum of opinions (Heitz et al., 2022)
- Ensure ethical and fair recommendations (Lunardi et al., 2020)
- Foster a healthy democratic debate (Giunchiglia et al., 2021)

Role of diversity may be overestimated and must be finely controlled to help reduce polarization

⇒ RQ1: Does diversity of recommendations bring a systematic gain?
Evaluation of NRS

*News characteristics:* short lifespan, high turnover...

*Democratic role of NRS* (Helberger, 2019)

Adaptation of recommendation and evaluation

⇒ **RQ2:** Is it sufficient to measure the influence of an NRS afterward with single-number metrics, or does this influence occur with some variations over time?
Experimental analysis
Large-scale dataset for news recommendation research

Information about news

Users’ interactions

5 weeks: October to November, 2019

Introduction

Experimental analysis

Take-home messages

Data selection ➔ 1,475 users & 20,541 news

From URL

News representation

Date of publication

User ID

History

Recommended

Accessed

LDA (Latent Dirichlet Allocation)

Week 1 to 4

Week 5
MIND dataset

Average diversity (Smyth & McClave, 2001):

\[
\text{Diversity}(i_1, i_2, ..., i_n) = \frac{\sum_{k=1}^{n} \sum_{j=1}^{n} (1 - \text{Similarity}(i_k, i_j))}{\frac{n}{2} \times (n - 1)}
\]

1. **Holistic analysis**
   - News of history
   - Recommended news
   - Accessed news
   - Unaccessed news

2. **Temporal analysis**
   - News accessed each week (1 to 5)
Analysis - **holistic**

Distribution of diversity among users

- High average diversity (0.75) & small standard deviation (0.04)
- Average diversity significantly lower
- Great variability
Analysis - *holistic*

**Conclusions:**

- Recommender system meets a predefined diversity level
- Impact of NRS on the news consumption
- Diversity input is not personalized

⇒ High diversity of recommendations does not systematically lead to a diverse news consumption *(RQ1)*
Analysis - *temporal*

Transition patterns remain stable over the weeks

⇒ Similar global impact of recommendations on users diversity of accessed news through weeks

*Sankey diagram: variation flows over weeks*
Analysis - *temporal*

**Conclusions:**

- Average diversity of accessed news differs over weeks
- Users observe diversity variations
- Impact is not the same for all users

3 types of users:
- Positively receptive users
- Negatively receptive users
- Resistant users

⇒ Single-number evaluations are insufficient

Need to take differences between users and temporal aspect into account (**RQ2**)

---

**Introduction**  
**Experimental analysis**  
**Take-home messages**
Overall conclusion

1. Diverse recommendations \(\not\Rightarrow\) diverse consumption
   
   Need to adapt the evaluation

2. NRS does not impact all users equally
   
   Different classes of user behavior
Take-home messages
Take-home messages

1. Need of well-established methodologies to model users’ diversity trajectories

2. Need of adapted diversity measures and personalized recommendation strategies

3. Lack of open datasets
Contact me!
celina.treuillier@loria.fr
References


References


