Applying exploratory factor analysis to the Serendipity 2018 dataset to validate a model of surprise in recommender systems

Andre Paulino de Lima
andre.p.lima@usp.br
University of São Paulo
São Carlos, São Paulo, Brazil

Marcelo Garcia Manzato
mmanzato@icmc.usp.br
University of São Paulo
São Carlos, São Paulo, Brazil

ABSTRACT

The research community on recommender systems has recently turned its attention to the beneficial aspects of serendipity, a property that has been positively associated with user satisfaction. However, measurement models of some component factors of serendipity, namely novelty, unexpectedness, and surprise, are mostly non-validated theoretical models. In this project, we use the Serendipity 2018 dataset to validate a model of surprise in the literature. This is attained by adopting a psychometric approach. First, we perform an exploratory factor analysis to determine the fittest factor model underlying the participant responses, and dissect the questionnaire that was employed to survey the participants to interpret the meaning of each resulting factor. Second, we use the induced factor model to obtain surprise scores for each participant response, and derive an empirical distribution of these scores. Finally, we compare the obtained empirical distribution to a set of distributions obtained from applying a theoretical model of surprise in the literature. Preliminary results show that (a) a model that postulates serendipity as an interaction between relevance and surprise better explains the covariances in the Serendipity 2018 dataset than a model in which serendipity is posited as an interaction between relevance, unexpectedness and novelty; and (b) the empirical distribution of surprise scores is seemingly compatible to distributions obtained from a theoretical model of surprise described in the literature.

CCS CONCEPTS

• **Information systems → Recommender systems**; *Novelty in information retrieval.*

KEYWORDS

Recommender System; Serendipity; Surprise; Novelty; Unexpectedness; Exploratory Factor Analysis; Scale Development

ACM Reference Format:
Andre Paulino de Lima and Marcelo Garcia Manzato. 2019. Applying exploratory factor analysis to the Serendipity 2018 dataset to validate a model of surprise in recommender systems. In *Proceedings of 1st ACM Latin American School on Recommender Systems (LARS'19).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

LARS'19, October 10–12, 2019, Fortaleza, Brazil

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Owing to its link to user satisfaction [14] and its potential to allay the filter bubble effect [17] and to improve explanations provided to the user about a recommendation [22], serendipity has received an increasing attention in the literature on recommender systems in recent years. Seen as a property of a recommendation, it is predominantly described as an interaction between other factors, namely relevance, novelty, unexpectedness, and surprise [3, 10, 13].

Although the methods to collect user feedback about the relevance of an item are well developed [12, 18, 21], the same cannot be said about the other factors. This state of affairs reflects, at least in part, the existence of competing definitions for each of the factors [13] and the large conceptual overlap across them [11]. In truth, most of the models of novelty, unexpectedness, and surprise are theoretical in character, as they have been posited by different authors seeking diverse perspectives [3], and a body of empirical evidence is required to confirm or guide the refinement of these models.

In this project, we seek to induce a model of surprise from the growing body of empirical evidence to assess and refine the theoretical models in the literature, specifically the models of surprise. We take advantage of the Serendipity 2018 dataset, which is the first public resource in our community that includes a large amount of user feedback about different components of serendipity. The strategy we follow consists of (a) applying exploratory factor analysis[1] to the Serendipity 2018 dataset to show that the instrument (i.e., the questionnaire used to collect responses from the participants) can efficiently obtain measures of relevance and surprise; (b) using the resulting factor model to obtain a measurement of surprise for each response collected in the dataset; and (c) employing the obtained measurements to assess a model of surprise in the literature [3].

In the remainder of this extended abstract, we briefly discuss different models of serendipity in the literature and their component factors (Section 2); we also shortly review some important aspects of how exploratory factor analysis has been employed in the HCI and IS literatures to uncover latent factors from data collected in user studies. In Section 3, we detail the methodology we intend to follow to validate a model of surprise in the literature. Finally, in Section 4, we present some promising preliminary results that speaks to the feasibility of the proposed methodology.

[1]Exploratory Factor Analysis (EFA) is a statistical technique used in many research areas, such as HCI [8, 16, 23], IS [19, 20], marketing research [9], and psychology [5, 6], to identify latent factors that can account for observed communalities and covariances among observed variables in empirical data. This technique is similar to PCA in that it is also relies on factorisation to reduce dimensionality of the observed data, but differs from it in assuming that a share of the observed covariance should be assigned to errors in measurement [1, 4]. In this project, exploratory factor analysis is applied to show that the responses given by the participants to the 9-items questionnaire used to collect the Serendipity 2018 dataset fits a model of serendipity as an interaction between relevance and surprise.

