Yelp Recommendation System Based on Collaborative Filtering

Sainan He, Jiaoyang Fu, Yameng Lio

Electrical and Computer Engineering, University of Waterloo, Canada,
`s66he, j45fu, y949li@uwaterloo.cau`

Abstract. Based on Yelp Data Challenge dataset, we aim to develop a predictive personalized recommendation system on users review star rating for restaurants, applying collaborative filtering algorithms. In particular, we implement and compare the performances of four algorithms including baseline, User-based and Item-based collaborative filtering and Singular Vaule Deccomposition (SVD). We evaluate our results by comparing our predicted rating to the actual rating using Root Mean Squared Error(RMSE) and Mean Absolute Error(MAE) metrics

Keywords: Recommendation System, Collaborative filtering, Singular Vaule Deccomposition (SVD), Yelp Data Challenge

1 Introduction

With rapid development of advanced technology, people nowadays can achieve the things that they desired faster and more effectively than ever. While, at the same time, the requirements for accurate, personalized and convenient services are increasing. Fortunately, Yelp contributes to providing reasonable recommendations of various businesses to users, e.g., restaurants.

Yelp collected review dataset which records how well each user rates for limited amount of restaurants. Based on these review history, it suggests favored restaurants to users. The problem is that Yelp seems to offer similar recommendations that are popular for various users. In fact, people may have diverse preferences to food. For instance, some users may prefer Asian food, while others may favor Mexican food. In fact, yelp did not consider these factors too much. Therefore, it seems still have some room to improve the recommendation system in the personalization aspect.

Collaborative filtering, a method making predictions based on a large dataset used by some recommendation system like Yelp, Amazon, and Netflix. It automatically predict about the interests of a user by collecting preference or tastes information from other users. Furthermore, there are numerous collarborative filtering methods, such as baseline, K- nearest neighbor, and matrix decomposition. This paper aims to compare how accurate each methods applying on Yelp Data Challenge through using Root metric Mean Squared Error(RMSE) and Mean Aboslute Error(MAE).

2 Literature Review

There are generaly three types of recommendation systems: content-based filtering, collaborative filtering and hybrid approaches. Content-based recommendation systems work with users profile and items characteristics. The feature used to bulding profiles are often a set of keywords. For example, a music recommendation system[1] implemented with content-based filtering, each song is assigned an attribute manually. If a users profile shows interests in songs with particular atrributes, similar songs will be recommended to the user. The limitations of these systems is that it always recommends similar items to user that he has already purchased and its difficult to recommend items for new users.

Collaborative filtering try to predict the utility of items for a particular user based on the items previously rated by other users with similar tastes and preferences. In[2] the author builds a recommendation system for a retail store using three kinds of collaborative filtering algorithms - memory based approach, matrix factorization and bigram matrix method. Collaborative filtering also has limitations that it is difficult to recommend items for new users, to recommend items which have not been rated before, and to recommend when rating information is insufficient.

Hybrid approach combines multiple techniques to overcome the limitations of individual systems. A restaurant recommendation system for yelp user[3] adopts hybrid approach by extracting collaborative and content-based features to identify customer and restaurant profiles. A hybrid cascade of K-nearest neighbor clustering, weighted bi-partite graph projection, and several other learning algorithms are proposed.

3 Data

We collect our data from Yelp recommendation Kaggle competition[4],This dataset contains 11,537 businesses, 8282 check-in sets, 43873 users, and 229907 reviews. We target Restaurant in the city of Phoenix as it is more reasonable to recommend restaurants for users in the same city.

3.1 Data Processing

The original data file is in json format. We firstly parse the raw data of imformation from users, businesses and reviews, respectively and merge them into one dataframe. Then we extract the records with features of Restaurants and Phoenix for futher analysis. After this approach, we have 17145 users, 1454 business and 52749 reviews.

Because many users only review a few restaurants, our dataset yields a sparsity of 99.79% which can be visualized from figure 1. To solve this problem, we create a smaller dataset which only consider restaurants reviewed by more than 50 users.

