Robustness of Recommender Systems

Neil Hurley
School of Computer Science and Informatics
Complex Adaptive Systems Laboratory
University College Dublin, Ireland
neil.hurley@ucd.ie

ABSTRACT

The possibility of designing user rating profiles to deliberately and maliciously manipulate the recommendation output of a collaborative filtering system was first raised in 2002. One scenario proposed was that an author, motivated to increase recommendations of his book, might create a set of false profiles that rate the book highly, in an effort to artificially promote the ratings given by the system to genuine users. Several attack models have been proposed and the performance of these attacks in terms of influencing the system predictions has been evaluated for a number of memory-based and model-based collaborative filtering algorithms. Moreover, strategies have been proposed to enhance the robustness of existing algorithms and new algorithms have been proposed with built-in attack resistance. This tutorial will review the work that has taken place in the last decade on robustness of recommendation algorithms and seek to examine the question of the importance of robustness in future research.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval ■ Information filtering

General Terms

Algorithms, Experimentation, Performance, Reliability

1. INTRODUCTION

The initial collaborative recommendation algorithms were based a benign world view in which a community of raters supply fair and honest opinions on the quality of the content it rates. However, the real world is made up of many different types of people, who may have motivations and agendas that run counter to the goal of providing accurate ratings to the community as a whole. The problem of *spamdexing*, in which users try to deliberately manipulate search engine rankings, has been combated since the early days of search engine design, famously leading to Google's link ranking algorithm, that provided robustness to one type of content-based attack, but did not make the problem go away. In the specific context of recommender system design, the possibility of designing user rating profiles to deliberately and maliciously manipulate the recommendation output of a collaborative filtering system was first raised in [11]. Since then, these attacks have been dubbed as *shilling* attacks [5] or *profile injection* attacks [16] and a body of research has emerged that has investigated the *robustness* of recommender system algorithms in the face of such attacks.

It is important to note that this research has focussed on the impact that maliciously designed rating profiles can have on the output of recommendation algorithms and on the detection and deletion of such profiles, based on the statistical characteristics of the ratings in the profiles. It is not concerned with extrinsic security measures that a real-world system can employ to verify the identity of its raters, or prevent automated rating input. Rather, it is concerned with the intrinsic ability of a recommendation algorithm to deal with tainted data.

RecSys'11, October 23–27, 2011, Chicago, Illinois, USA.
ACM 978-1-4503-0683-6/11/10.

2. TUTORIAL OUTLINE

In this tutorial, we review the research in robust recommendation that has emerged over the last decade. In general, this work breaks down in the following manner:

1. Early work such as [11, 10, 5, 12, 2], that focussed on identifying different profile injection attack strategies and empirically evaluating their effect on memory-based collaborative filtering algorithms;

2. Work such as [1, 16, 6, 4] that has focussed on detecting attack profiles in order to filter them from the database;

3. Work such as [9, 8, 3] that has extended robustness analysis to model-based algorithms;

4. Work such as [13, 7, 14, 15] that has examined manipulation-resistant recommendation algorithms, and provided some theoretical analysis of the cost-effectiveness of attacks.

The tutorial will overview the main results that have emerged from this research and discuss its implications for recommender system design in general. Finally, robustness will be discussed in relation to other related concepts such as trust and privacy.

3. ACKNOWLEDGMENTS

My work is supported by the Science Foundation Ireland under Grant No. 08/SRC/I1407: Clique: Graph & Network Analysis Cluster.

4. REFERENCES

[1] R. Burke, B. Mobasher, and C. Williams. Classification features for attack detection in collaborative recommender systems. In *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining*, pages 17–20, 2006.

[2] R. Burke, B. Mobasher, R. Zabicki, and R. Bhaumik. Identifying attack models for secure recommendation. In *Beyond Personalization: A Workshop on the Next Generation of Recommender Systems*, 2005.

[3] Z. Cheng and N. Hurley. Robust collaborative recommendation by least trimmed squares matrix factorization. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 2, pages 105 –112, oct. 2010.

[4] N. Hurley, Z. Cheng, and M. Zhang. Statistical attack detection. In *Proceedings of the third ACM conference on Recommender systems*, RecSys '09, pages 149–156, New York, NY, USA, 2009. ACM.

[5] S. K. Lam and J. Riedl. Shilling recommender systems for fun and profit. *In Proceedings of the 13th International World Wide Web Conference*, pages 393–402, May 17–20 2004.

[6] B. Mehta, T. Hofmann, and P. Fankhauser. Lies and propaganda: Detecting spam users in collaborative filtering. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 14–21, 2007.

[7] B. Mehta and W. Nejdl. Attack resistant collaborative filtering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 75–82, New York, NY, USA, 2008. ACM.

[8] B. Mehta and W. Nejdl. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User-Adapted Interaction*, 19(1-2):65–97, 2009.

[9] B. Mobasher, R. D. Burke, and J. J. Sandvig. Model-based collaborative filtering as a defense against profile injection attacks. In *AAAI*. AAAI Press, 2006.

[10] M. P. O'Mahony, N. J. Hurley, and C. C. M. Silvestre. An evaluation of neighbourhood formation on the performance of collaborative filtering. *Artificial Intelligence Review*, 21(1):215–228, March 2004.

[11] M. P. O'Mahony, N. J. Hurley, and G. C. M. Silvestre. Promoting recommendations: An attack on collaborative filtering. In A. Hameurlain, R. Cicchetti, and R. Traunmüller, editors, *DEXA*, volume 2453 of *Lecture Notes in Computer Science*, pages 494–503. Springer, 2002.

[12] M. P. O'Mahony, N. J. Hurley, and G. C. M. Silvestre. Recommender systems: Attack types and strategies. *In Proceedings of the 20th National Conference on Artificial Intelligence (AAAI-05)*, pages 334–339, July 9–13 2005.

[13] P. Resnick and R. Sami. The influence limiter: provably manipulation-resistant recommender systems. In *RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems*, pages 25–32, New York, NY, USA, 2007. ACM.

[14] B. V. Roy and X. Yan. Manipulation robustness of collaborative filtering. *Management Science*, 56(11):1911–1929, 2010.

[15] L.-H. Vu, T. G. Papaioannou, and K. Aberer. Impact of trust management and information sharing to adversarial cost in ranking systems. In *Trust Management IV - 4th IFIP WG 11.11 International Conference, IFIPTM 2010, Morioka, Japan, June 16-18, 2010. Proceedings*, volume 321, pages 108–124. Springer, 2010.

[16] C. Williams, B. Mobasher, and R. Burke. Defending recommender systems: detection of profile injection attacks. *Service Oriented Computing and Applications*, pages 157–170, 2007.