

# Gaussian Process Regression for Movie Recommendations

## Group 7

National University of Singapore

### Abstract

In recent years, online streaming and video on-demand services have grown exponentially. The online nature of such services necessarily multiplies the number of products that can be shown to the consumer. Consequently, there is a need for such services to determine a selection which should be recommended. In order to produce accurate predictions, there is a need to consider users' preferences, item profiles and correlations between items and users. In this paper, we attempt to apply Gaussian Process (GP) regression on movie ratings to generate recommendations.

## 1 Introduction

Today, it is easy to obtain online information pertaining to a movie and its critics through a search on Google, IMDB, Rotten Tomatoes, etc. Hence, modern day movie watchers would often research online about a movie before buying tickets in their local cinema or paying for online streams on movie streaming platforms such as Netflix, Amazon Video, HBO, or Hulu.

The rise of on-demand video services requires an accurate movie recommendation system in order to help users purchase the movies they like. As movie-watching shifts into the online world, consumers are faced with more choices as compared to the traditional way of browsing limited movies available in physical dvd rental stores or the theatre. Within the online ecosystem, services like Netflix are faced with the double-edged sword of being able to provide virtually any movie in existence to the consumer. Letting the consumer conduct an exhaustive search on this massive search space is certainly a computational burden that we would like to avoid.

## 2 Motivating Application

Moreover, motivating reasons such as word of the mouth, reviews and advertising are insufficient in the context of digitized movies. To cater to the consumer who faces the paradox of choice, there is a need for quality recommender systems which can enable movie watchers to discover movies they love on their own.

Suppose a Gaussian process model can generate accurate predictive means of user ratings for a particular movie. We can then provide a good selection of recommended movies based on predicted ratings. This selection can be made on the basis of the predictive uncertainty that the Gaussian process model usefully provides. For example, we might obtain similarly high ratings for 2 movies, A and B. If the predictive uncertainty given by the Gaussian process model is such that  $\sigma_A^2 > \sigma_B^2$ , then we are more likely to recommend movie B, or rank it higher on our recommendation list. Alternatively, it is possible to use multiple models and weight our predictions inversely proportional to the variance. As a consequence, users will benefit by deriving greater enjoyment.

With increased user satisfaction and hence higher utilisation of the system, we are also likely to obtain even more data from new ratings. This will allow us to further exploit the correlations among user preferences, as well as the Bayesian nature of Gaussian process models by updating the model with new data.

Moreover, movie preference is highly personal and complex to model. Director, script, cast, score, visual effects and genres are just of some obvious features of a movie. To give a trivial example, an individual might not like mainstream A-listers action movies such as Fast & Furious, Bourne or James Bond, but he may like Marvel action movies because he likes superhero comics. Even then, he might not like the newer Spiderman movies because of the main lead. He might love DC Comic Batman trilogy directed by Christopher Nolan but still be disappointed by Batman v Superman: Dawn of Justice.

It is clear that modelling human behaviour and preference is complex. Using a parametric model could unintentionally constrain our model. Since the GP model is a distribution over an infinite number of functions, its non-parametric attribute allows us to better model the full spectrum of highly variable human preferences.

To tackle this problem, we propose a movie recommender system based on the Gaussian Process (GP) model. Given a list of movies that an individual has not rated before, we can use GP to compute a predictive mean rating that the individual is likely to give and a predictive uncertainty which gives us insight on how accurate our prediction is.

### 3 Technical Approach

A Gaussian process is a collection of random variables, any finite subset of which have a multivariate Gaussian distribution. It is completely specified by a mean function  $\mu(\mathbf{x})$  and the covariance function  $k(\mathbf{x}, \mathbf{x}')$ . For a real process  $f(\mathbf{x})$ :

$$\begin{aligned}\mu(\mathbf{x}) &= E[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= E[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]\end{aligned}$$

The GP can then be denoted as:

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

We assume that our data  $\mathcal{D} = \{(x_1, y_1), \dots, (x_i, y_i)\}$  are such that  $y_i$  are noisy observations originating from a GP-distributed random function  $f(\mathbf{x}_i)$  such that:

$$\begin{aligned}y_i &= f(\mathbf{x}_i) + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, \sigma_n^2)\end{aligned}$$

Given  $\mathbf{y} = [y_1 y_2 \dots y_i]^\top$ , suppose we have a new input  $\mathbf{x}_*$  for which we would like to obtain a prediction for. In other words, we would like to infer  $p(f_* | \mathbf{y})$ . Then, the predictive mean and variance from the GP can be given by:

$$\begin{aligned}E[f_* | \mathbf{y}] &= \mu(\mathbf{x}_*) + \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ V[f_* | \mathbf{y}] &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*\end{aligned}$$

#### Choosing Model Parameters

The GP function is mainly characterized by its covariance function after normalizing the data to attain a mean of 0. Our choice of kernel for both experimental models, per user and per movie, was the radial basis function(RBF) kernel

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2} \sum_{i=1}^d \frac{(x_i - x'_i)^2}{\ell_i^2}\right)$$

where the hyperparameters  $\sigma^2$  is the variance,  $\ell$  is the length-scale,  $d$  is the dimension of the input vector  $x_i$ .

The choice of the rbf kernel was due to its ability to produce similar predictive ratings for two inputs with similar features based on their distance. If two users have contrasting preferences, it is highly unlikely that both will assign the same rating to their opposing's favorite movie.

### 4 Experimental Setup

#### Dataset

Our primary dataset is the MovieLens 1M Dataset which is a stable benchmark dataset widely used for evaluating recommender systems. The primary dataset contains around 1,000,000 anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users who joined MovieLens<sup>1</sup> in 2000. Besides ratings, the following attributes were extracted from this dataset:

<sup>1</sup>MovieLens is a movie recommender system created by GroupLens Research.

**movie\_id** movie identification number

**title** movie title

**genres** a list of genres that the movie belongs to

We also utilised user demographic information such as age, gender and occupation available from the dataset.

To supplement this dataset, we combined it with the MovieLens+IMDb/Rotten Tomatoes Dataset released by HetRec 2011. Our motivation for doing so was to obtain movie critic ratings from Rotten Tomatoes. As the project progressed, we also extracted the relevant directors and actors for feature engineering purposes.

#### Preprocessing

#### Feature Engineering

### 5 Experimental Evaluation

We have discovered truly remarkable results which this margin is too small to contain

### 6 References

Note: max 6 pages