Genre prediction with structured models

Bar Eini , Rom Gutmon, Dafna Schumacher| Machine Learning, Spring 2018 | IE&M, Technion

Introduction

In recent years, the rapid growth of consumable media-content led to increasing interest in customized viewing recommendations. In this project, we focus on viewing patterns. Our goal is to predict specific genres in viewing sequences, where the number of possible genres is large. Through the different sections of our work, we explore various structural connections. To this end, we implement multiple MEMM based models.

Data

The *FourthWall Media* dataset contains viewing records of households' devices across the US, as well as demographic information. In this work, we focused on a specific week and selected devices which are active throughout the entire week.

We define a *view node* as the recoded information about a device while watching a program. Two types' of events are Separating view nodes: switching channels and changing of programs. A view node is valid view node the duration between events is *significant*.

Models

Baselines

1. Most common - A naïve predictor which predicts the most common genre.

2. Multiclass Perceptron – Multiclass perceptron provides a prediction considering demographic and prior knowledge features, but not structured information.

MEMM - Maximum Entropy Markov Model is a discriminative model which provides a structured prediction and estimates $P(y_i|x_1,\dots,x_n,y_{i-2},y_{i-1})$.

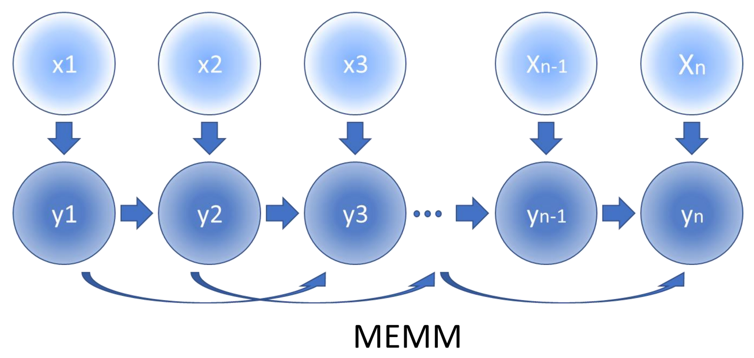
Learning – $loss: L(w) = \sum_{i=1}^{n} \log p(y_i|x_i;w) - \frac{\lambda}{2}||w||^2$

. implemented with and without demographics.

y_i- represent program's genre.

x_i are the features

Inference - Viterbi algorithm with slight modifications: only genres which were previously seen are considered.



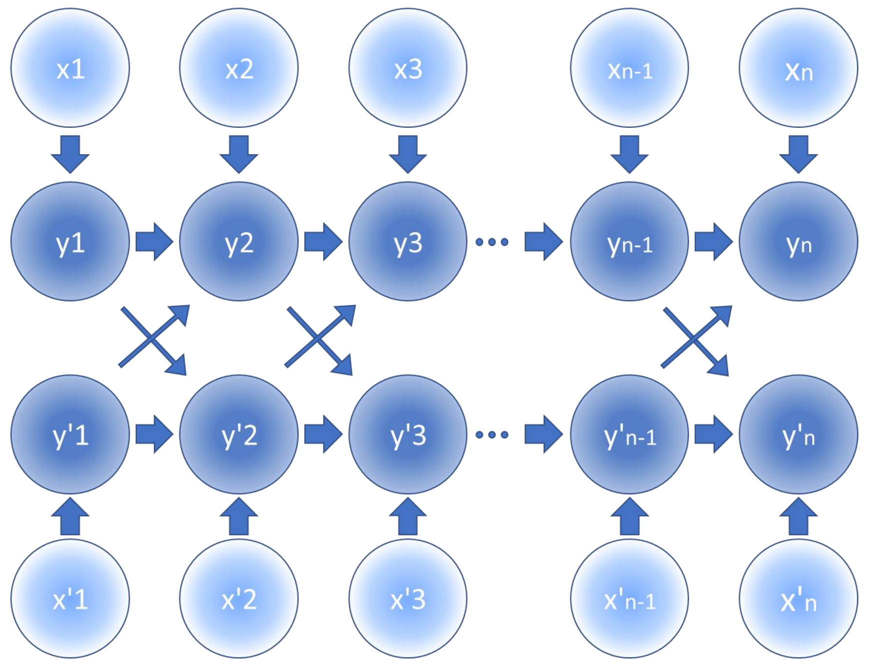
Advanced models – MEMM based models including a paired device.