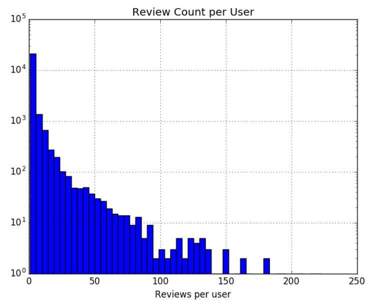

Fig. 1. User review count

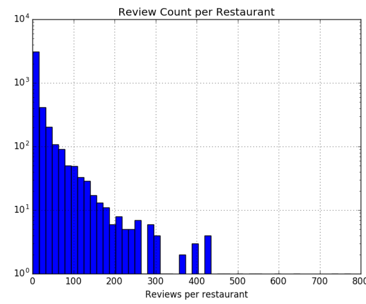

Fig. 2. Restaurants review count

3.2 Traing and Testing sets

In order to evaluate the performance of our recommendation system, we divided our data into training and testing datasets. We extract a list of all restaurants each user has rated, take 80% of this list as training set and 20% as test set.

4 Methods

4.1 Baseline

Our baseline model is similar to the model implemented by[5] which is a mean predictor and accounts for the user and item effects.

$$b_{ur} = \mu + b_u + b_r \ . \tag{1}$$

Here μ is the mean rating of reviews of all business by all users. The parameter b_u indicates the difference between the average rating of user u and μ. The parameter b_i indicates the difference between the average rating of business i and μ. This will normalize the widely noticed tendency of some user giving higher rating than others and some restaurants getting higher ratings than others.

4.2 User-based Collaborative Filtering

4.3 User-based Collaborative Filtering

4.4 Singular Vaule Deccomposition

5 Experiments and Results

5.1 Evaluation Metrics

Our goal is to predict the rating a user would give to a restaurant. We predict the rating that user has not rated in the training dataset, but the true rating is stored in the test dataset. We use the root-mean-square error and mean-absolute error for evaluation.

$$RMSE = \sqrt{\frac{\sum \left(r'_{u,i} - r_{u,i}\right)^2}{N}} \tag{2}$$

$$MAE = \sqrt{\frac{\sum \left|r'_{u,i} - r_{u,i}\right|}{N}} \tag{3}$$

Here $r_{u,i}$ is the predicted rating from user u on item i and $r'_{u,i}$ is the true rating; N is the size of test dataset.

6 Results

The performance of our baseline predictor is $RMSE =$, $MAE =$.

References

1. MJ Pazzani, D Billsus.: Content-based Recommendation Systems. In The adaptive web, pp. 325341. Springer (1981)
2. Bruno Pradel, Savaneary Sean, Julien Delporte.: A Case Study in a Recommender System Based on Purchase Data. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 377-385 (2011)
3. Sumedh Sawant, Gina Pai.: Yelp Food Recommendation System
4. https://www.kaggle.com/c/yelp-recsys-2013
5. Yehuda Koren.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '08), pp. 426-434 (2008)

Hamiltonian Mechanics2

Ivar Ekeland[1] and Roger Temam[2]

[1] Princeton University, Princeton NJ 08544, USA
[2] Université de Paris-Sud, Laboratoire d'Analyse Numérique, Bâtiment 425, F-91405 Orsay Cedex, France

Abstract. The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. . . .

Keywords: graph transformations, convex geometry, lattice computations, convex polygons, triangulations, discrete geometry

1 Fixed-Period Problems: The Sublinear Case

With this chapter, the preliminaries are over, and we begin the search for periodic solutions to Hamiltonian systems. All this will be done in the convex case; that is, we shall study the boundary-value problem

$$\begin{aligned} \dot{x} &= JH'(t,x) \\ x(0) &= x(T) \end{aligned}$$

with $H(t,\cdot)$ a convex function of x, going to $+\infty$ when $\|x\| \to \infty$.

1.1 Autonomous Systems

In this section, we will consider the case when the Hamiltonian $H(x)$ is autonomous. For the sake of simplicity, we shall also assume that it is C^1.

We shall first consider the question of nontriviality, within the general framework of (A_∞, B_∞)-subquadratic Hamiltonians. In the second subsection, we shall look into the special case when H is $(0, b_\infty)$-subquadratic, and we shall try to derive additional information.

