TV Ratings Prediction with Time Weighting Based Regression (TWR)

Master Thesis Defense

Student — Ting-Wei Ku (Martin)

Advisor — Prof. Shou-De Lin

Committee — Prof. Pu-Jen Cheng, Ph.D. Cheng-Te Li

National Taiwan University
Department of Computer Science & Information Engineering
Machine Discovery & Social Network Mining Lab

Feb 2, 2015

Agenda

1. Thesis Goal: Improve TV Ratings Prediction with MY NOVELTY
2. Related Work
3. Solution: Time Weighting Based Regression (TWR)
4. Experiments
5. Conclusion

Why TV ratings prediction?

It is an important, complex, and real-world problem with money.

- It's important because TV ratings decide **price of advertising time**.
- It's complex because...
  - TV ratings are **aggregate** measure of **many people's choices**.
  - TV is **competing** with many platforms/services (mobile/YouTube).

MY NOVELTY (Contribution) is TWR

Key idea of TWR

Fit regression model with time-weighted instances.

- **Example**: Given x is a time series of ratings,
  - (x1, x2, x3, x4=y4), t=4, weight=4
  - (x2, x3, x4, x5=y5), t=5, weight=5
  - (x3, x4, x5, x6=y6), t=6, weight=6
  - . . . more weighted training instances
  - (x6, x7, x8, x9=y9), t=9, testing instance
- **Assumption**: Intuitively, newer instances are more important.

We'll show how **effective** this **simple** solution is via experiments later.

Agenda

TV Ratings Prediction (1/3)

- Forecasting television ratings
  (IJF 2011, Danaher et al.)
  - Compared 8 regression models such as Bayesian model averaging
  - Suggested features such as seasonal factors and program genre
  - Found that modeling ratings directly is better than as total_audience×channel_share
  - Relatively large data: 5,000 programs and 48,000 ratings from 2004-2008
- Using a nested logit model to forecast television ratings
  (IJF 2012, Danaher et al.)
  - Applied nested logit model to TV ratings prediction
  - Same relatively large data

Both works are not compared to ours due to key difference in data.

TV Ratings Prediction (2/3)

- Predicting TV audience rating with social media
  (SocialNLP 2013, Hsieh et al.)
  A predicting model of TV audience rating based on the Facebook
  (SocialCom 2013, Cheng et al.)
  - Introduced Facebook features such as # of likes on the fan page
  - Fit data with neural network
  - 4 weekly dramas (78 ratings) broadcast in TW

Key difference between they and us

We only use historical ratings as features, i.e., no external features at all.

TV Ratings Prediction (3/3)

- A weight-sharing gaussian process model using web-based information for audience rating prediction
  (TAAI 2014, Huang et al.)
  - Proposed a novel GP model
  - Introduced Google Trends features (search-term frequency)
  - 4 daily dramas (336 ratings) broadcast in TW

Key difference between they and us

We only use historical ratings as features, i.e., no external features at all. Besides, we only focus on weekly dramas broadcast in TW.

Agenda

What is TWR?

Why TWR?

How does TWR work?

Agenda

What TV ratings to predict?

- **Data**: 8 real-world weekly dramas (170 ratings) broadcast in TW
  - Originally from SET but now also available at Wikipedia
- Predict next ratings of each drama (1-step forecasting)
- Start making predictions from the 6th episode