Device pairing attempts to infer additional information from genre on the paired device y'_{i-1} and calculate $P(y_i|x_i,y_{i-1},y_{i-2},y'_{i-1})$.

2-Step inference –

Step1: Plain Viterbi.

Step2: Embedding predicted genres and additional Viterbi based on y'_{i-1}



Device Paring types:

1. **In-household paring** – Devices within the same household.

Aims to capture household's dynamics.

2. **Preference-based paring** – Devices with similar preferences. Similarity is calculated as following: Θ_i - genre combination , θ – atomic genre

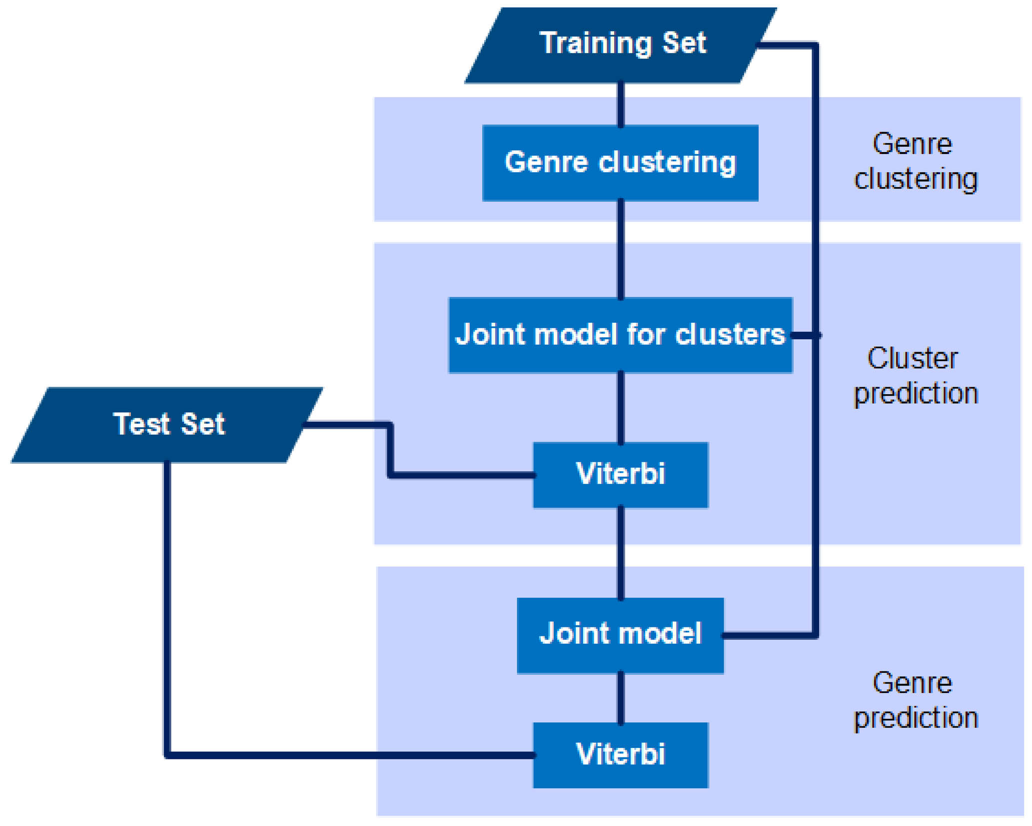
Distance between devices A and B	$\sum_{\Theta_i \in A} \sum_{\Theta_j \in B} \Delta(\Theta_i,\Theta_j) * 1_{(\Theta_{i,j} \notin B \cap A)}$
Genre combinations distance	$\Delta(\Theta_i,\Theta_j) = \frac{\sum_{\{\theta:(\theta \in \Theta_i) \oplus (\theta \in \Theta_j)\}} score(\theta)}{\sum_{\{\theta:(\theta \in \Theta_i) \vee (\theta \in \Theta_j)\}} score(\theta)}$
Genre score is inversed to its frequency	$score(\theta_k) = \frac{1}{freq(\theta_k)}$

3. **Joint model – Both devices**

Cluster based procedure –

Investigating relationships between atomic genres and genre combinations.

Step1: Genre distance metric above for clustering. The clustering procedure is performed with Agglomerative clustering.