The General Case: Nontriviality. We assume that H is (A_∞, B_∞)-subquadratic at infinity, for some constant symmetric matrices A_∞ and B_∞, with $B_\infty - A_\infty$ positive definite. Set:

$$\gamma := \text{smallest eigenvalue of} \quad B_\infty - A_\infty \tag{1}$$

$$\lambda := \text{largest negative eigenvalue of} \quad J\frac{d}{dt} + A_\infty \ . \tag{2}$$

Theorem 21 tells us that if $\lambda + \gamma < 0$, the boundary-value problem:

$$\begin{aligned} \dot{x} &= JH'(x) \\ x(0) &= x(T) \end{aligned} \tag{3}$$

has at least one solution $\overline{x}$, which is found by minimizing the dual action functional:

$$\psi(u) = \int_o^T \left[\frac{1}{2}\left(\Lambda_o^{-1} u, u\right) + N^*(-u)\right] dt \tag{4}$$

on the range of Λ, which is a subspace $R(\Lambda)_L^2$ with finite codimension. Here

$$N(x) := H(x) - \frac{1}{2}\left(A_\infty x, x\right) \tag{5}$$

is a convex function, and

$$N(x) \le \frac{1}{2}\left(\left(B_\infty - A_\infty\right) x, x\right) + c \quad \forall x \ . \tag{6}$$

Proposition 1. *Assume $H'(0) = 0$ and $H(0) = 0$. Set:*

$$\delta := \liminf_{x \to 0} 2N(x) \left\|x\right\|^{-2} \ . \tag{7}$$

If $\gamma < -\lambda < \delta$, the solution $\overline{u}$ is non-zero:

$$\overline{x}(t) \ne 0 \quad \forall t \ . \tag{8}$$

Proof. Condition (7) means that, for every $\delta' > \delta$, there is some $\varepsilon > 0$ such that

$$\|x\| \le \varepsilon \Rightarrow N(x) \le \frac{\delta'}{2} \left\|x\right\|^2 \ . \tag{9}$$

It is an exercise in convex analysis, into which we shall not go, to show that this implies that there is an $\eta > 0$ such that

$$f \left\|x\right\| \le \eta \Rightarrow N^*(y) \le \frac{1}{2\delta'} \left\|y\right\|^2 \ . \tag{10}$$

Fig. 1. This is the caption of the figure displaying a white eagle and a white horse on a snow field

Since u_1 is a smooth function, we will have $\|hu_1\|_\infty \leq \eta$ for h small enough, and inequality (10) will hold, yielding thereby:

$$\psi(hu_1) \leq \frac{h^2}{2}\frac{1}{\lambda}\left\|u_1\right\|_2^2 + \frac{h^2}{2}\frac{1}{\delta'}\left\|u_1\right\|^2 \ . \tag{11}$$

If we choose δ' close enough to δ, the quantity $\left(\frac{1}{\lambda}+\frac{1}{\delta'}\right)$ will be negative, and we end up with

$$\psi(hu_1) < 0 \qquad \text{for} \ \ h \neq 0 \ \ \text{small} \ . \tag{12}$$

On the other hand, we check directly that $\psi(0) = 0$. This shows that 0 cannot be a minimizer of ψ, not even a local one. So $\overline{u} \neq 0$ and $\overline{u} \neq \Lambda_o^{-1}(0) = 0$. □

Corollary 1. *Assume H is C^2 and (a_∞, b_∞)-subquadratic at infinity. Let ξ_1, $\ldots, \xi_N$ be the equilibria, that is, the solutions of $H'(\xi) = 0$. Denote by ω_k the smallest eigenvalue of $H''(\xi_k)$, and set:*

$$\omega := \text{Min}\left\{\omega_1, \ldots, \omega_k\right\} \ . \tag{13}$$

If:

$$\frac{T}{2\pi}b_\infty < -E\left[-\frac{T}{2\pi}a_\infty\right] < \frac{T}{2\pi}\omega \tag{14}$$

then minimization of ψ yields a non-constant T-periodic solution $\overline{x}$.

