By,
Rahul Crunal Kalaria
r.c.kalaria@student.tudelft.nl

Exploring the use of demographic data and domain-specific similarity measures to address the cold-start problem in recommender systems

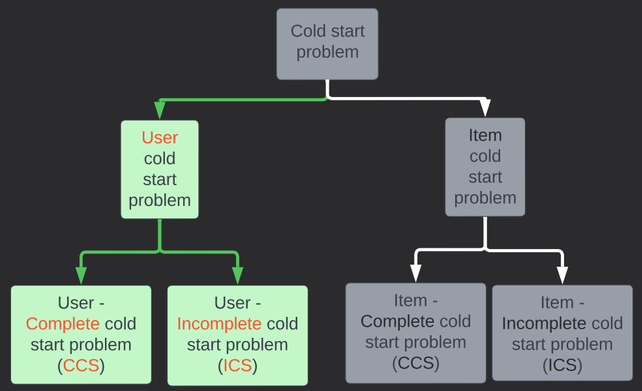
Supervisors:
Frans Oliehoek
Aleksander Czechowski
Oussama Azizi
Davide Mambelli

01. The Cold-start problem

Recommender systems (RS's) are becoming increasingly important for many businesses as they can aid in the creation of unique and personalised experiences for every user.

There are plenty of new approaches to develop RS's. However, collaborative filtering (CF) methods are some of the most widely used and successful systems [1].

Despite the largescale successes of CF models, they often suffer from complete cold start (CCS) problems where no ratings records are available and incomplete cold start (ICS) problems where only a small amount of ratings are available [2].



02. Research questions

This research focuses on the user side of the ICS and CCS.

Main research question:
Can the combined use of user demographic data and a domain-aware similarity measure solve the CCS and ICS problems in collaborative filtering based recommender systems?'

Sub research questions:
- Strengths and limitations of using user demographic data to solve the cold-start problem.
- Strengths and limitations of using domain aware similarity measure to solve the cold-start problem.

03. Method

A reverse ablation study. First, a baseline model will be implemented. Then in an iterative fashion the baseline will be augmented with various ideas and concepts which are designed to deal with the CCS and ICS problems.

Research steps:
- Create a baseline CF (BCF) recommender system.
- Create a demographic filtering model (DCF) which incorporates the use of user demographics to deal with the CCS problem [3].
- Create a PIP collaborative filtering model (PIPCF) which incorporates the use of a domain-specific similarity measure to deal with the ICS problem [4].
- Create a final model which deals with CCS and ICS by combining DCF and PIPCF models

