AI for personalization: from predictive to generative modeling.

Gabriel Bénédict

AI for personalization: from predictive to generative modeling.

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. G.T.M. ten Dam
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in
de Aula der Universiteit
op vrijdag 01 January 2024, te 14:00 uur

door

Gabriel Bénédict

geboren te Chavannes-Pres-Renens, Zwitserland

Promotiecommissie

Promotor:
Prof. dr. M. de Rijke   Universiteit van Amsterdam
Co-promotor:
dr. D. Odijk   RTL NL

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

This research was (partially) funded by Bertelsmann SE & Co. KGaA and by the Hybrid Intelligence Center, a 10-year program funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, `https://hybrid-intelligence-centre.nl`.

Cover by Off Page, Amsterdam
Printed by Off Page, Amsterdam

ISBN:

Certains pensent qu'ils font un voyage, en fait, c'est le voyage qui vous fait ou vous défait.

– Nicolas Bouvier

Acknowlegements

Gabriel Bénédict
Amsterdam
January 2024

Table of Contents

Chapter 1

Introduction

write

Define a streaming platform

The user journey on a streaming platform is a unique experimentation platform: survey, sampling, modeling, (adaptive) testing, behavioral analysis.

personalization

pipeline

generative recommendation

multilabel classification

intent

RADio

1.1 Research Outline and Questions

We scope the thesis around four research questions, each corresponding to a chapter in the thesis.

Personalization on streaming platforms is oftentimes perceived as a purely predictive phenomenon: we propose to view it as a comprehensive and responsible generative approach, throughout a pipeline. We introduce RecFusion to issue recommendations in a generative way with diffusion models, as part of the nascent Generative Information Retrieval field. For these recommendations, we propose a method to generate personalized stills from movies, with sigmoidF1. We show that the resulting interactions on platforms are also dependent on implicit data hidden from a web analytics platform, with our intent-satisfaction analysis. At the end of the pipeline, we propose to ensure normative diversity in the issued recommendations with RADio, a normative metrics framework.

Our first research question addresses the entry point of the user journey on a personalized platform, namely recommendations (on the main page).

RQ1 Can we use diffusion to do recommendation in the classical user-item matrix setting?

Assuming recommendations are to be linked with a certain form of creativity, we harness here the power of generative models. The emergent diffusion models field has been applied to images, music and other modalities. Unlike these, the classical recommendation setting of the user-item matrix does not entail spatial relationships between data points: contrary to pixels on an image, there is no information encoded in the allocation of users and item on a matrix. Can diffusion still be applied in this setting? The binary nature of the classical user-item matrix formulation inspired a second more theoretical research question: *can Bernoulli diffusion be a suitable forward and backward process theoretically and empirically on binary data?*

RQ2 Is there a way we can generate personalized posters and stills for each item on a streaming platform?

Can we formulate loss function that accommodates for multilabel classification at training time and operates on the whole batch to balance confusion matrix entries?

RQ3 Are users' intents together with their behavioral data useful signals to predict or explain satisfaction on a video streaming platform?

Can we use Bayesian posterior draws to meaningfully draw conclusions from data?

RQ4 Can we formulate a divergence metric that measures the normative diversity of recommendations?

This work applies to news recommendation but trivially generalizes to any domain that has categories (e.g. video streaming with movie genres).

Can we formulate a divergence metric that is distributional and rank-aware?

1.2 Main Contributions

In this section, we summarize the main contributions of this thesis. We separate theoretical from artifact (tool and experimental design) contributions.

[Gab: I found this 'artifact' expression here, but I am not sure it is commonly used]

Theoretical Contributions

- Diffusion applied on unstructured data, where there is no spatial dependency (Chapter 2).
- Diffusion for binary and/or 1D data: A demonstration that KL divergence is also suited for binary data and that the Bernouilli Markov Process has the same properties as its Gaussian counterpart (Chapter 2).
- A multilabel loss function that accounts for all examples in a batch (Chapter 3).
- An F1 score surrogate as a loss function (Chapter 3).
- An account of the current limitations and underreporting of thresholding at inference time (Chapter 3).