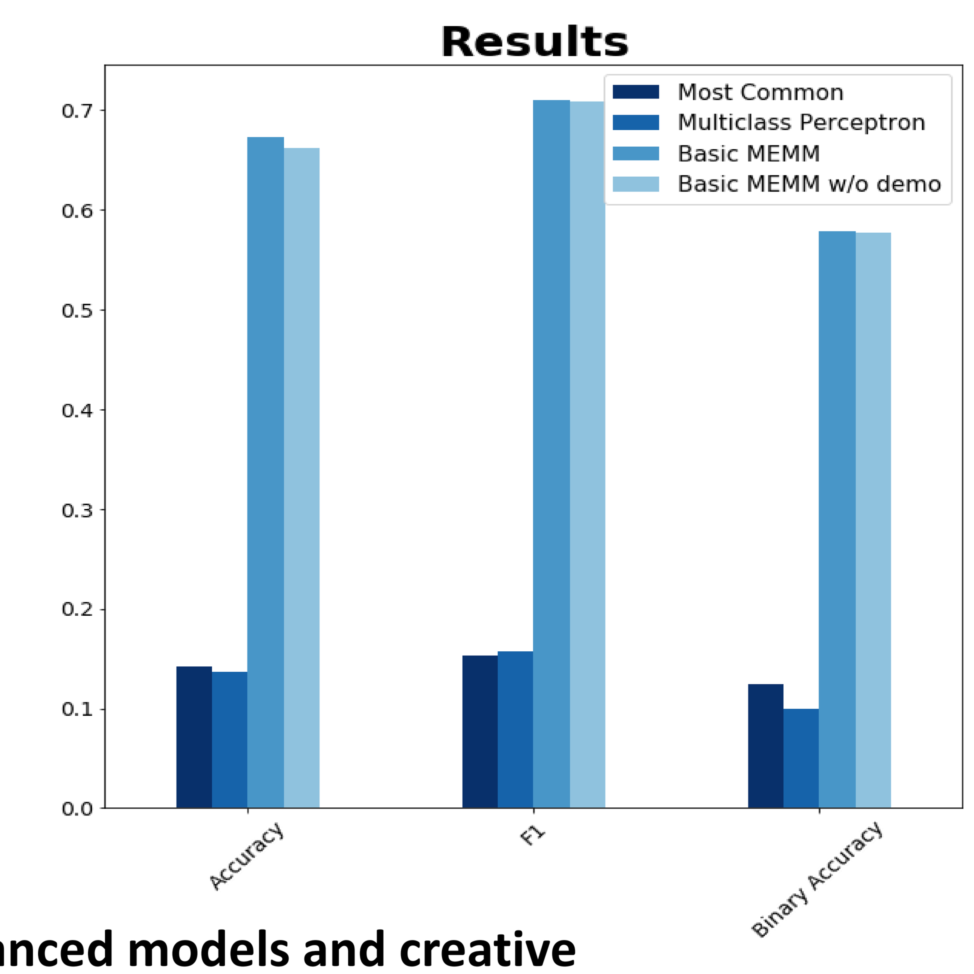
Step2: Predict programs' clusters with a *Joint model*. Inference – Viterbi.

Step3: Predict genres after embedding predicted clusters.

Evaluation & Results

- **Basic Section – Structured VS Unstructured**

Model	Binary Accuracy	Full Accuracy	F1
Most common	0.124	0.142	0.153
MultiClass preceptron	0.099	0.137	0.157
MEMM (w/ demographic features)	**0.578**	**0.665**	**0.71**
MEMM (w/o demographic features)	0.576	0.662	0.708



- **Advanced models and creative**

Model	Binary Accuracy	Full Accuracy	F1
MEMM (basic)	0.578	0.672	0.71
Random device paring	0.58	0.672	0.714
In-household device paring	**0.584**	0.676	0.7243
Preference based device paring	0.582	0.677	0.725
Joint model	**0.584**	**0.679**	**0.727**
Creative model	**0.589**	**0.682**	**0.73**

Conclusion

Overall, our good results of the basic section show that predicting next view nodes' genre can benefit from structured nature of the viewing patterns of households' devices. On the other hand, enhancing the model via device pairing achieves only slight improvements comparing to the basic model.

Future work: The paring of two devices could be improved: In-household paring with the addition of device affiliation and preference based paring with a different similarity metric. We believe that the cluster based procedure could be also benefit from a different similarity metric.