Explaining Reviews and Ratings with PACO: Poisson Additive Co-Clustering

Chao-Yuan Wu*
Department of Computer Science
University of Texas at Austin, Austin, TX
cywu@cs.utexas.edu

Alex Beutel*
Computer Science Department
Carnegie Mellon University, Pittsburgh, PA
abeutel@cs.cmu.edu

Amr Ahmed
Strategic Technologies
Google, Mountain View, CA
amra@google.com

Alexander J. Smola
Machine Learning Department
Carnegie Mellon University, Pittsburgh, PA
alex@smola.org

ABSTRACT

Understanding a user's *motivations* provides valuable information beyond the ability to recommend items. Quite often this can be accomplished by perusing both ratings and review texts. Unfortunately matrix factorization approaches to recommendation result in large, complex models that are difficult to interpret. In this paper, we attack this problem through succinct additive co-clustering on both ratings and reviews. Our model yields accurate and interpretable recommendations.

Keywords

Co-clustering; recommendation systems; joint modeling

1. INTRODUCTION

Recommender systems often aim to generate suggestions while simultaneously *explaining* why a certain recommendation was made. We address this problem by extending ACCAMS [1] to include a novel *additive* language description in the form of a sum of *Poisson* distributions. This allows us to use backfitting for documents rather than just in a regression setting, and enables new applications.

With this approach we make a number of contributions:

- We design an additive co-clustering model, PACO, that can sum over both Gaussian and Poisson distributions. PACO jointly learns a model of reviews and ratings, giving the ability to interpret our model.
- We describe an efficient technique for sampling from a sum of Gaussian and Poisson random variables.
- We give empirical evidence across multiple datasets that PACO predicts ratings better than HFT [3] and JMARS [2]. Additionally, our method predicts reviews better than HFT, and achieves nearly as high quality prediction as JMARS, while being far faster and simpler. As seen in Figure 1, PACO outperforms both models in jointly predicting ratings and reviews.

*These authors contributed equally.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889400.

2. ADDITIVE CO-CLUSTERING MODEL

In this section we give a high-level description on how to extend ACCAMS to jointly model ratings and reviews. See [1] for background and the complete paper [4] for details. We consider modeling a set of observed entries (u, m), each of which is comprised of a rating and a review that user u gives to item m. In the generative model of ACCAMS, each block in a co-clustering generates a Gaussian-distributed rating, a sum of which across co-clusterings gives the final rating. In PACO, each block further emits a Poisson-distributed word count, a sum of which across co-clusterings gives the final count $n_{u,m,x}$ of word x for review (u, m), i.e.

$$n_{u,m,x} \sim \text{Poi}\left(\lambda_{u,m,x}\right) \text{ and } \mu_x^{(*)} \sim \text{Gamma}(\alpha, \beta) \qquad (1)$$

where

$$\lambda_{u,m} = \mu^{(0)} + \mu^{(m)} + \left[\sum_{\ell=1}^{S} \mu^{(\ell)}_{\mathbf{c}_u^{(\ell)}, \mathbf{d}_m^{(\ell)}} + \mu^{(u,\ell)}_{\mathbf{c}_u^{(\ell)}} + \mu^{(m,\ell)}_{\mathbf{d}_m^{(\ell)}}\right] \qquad (2)$$

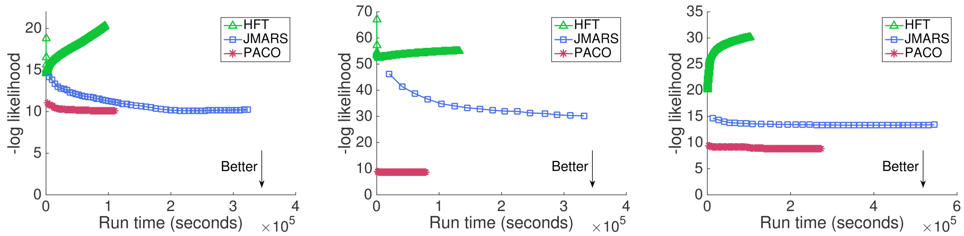
and $\mathbf{c}_u$ and $\mathbf{d}_m$ are cluster assignments for (u, m). Here we further extend each co-clustering to have a user-clustering-specific language model $\mu^{(u,\ell)}_{\mathbf{c}_u^{(\ell)}}$ and an item-clustering-specific language model $\mu^{(m,\ell)}_{\mathbf{d}_m^{(\ell)}}$. In addition, we add a global item language model, $\mu^{(m)}$, and a global background language model, $\mu^{(0)}$. The text of the review is modeled as a combination of these Poisson language models.