2 RELATED WORK

There are two predominant models of serendipity in the literature on recommender systems: one describes serendipity by means of a formative model[2] with three factors, namely relevance, novelty and unexpectedness [13], and the other describes serendipity by means of a formative model that combines two factors: relevance and surprise [10]. As mentioned, current methods to measure relevance are reasonably well developed [12, 18, 21], but there is a lack of validated instruments (e.g., questionnaires) that can be used to measure the remaining factors. Moreover, there is a substantial overlap among the conceptual definitions of novelty, unexpectedness and surprise [3, 11, 14], which makes it difficult to draw discerning boundaries and to develop instruments to measure them separately. An attempt has recently been made by the authors in [13], who employed a 9-items questionnaire to capture the feedback of 481 users of the online movie recommender systems Movielens about different notions of novelty and unexpectedness. One of the contributions of that study is the Serendipity 2018 dataset.

Being of central importance to this project and relatively unfamiliar to researchers, some words are necessary about the use of factor analysis in the neighbouring research fields of HCI and IS. In HCI, it has been used to refine the instruments used to measure usability [15, 16], and to propose an investigative process that could lead to higher research productivity (i.e., a disciplined process comprising the formal steps of conceptualisation, operationalisation, and measurement) [8, 23]. In IS, it has been used to develop and validate models of constructs of theoretical interest [19]. These benefits are enabled by (a) the procedural framework underlying factor analysis, which imposes discipline in theorisation, and (b) the statistical tools that allows one to select an explanatory (factor) model that better fits the available empirical data. We believe that the use of factor analysis can benefit the investigation of serendipity and its component factors because we share with our neighbouring fields the challenge of developing and validating models of interactional constructs and now we may have a body of empirical evidence that is large enough to assess some models in the literature[3].

3 PROPOSED METHODOLOGY

First, we perform an exploratory factor analysis of the Serendipity 2018 dataset to determine the fittest latent model underlying the participant responses, and dissect the questionnaire that was employed to survey the participants to interpret the meaning of the resulting factors. Second, we use the induced factor model to obtain normalised measurements of surprise for each participant response. These measurements are produced by computing the factor scores for the latent factor [4] that is related to surprise. Moreover, an empirical distribution of surprise scores is derived from the obtained measurements. Finally, the empirical distribution is compared to distributions obtained from a model of surprise in the literature [3] under different assumptions on how items are represented (i.e., item vectors are produced by means of factorisation of the rating matrix and by applying dimensionality reduction to the document-term matrix) and compared (i.e., item vectors are compared using diverse distance functions). We expect that the empirical and some theoretical distributions will approximately match, which will add support to the validity and predictive power of the assessed model.

4 PRELIMINARY RESULTS

Preliminary results indicate that a factor model that posits serendipity as an interaction between relevance and surprise (Figure 1) better explains the communalities and specific covariances observed in the Serendipity 2018 dataset than a model with relevance, unexpectedness and novelty. With respect to the empirical distribution of surprise scores (Figure 2, left), it is seemingly compatible to some theoretical distributions obtained from the Movielens 1M dataset (Figure 2, right; from a previous study conducted by the first author [3]) by visual inspection, but a refined comparison is being conducted.

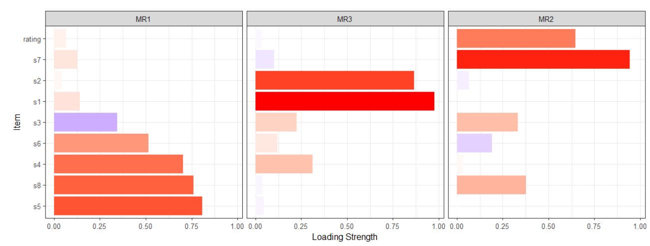

Figure 1: Factor loadings obtained from the Serendipity 2018 dataset. Red bars indicate positive loading and blue bars indicate negative loading. The MR1 factor clusters the questionnaire items that are related to surprise, MR2 clusters the questions related to relevance, and MR3 captures the impact that the Movielens systems exerted on the user experience.

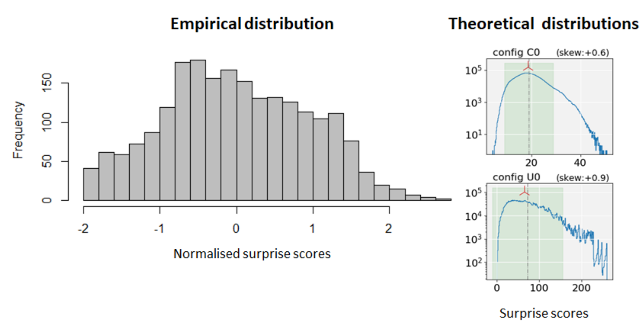

Figure 2: Empirical and theoretical surprise distributions. The latter were obtained from the Movielens 1M dataset [3].