- a proposal of typical intents for a video streaming that we divide into explorative and decisive categories (Chapter 4).
- A diversity metric that is versatile to any normative concept and expressed as the divergence between two (discrete) distributions, rank-aware and mathematically grounded in distributional divergence statistics (Chapter 5).

Artifact Contributions

- Frequentist logistic regression model, we test Bayesian multilevel models for visualization and explanations, along with random forests for improved accuracy (Chapter 4).
- A reproducibility study from music to video streaming platforms of intent-satisfaction modeling (this time with code and synthetic data) (Chapter 4).
- In-app survey design for a medium size streaming platform (∼1 million users) and corresponding synthetic data (Chapter 4).
- A metadata enrichment pipeline (e.g., sentiment analysis, named entity recognition) to extract normative concepts from news articles (Chapter 5).

1.3 Thesis Overview

1.4 Origins

We list the publications that are the origins of each chapter below.

Chapter 1 is based on the following paper:

- Gabriel Bénédict, Olivier Jeunen, Samuele Papa, Samarth Bhargav, Daan Odijk, and Maarten de Rijke. Recfusion: A binomial diffusion process for 1d data for recommendation, 2023.

GB wrote the first draft, code, experiments and mathematical formulations. GB was helped all along the way via discussion with the coauthors. Coauthors then edited the first draft together with GB. GB did most of the writing.

Chapter 2 is based on the following paper:

- Gabriel Bénédict, Hendrik Vincent Koops, Daan Odijk, and Maarten de Rijke. sigmoidf1: A smooth f1 score surrogate loss for multilabel classification. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL `https://openreview.net/forum?id=gvSHaaD2wQ`.

GB wrote the first draft, code, experiments and mathematical formulations. GB was helped all along the way via discussion with the coauthors. Coauthors then edited the first draft together with GB. GB did most of the writing.

Chapter 3 is based on the following paper:

- Gabriel Bénédict, Daan Odijk, and Maarten de Rijke. Intent-satisfaction modeling: From music to video streaming. *ACM Trans. Recomm. Syst.*, 1(3), aug 2023. doi: 10.1145/3606375. URL `https://doi.org/10.1145/3606375`.

GB wrote the first draft, code, experiments and mathematical formulations. GB was helped all along the way via discussion with the coauthors. Coauthors then edited the first draft together with GB. GB did most of the writing.

Chapter 4 is based on the following paper:

- Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. Radio – rank-aware divergence metrics to measure normative diversity in news recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, page 208–219, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392785. doi: 10.1145/3523227.3546780. URL `https://doi.org/10.1145/3523227.3546780`.

GB, together with SV, wrote the first draft, code, experiments and mathematical formulations. GB was helped all along the way via discussion with the coauthors. Coauthors then edited the first draft together with GB. GB and SV did most of the writing.

The writing of this thesis also benefited from work on the following publications:

- Gabriel Bénédict, Ruqing Zhang, and Donald Metzler. Gen-ir @ sigir 2023: The first workshop on generative information retrieval, 2023.
- Ali Vardasbi, Gabriel Bénédict, Shashank Gupta, Maria Heuss, Pooya Khandel, Ming Li, and Fatemeh Sarvi. The university of amsterdam at the trec 2021 fair ranking track. *TREC Fair Ranking*, 2021.
- Gabriel Bénédict. Generative adversarial networks, 2021. URL `https://spectra.mathpix.com/article/2021.09.00009/gans`.

CHAPTER 2

Generative Recommendations with Diffusion

RQ1: Can we use diffusion to do recommendation in the classical user-item matrix setting?

Reproducibility

To facilitate the reproducibility of the work in this chapter, our code is available at `https://github.com/gabriben/recfusion`.

This chapter is under submission at International Conference on Learning Representations (ICLR) under the title "RecFusion: A Binomial Diffusion Process for 1D Data for Recommendation" [?].