We recall once more that by the integer part $E[\alpha]$ of $\alpha \in \mathbb{R}$, we mean the $a \in \mathbb{Z}$ such that $a < \alpha \leq a + 1$. For instance, if we take $a_\infty = 0$, Corollary 2 tells us that $\overline{x}$ exists and is non-constant provided that:

$$\frac{T}{2\pi}b_\infty < 1 < \frac{T}{2\pi} \tag{15}$$

or

$$T \in \left(\frac{2\pi}{\omega}, \frac{2\pi}{b_\infty}\right) \ . \tag{16}$$

Proof. The spectrum of Λ is $\frac{2\pi}{T}\mathbb{Z} + a_\infty$. The largest negative eigenvalue λ is given by $\frac{2\pi}{T}k_o + a_\infty$, where

$$\frac{2\pi}{T}k_o + a_\infty < 0 \leq \frac{2\pi}{T}(k_o + 1) + a_\infty \ . \tag{17}$$

Hence:

$$k_o = E\left[-\frac{T}{2\pi}a_\infty\right] \ . \tag{18}$$

The condition $\gamma < -\lambda < \delta$ now becomes:

$$b_\infty - a_\infty < -\frac{2\pi}{T}k_o - a_\infty < \omega - a_\infty \tag{19}$$

which is precisely condition (14). □

Lemma 1. *Assume that H is C^2 on $\mathbb{R}^{2n}\setminus\{0\}$ and that $H''(x)$ is non-degenerate for any $x \neq 0$. Then any local minimizer $\widetilde{x}$ of ψ has minimal period T.*

Proof. We know that $\widetilde{x}$, or $\widetilde{x}+\xi$ for some constant $\xi \in \mathbb{R}^{2n}$, is a T-periodic solution of the Hamiltonian system:

$$\dot{x} = JH'(x) \ . \tag{20}$$

There is no loss of generality in taking $\xi = 0$. So $\psi(x) \geq \psi(\widetilde{x})$ for all $\widetilde{x}$ in some neighbourhood of x in $W^{1,2}\left(\mathbb{R}/T\mathbb{Z};\mathbb{R}^{2n}\right)$.

But this index is precisely the index $i_T(\widetilde{x})$ of the T-periodic solution $\widetilde{x}$ over the interval $(0,T)$, as defined in Sect. 2.6. So

$$i_T(\widetilde{x}) = 0 \ . \tag{21}$$

Now if $\widetilde{x}$ has a lower period, T/k say, we would have, by Corollary 31:

$$i_T(\widetilde{x}) = i_{kT/k}(\widetilde{x}) \geq k i_{T/k}(\widetilde{x}) + k - 1 \geq k - 1 \geq 1 \ . \tag{22}$$

This would contradict (21), and thus cannot happen. □

Notes and Comments. The results in this section are a refined version of 1980; the minimality result of Proposition 14 was the first of its kind.

To understand the nontriviality conditions, such as the one in formula (16), one may think of a one-parameter family x_T, $T \in \left(2\pi\omega^{-1}, 2\pi b_\infty^{-1}\right)$ of periodic solutions, $x_T(0) = x_T(T)$, with x_T going away to infinity when $T \to 2\pi\omega^{-1}$, which is the period of the linearized system at 0.

Table 1. This is the example table taken out of *The TEXbook,* p. 246

Year	World population
8000 B.C.	5,000,000
50 A.D.	200,000,000
1650 A.D.	500,000,000
1945 A.D.	2,300,000,000
1980 A.D.	4,400,000,000

Theorem 1 (Ghoussoub-Preiss). *Assume $H(t,x)$ is $(0,\varepsilon)$-subquadratic at infinity for all $\varepsilon > 0$, and T-periodic in t*

$$H(t,\cdot) \qquad \text{is convex} \quad \forall t \tag{23}$$

$$H(\cdot,x) \qquad \text{is} \quad T\text{−periodic} \quad \forall x \tag{24}$$

$$H(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to \infty \quad \text{as} \quad s \to \infty \tag{25}$$

$$\forall \varepsilon > 0 \ , \quad \exists c \ : \ H(t,x) \leq \frac{\varepsilon}{2} \left\| x \right\|^2 + c \ . \tag{26}$$