2.1 Inference

We offer an efficient Gibbs sampling procedure to learn the PACO model. The collapsed Gibbs sampler for Gaussian distributions is described in [1]. In our complete paper [4], we give the precise equations for sampling a given $\mu^{(\ell)}_{a,b,x}$ conditioned on all other language models in PACO. The key idea is a novel algorithm that parametrizes sampling from the sum of Poisson distributions as an efficient sampling from a multinomial distribution.

$$\{\hat{n}_{u,m,x}\} \sim \text{Multi}\left(\frac{\left\{\mu^{(*)}_{u,m,x}\right\}}{\lambda_{u,m,x}}, n_{u,m,x}\right). \qquad (3)$$

This allows the model to attribute a given word of the review to each language model (user, movie or cluster).



(a) Amazon ne foods (b) RateBeer (c) Yelp

	RMSE	Perplexity
HFT	2.2328	8.1747
JMARS	2.2947	7.4610
PACO	2.1877	7.5540

(d) IMDb

Figure 1: Negative log likelihood. PACO better jointly predicts ratings and reviews than state-of-the-art JMARS [2] and HFT [3] on Amazon Fine Food, Yelp and RateBeer datasets. The joint predictive power is capture by the normalized negative log likelihood. Lower is better. (d) shows detailed results on IMDb dataset. More comparisons are given in [4].

Subset of items in cluster	Cluster words
Entrapment, Mission: Impossible III, Zombie, Snake Eyes, Starsky & Hutch, New England Patriots vs. Minnesota Vikings, I Am Legend, Chaos	action, good, character, thought, story, plot, scene, expected, average, movies, game, scenes, lack, massive, destruction, entertained, suspenseful, audience, seats, batman
Gargantua, Random Hearts, Chocolate: Deep Dark Secrets, Blackout, The Ventures of Marguerite, Irresistible, Ghosts of Girlfriends Past, Youth Without Youth	like, good, bad, time, movies, people, acting, plot, watch, horror, watching, worst, scenes, pretty, awful, effects, scene, characters, thought, story, actors, worse, films, terrible, special, lot, fun

Table 1: Discovered clusters of items and associated topics for IMDb.

3. EXPERIMENTS

Dataset	Yelp	Food	RateBeer	IMDb
# items	60,785	74,257	110,369	117,240
# users	366,715	256,055	29,265	452,627
# observations	1,569,264	568,447	2,924,163	1,462,124
# unigrams	9,055	9,088	8,962	9,182
avg. review length	45.20	31.55	28.57	88.30

Table 2: Datasets used in experiments.

To extensively test our model, we select four datasets about movies, beer, businesses, and food. All four datasets come from di erent websites and communities, thus capturing di erent styles and patterns of online ratings and reviews. We evaluate performance of rating prediction based on RMSE, and review text prediction based on perplexity. An overview of our results can be seen in Figure 1, and detailed results for IMDb is shown in Table 1(d). Complete results for all four datasets are provided in the complete paper. We see PACO outperforms HFT and JMARS in rating prediction and achieves nearly as high quality review prediction as JMARS, while being far faster and simpler.

In addition to quantitatively evaluating our method, we also want to empirically demonstrate that the patterns surfaced would be useful to the human eye. We see PACO is able to nd meaningful item clusters (Table 1), learn item-speci c words (Table 4), and predict words matching the sentiment of the predicted rating (Table 3).

4. CONCLUSION

We presented PACO, an additive novel Poisson co-clustering algorithm for explainable recommendations that is fast, succinct, interpretable and showed competitive results with state-of-the-art joint models.

Acknowledgements: This research was supported by funds from Google, a Facebook Fellowship, and the National Science Foundation under Grant No. DGE-1252522, CNS-1314632 and IIS-1408924. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties.

Rating	Words
0.022	great, love, movies, story, life, watch, time, people, character, characters, best, films, scene, watching, real, world, acting
-0.018	bad, good, plot, like, worst, money, waste, acting, script, movies, minutes, horrible, boring, thought, stupid, people

Table 3: Blocks predict words matching the sentiment of the predicted rating.

Item	Item-specific words
The Dark Knight	batman, joker, dark, ledger, knight, heath, nolan, best, performance, bale, action, dent
Silent Hill	game, silent, hill, games, horror, video, rose, town, like, played, plot, scary, story, monsters

Table 4: Item-specific words capture concepts highly specific to the individual item.

References

[1] A. Beutel, A. Ahmed, and A. J. Smola. ACCAMS: Additive Co-Clustering to Approximate Matrices Succinctly. In *World Wide Web*, pages 119{129, 2015.

[2] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). In *Knowledge Discovery and Data Mining*, pages 193{202. ACM, 2014.

[3] J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *ACM Recommender Systems*, pages 165{172. ACM, 2013.

[4] C.Y. Wu, A. Beutel, A. Ahmed, and A. J. Smola. Explaining reviews and ratings with PACO Poisson Additive Co-Clustering In *http://arxiv.org/abs/1512.01845*