[2]In structural equation modelling, a formative model describes a latent factor (i.e., a hypothetical construct that is represented by a non-observable variable) that is caused by observable variables (i.e., variables that can be directly measured). For example, socioeconomic status (SES) is a construct in economic and social sciences that is described by means of a formative model that combines household income, level of education, and occupational prestige, which are variables that can be measured (usually by means of a questionnaire) [2]. This is in contrast with reflective models, in which the latent factor is postulated as being the cause of the observable variables. For example, intelligence, as defined in the Stanford-Binet intelligence test (SB5), is a construct in psychology that is described by means of a reflective model that combines factors that measure fluid reasoning, quantitative processing ability, and visual-spatial processing ability, among others [5]. These factors are posited as being the manifestations (or indicators) of intelligence. The choice of a model (as formative or reflective) has non-negligible effects on the results of a factor analysis [19, 20].

[3]In other words, the number of responses in the Serendipity 2018 dataset ($N = 2150$) meets the requirements for factor analysis according to criteria specified in [7].

ACKNOWLEDGMENTS

The first author acknowledges the financial support provided by the CAPES research agency.

REFERENCES

[1] Ethem Alpaydin. 2014. *Introduction to Machine Learning* (3rd ed.). MIT Press, Cambridge, MA, USA.

[2] Kenneth Bollen and Richard Lennox. 1991. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* 110, 2 (1991), 305. https://doi.org/10.1037/0033-2909.110.2.305

[3] Andre Paulino de Lima and Sarajane Marques Peres. 2019. Effect of item representation and item comparison models on metrics for surprise in recommender systems. In *Proceedings of the 21st International Conference on Enterprise Information Systems*, Vol. 1. SciTePress, 513–524. https://doi.org/10.5220/0007677005130524

[4] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics using R.* Sage Publications, London, UK.

[5] R Michael Furr and Verne R Bacharach. 2014. *Psychometrics: An Introduction* (2nd ed.). Sage Publications, Thousand Oaks, California, USA.

[6] Richard L. Gorsuch. 1974. *Factor Analysis.* Saunders, Philadelphia, PA, USA.

[7] Janke K. Holt. 2014. *A comparison between factor analysis and item response theory modeling in scale analysis.* Ph.D. Dissertation. University of Groningen.

[8] Matt C. Howard. 2016. A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction* 32, 1 (2016), 51–62. https://doi.org/10.1080/10447318.2015.1087664

[9] Cheryl B. Jarvis, Scott B. MacKenzie, and Philip M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30, 2 (2003), 199–218. https://doi.org/10.1086/376806

[10] Marius Kaminskas and Derek Bridge. 2014. Measuring surprise in recommender systems. In *Proceedings of the Workshop on Recommender Systems Evaluation: Dimensions and Design (REDD 2014) at the 8th ACM Conference on Recommender Systems (RecSys'14).* https://www.insight-centre.org/content/measuring-surprise-recommender-systems-0

[11] Marius Kaminskas and Derek Bridge. 2016. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (Dec. 2016). https://doi.org/10.1145/2926720

[12] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating recommender systems with user experimentation. In *Recommender Systems Handbook* (2nd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, USA, Chapter 9, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9

[13] Denis Kotkov, Joseph A. Konstan, Qian Zhao, and Jari Veijalainen. 2018. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18).* ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/3167132.3167276

[14] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192. https://doi.org/10.1016/j.knosys.2016.08.014

[15] James R. Lewis. 2014. Usability: Lessons learned and yet to be learned. *International Journal of Human-Computer Interaction* 30, 9 (2014), 663–684. https://doi.org/10.1080/10447318.2014.930311

[16] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103. https://doi.org/10.1007/978-3-642-02806-9_12

[17] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14).* ACM, New York, NY, USA, 677–686. https://doi.org/10.1145/2566486.2568012

[18] Tien T. Nguyen, Daniel Kluver, Ting-Yu Wang, Pik-Mai Hui, Michael D. Ekstrand, Martijn C. Willemsen, and John Riedl. 2013. Rating support interfaces to improve user experience and recommender accuracy. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13).* ACM, New York, NY, USA, 149–156. https://doi.org/10.1145/2507157.2507188

[19] Stacie Petter, Detmar Straub, and Arun Rai. 2007. Specifying formative constructs in information systems research. *MIS Quaterly* 31, 4 (Dec. 2007), 623–656. http://dl.acm.org/citation.cfm?id=2017356.2017358

[20] Nicholas Roberts and Jason Thatcher. 2009. Conceptualizing and testing formative constructs: Tutorial and annotated example. *SIGMIS Database* 40, 3 (July 2009), 9–39. https://doi.org/10.1145/1592401.1592405

[21] Alan Said and Alejandro Bellogín. 2018. Coherence and inconsistencies in rating behavior: Estimating the magic barrier of recommender systems. *User Modeling and User-Adapted Interaction* 28, 2 (01 Jun 2018), 97–125. https://doi.org/10.1007/s11257-018-9202-0

[22] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook* (2nd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, Chapter 10, 353–382. https://doi.org/10.1007/978-1-4899-7637-6_10

[23] Noam Tractinsky. 2018. The usability construct: A dead end? *Human-Computer Interaction* 33, 2 (2018), 131–177. https://doi.org/10.1080/07370024.2017.1298038