Assume also that H is C^2, and $H''(t,x)$ is positive definite everywhere. Then there is a sequence x_k, $k \in \mathbb{N}$, of kT-periodic solutions of the system

$$\dot{x} = JH'(t,x) \tag{27}$$

such that, for every $k \in \mathbb{N}$, there is some $p_o \in \mathbb{N}$ with:

$$p \geq p_o \Rightarrow x_{pk} \neq x_k \ . \tag{28}$$

□

*Example 1 (*External forcing*).* Consider the system:

$$\dot{x} = JH'(x) + f(t) \tag{29}$$

where the Hamiltonian H is $(0, b_\infty)$-subquadratic, and the forcing term is a distribution on the circle:

$$f = \frac{d}{dt} F + f_o \qquad \text{with} \ \ F \in L^2 \left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n} \right) \ , \tag{30}$$

where $f_o := T^{-1} \int_o^T f(t)dt$. For instance,

$$f(t) = \sum_{k \in \mathbb{N}} \delta_k \xi \ , \tag{31}$$

where δ_k is the Dirac mass at $t = k$ and $\xi \in \mathbb{R}^{2n}$ is a constant, fits the prescription. This means that the system $\dot{x} = JH'(x)$ is being excited by a series of identical shocks at interval T.

Definition 1. *Let $A_\infty(t)$ and $B_\infty(t)$ be symmetric operators in $\mathbb{R}^{2n}$, depending continuously on $t \in [0,T]$, such that $A_\infty(t) \leq B_\infty(t)$ for all t.*

A Borelian function $H : [0,T] \times \mathbb{R}^{2n} \to \mathbb{R}$ is called (A_∞, B_∞)-subquadratic at infinity if there exists a function $N(t,x)$ such that:

$$H(t,x) = \frac{1}{2} \left(A_\infty(t)x, x \right) + N(t,x) \tag{32}$$

$$\forall t \ , \quad N(t,x) \qquad \text{is convex with respect to} \ \ x \tag{33}$$

$$N(t,x) \geq n\left(\|x\| \right) \qquad \text{with} \ \ n(s)s^{-1} \to +\infty \ \ \text{as} \ \ s \to +\infty \tag{34}$$

$$\exists c \in \mathbb{R} \ : \quad H(t,x) \leq \frac{1}{2} \left(B_\infty(t)x, x \right) + c \quad \forall x \ . \tag{35}$$

If $A_\infty(t) = a_\infty I$ and $B_\infty(t) = b_\infty I$, with $a_\infty \leq b_\infty \in \mathbb{R}$, we shall say that H is (a_∞, b_∞)-subquadratic at infinity. As an example, the function $\|x\|^\alpha$, with $1 \leq \alpha < 2$, is $(0, \varepsilon)$-subquadratic at infinity for every $\varepsilon > 0$. Similarly, the Hamiltonian

$$H(t,x) = \frac{1}{2} k \left\| k \right\|^2 + \left\| x \right\|^\alpha \tag{36}$$

is $(k, k + \varepsilon)$-subquadratic for every $\varepsilon > 0$. Note that, if $k < 0$, it is not convex.

Notes and Comments. The first results on subharmonics were obtained by Rabinowitz in 1985, who showed the existence of infinitely many subharmonics both in the subquadratic and superquadratic case, with suitable growth conditions on H'. Again the duality approach enabled Clarke and Ekeland in 1981 to treat the same problem in the convex-subquadratic case, with growth conditions on H only.

Recently, Michalek and Tarantello (see Michalek, R., Tarantello, G. 1982 and Tarantello, G. 1983) have obtained lower bound on the number of subharmonics of period kT, based on symmetry considerations and on pinching estimates, as in Sect. 5.2 of this article.

References

Clarke, F., Ekeland, I.: Nonlinear oscillations and boundary-value problems for Hamiltonian systems. Arch. Rat. Mech. Anal. 78, 315–333 (1982)

Clarke, F., Ekeland, I.: Solutions périodiques, du période donnée, des équations hamiltoniennes. Note CRAS Paris 287, 1013–1015 (1978)

Michalek, R., Tarantello, G.: Subharmonic solutions with prescribed minimal period for nonautonomous Hamiltonian systems. J. Diff. Eq. 72, 28–55 (1988)

Tarantello, G.: Subharmonic solutions for Hamiltonian systems via a $\mathbb{Z}_p$ pseudoindex theory. Annali di Matematica Pura (to appear)

Rabinowitz, P.: On subharmonic solutions of a Hamiltonian system. Comm. Pure Appl. Math. 33, 609–633 (1980)