Chapter 3

Metrics as Losses

In this chapter, we address the following research question:

RQ2: Is there a way we can generate personalized posters and stills for each item on a streaming platform?

Reproducibility

To facilitate the reproducibility of this work, our code is available at `https://github.com/gabriben/metrics-as-losses`.

This chapter was published at the Transactions of Machine Learning Research (TMLR) under the title "sigmoidF1: A Smooth F1 Score Surrogate Loss for Multilabel Classification" [1].

Chapter 4

Intent, Behavior and Satisfaction

In this chapter, we address the following research question:

RQ3: Are users' intents together with their behavioral data useful signals to predict or explain satisfaction on a video streaming platform?

Reproducibility

To facilitate the reproducibility of the work in this chapter, our code is available at `https://github.com/rtlnl/streaming-intent-model`.

This chapter was published at the ACM Transactions on Recommender Systems (TORS) under the title "Intent-Satisfaction Modeling: From Music to Video Streaming" [2]

Chapter 5

Normative Diversity

In this chapter, we address the following research question:

RQ4: Can we formulate a divergence metric that measures the normative diversity of recommendations?

Reproducibility

To facilitate the reproducibility of the work in this chapter, our code is available at `https://github.com/svrijenhoek/RADio`.

This chapter was published at the ACM Conference on Recommender Systems (RecSys 2022) under the title "RADio – Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations" [7], , where it won a best paper runner up award.

Conclusions

In this thesis...

6.1 Main Findings

In this section, we describe our main findings across the three parts of the thesis.

The first part of this thesis focused on ... In Chapter 2, we asked our first research question:

RQ1 Can we use diffusion to do recommendation in the classical user-item matrix setting?

The answer to **RQ1** is yes:

In Chapter 3, we then turned to our next research question:

RQ2 Is there a way we can generate personalized posters and stills for each item on a streaming platform?

The answer to **RQ2** is yes:

In Chapter 4, we investigated the following research question:

RQ3 Are users' intents together with their behavioral data useful signals to predict or explain satisfaction on a video streaming platform?

The answer to **RQ3** is yes:

We asked our final research question in Chapter 5:

RQ4 Can we formulate a divergence metric that measures the normative diversity of recommendations?

We answered **RQ4** by ...

6.2 Future Directions

In this section, we describe some limitations of the methods proposed in this thesis and identify potential avenues for future work.

Final thoughts

Bibliography

[1] Gabriel Bénédict, Hendrik Vincent Koops, Daan Odijk, and Maarten de Rijke. sigmoidf1: A smooth f1 score surrogate loss for multilabel classification. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=gvSHaaD2wQ. (Cited on page 7.)

[2] Gabriel Bénédict, Daan Odijk, and Maarten de Rijke. Intent-satisfaction modeling: From music to video streaming. *ACM Trans. Recomm. Syst.*, 1(3), aug 2023. doi: 10.1145/3606375. URL https://doi.org/10.1145/3606375. (Cited on page 9.)

[3] Gabriel Bénédict. Generative adversarial networks, 2021. URL https://spectra.mathpix.com/article/2021.09.00009/gans.

[4] Gabriel Bénédict, Olivier Jeunen, Samuele Papa, Samarth Bhargav, Daan Odijk, and Maarten de Rijke. Recfusion: A binomial diffusion process for 1d data for recommendation, 2023.

[5] Gabriel Bénédict, Ruqing Zhang, and Donald Metzler. Gen-ir @ sigir 2023: The first workshop on generative information retrieval, 2023.

[6] Ali Vardasbi, Gabriel Bénédict, Shashank Gupta, Maria Heuss, Pooya Khandel, Ming Li, and Fatemeh Sarvi. The university of amsterdam at the trec 2021 fair ranking track. *TREC Fair Ranking*, 2021.

[7] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. Radio – rank-aware divergence metrics to measure normative diversity in news recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, page 208–219, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392785. doi: 10.1145/3523227.3546780. URL https://doi.org/10.1145/3523227.3546780. (Cited on page 11.)

Summary

Samenvatting