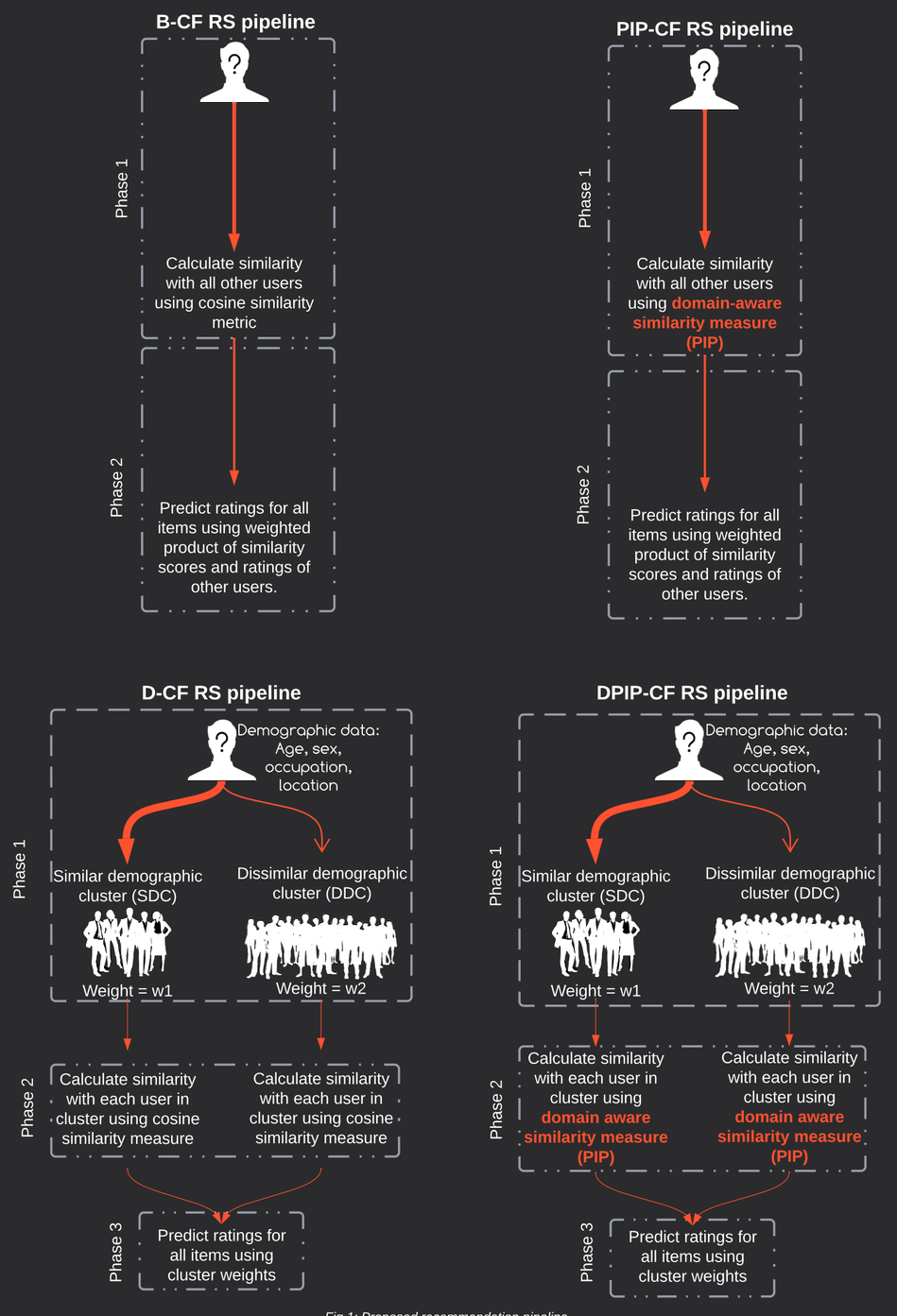


Fig 1: Proposed recommendation pipeline

04. Results

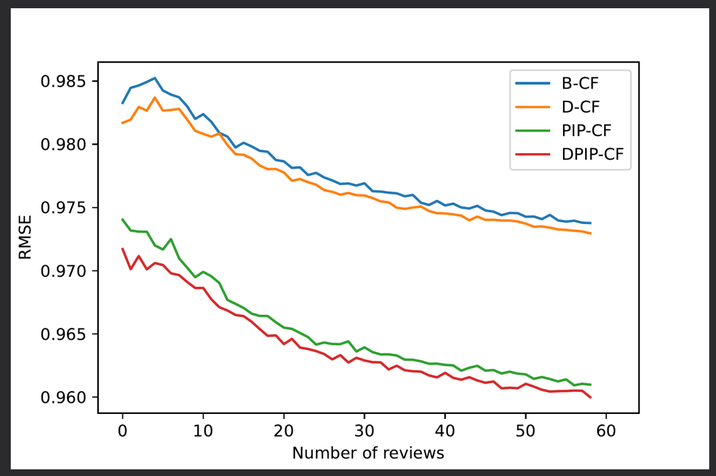


Fig 2: performance comparison of various models using RMSE metric

05. Conclusion

- Use of demographic data shows greatly imporoved performance under CCS conditions.
- Use of domain-aware simialrity measure shows improved performance under ICS conditions
- Use of both demographic and domain-aware similarity measure shows the best performance for CCS and ICS conditions.

06. Potential future work

- Use of a **content-based recommender** system as a subroutine to prevent scenarios in which there are no other users who have rated the same items as the test user. The subroute can be used to find users who have rated **similar** items and not just users who have rated the exact same items.
- Dynamically scale the weights (w1 and w1) for the final DPIP-CF model based on how many items the test user has reviewed. If the test user is "fresh" (i.e has every few reviews), w1 should be larger then w1, else w1 and w2 should have roughly the same weight.

07. References

[1] H. Al-bashiri, M. Abdulgabber, A. Romli, and F. Hujainah, "Collaborative filtering recommender system: Overview and challenges,"Advanced Science Letters, vol. 23, pp. 9045–9049, Sep. 2017. doi:10.1166/asl.2017.10020.
[2] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items,"Expert Systems with Applications, vol. 69, pp. 29–39, 2017,issn: 0957-4174.doi:https://doi.org/10.1016/j.eswa.2016.09.040. [Online].Available:https://www.sciencedirect.com/science/article/pii/S0957417416305309.
[3] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems,"Expert Systems with Applications, vol. 41, no. 4, Part 2, pp. 2065–2073, 2014, issn: 0957-4174.doi:https://doi.org/10.1016/j.eswa.2013.09.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417413007240.
[4] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem,"Information Sciences, vol. 178, no. 1, pp. 37–51, 2008,issn: 0020-0255.doi: https : / / doi . org / 10 . 1016 / j . ins . 2007 . 07 . 024. [Online]. Available:https : / / www . sciencedirect.com/science/article/pii/S0020025507003751

TU Delft