Time Series Plot of Data

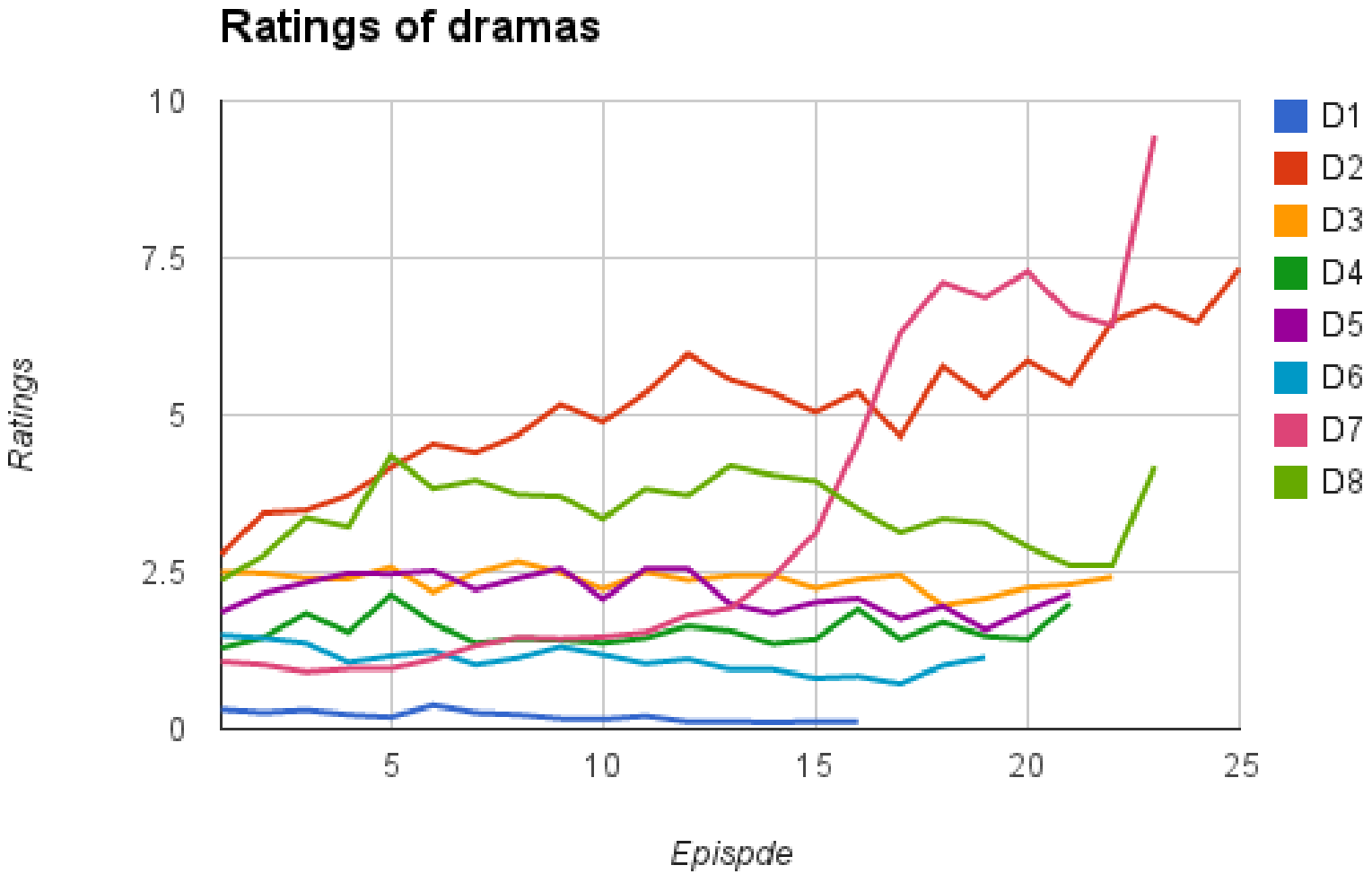

Figure 1: Time series plot of ratings

Box Plot of Data

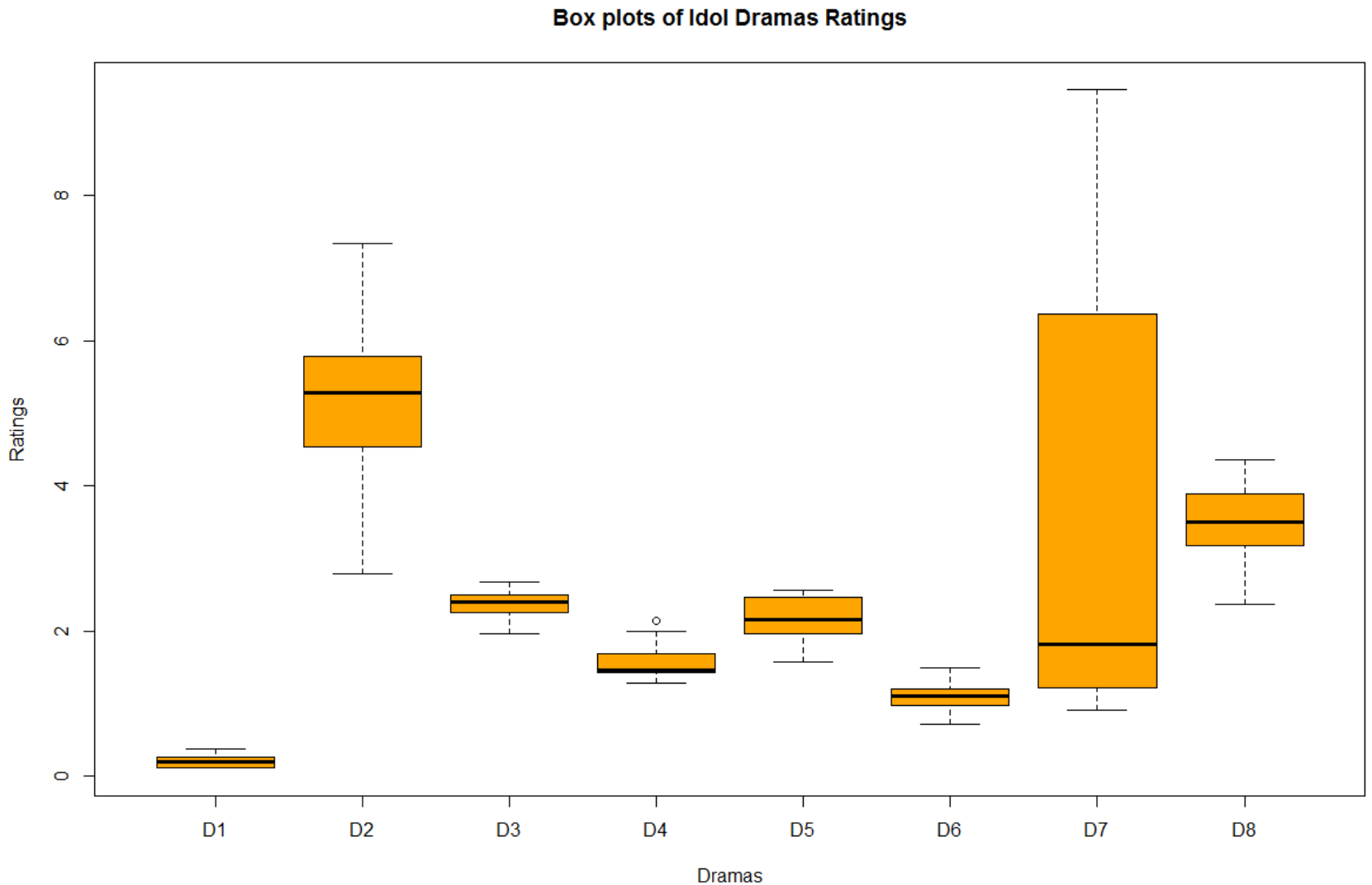

Figure 2: Box plot of ratings

Basic Info of Data

Drama	# Episode	Start	Avg(ratings)	Std(ratings)
D1	16	2013/2/28	0.21	0.08
D2	25	2011/8/21	5.12	1.09
D3	22	2012/2/19	2.38	0.16
D4	21	2013/1/6	1.57	0.23
D5	21	2013/6/9	2.16	0.3
D6	19	2010/12/5	1.1	0.21
D7	23	2010/11/5	3.36	2.75
D8	23	2012/7/22	3.47	0.56

Agenda

Code example

```
if (weight_type == 'equal') {
  # this is known as bagging
  case_weights <- rep(1 / num_cases, num_cases)
} else if (weight_type == 'linear') {
  case_weights <- seq(1, num_cases)
} else if (weight_type == 'exp') {
  case_weights <- exp(1:num_cases)
} else if (weight_type == 'exp3') {
  alpha <- 3
  case_weights <- (exp(1)^alpha)^(1:num_cases)
} else {
  # decide weight type automatically via validation error
}
```

Thank you!