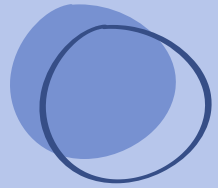


Infinite Possibilities in User Review:

A Recommendation System based on Sentimental Analysis and Neural Collaborative Filtering for STEAM Gamers

120020275 Zhang Jingheng
120020401 Zhang Jiawei
120040081 Lu Yu
120050052 Bian Yingxuan
120090078 Tang Junzhe
120090089 Wang Yuqi

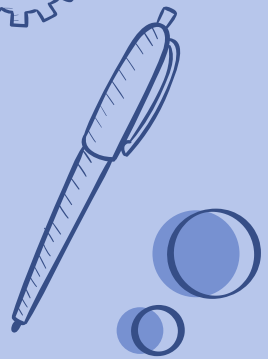
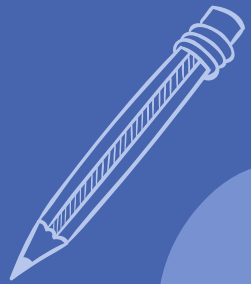
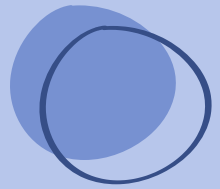


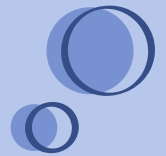
TABLE OF CONTENTS



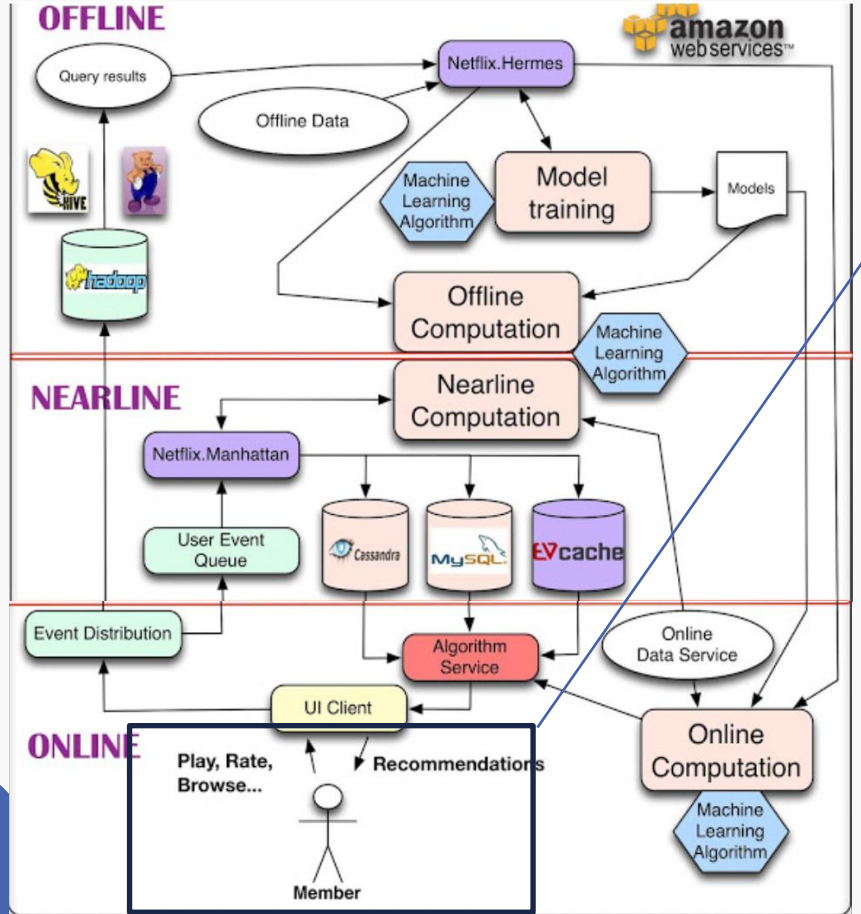
Introduction

01

Backgrounds – Missing Data

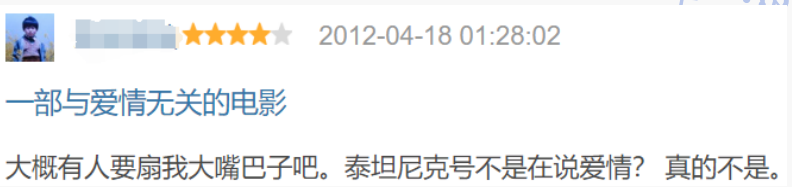


Traditional User-Based Collaborative Filtering:

1. Both A and B love Romantic Films.
(A and B share high Similarity in playing, rating, browsing numerous romantic films.)
2. A likes "Titanic"
(A gives a 5/5 rating)
3. Recommend "Titanic" to B

Question:

Does your history of playing, rating, and browsing tell ALL information about your preferences?



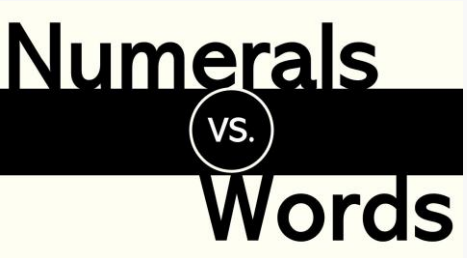
Backgrounds-Why STEAM

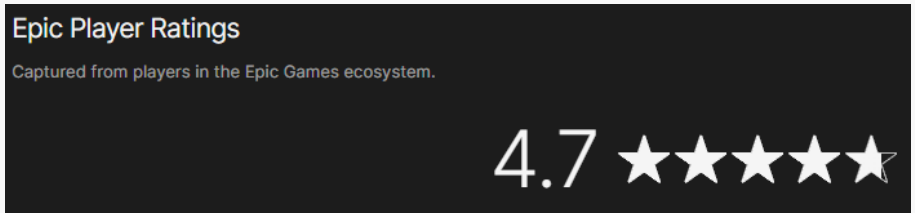
What is Steam?

- Steam is the **biggest** online, cross-platform game distribution system.
- Steam has 120 million monthly active users. 62.6 million of them use Steam everyday.
- Steam catalog includes **more than 50,000 games**.

Why is Steam interesting for Marketing Research?

- A conflict in recommendation algorithm: Numerals or Words?
- Steam's recommendation system and feedback system **DOES NOT** use ratings.
- Instead, steam highly relies on **user comments** and other users' **reactions to comments**.





Research Questions

- **How to predict the future degree of satisfaction of steam users on different games based on their comments?**
- **How to recommend games to the potential customers by sentimental analysis and neural collaborative filtering?**

Will war lovers buy games of anime girls?

Literature Review

02

2 Literature Review

Recommendation system is a key success factor in the competition of online industries

Gallagher, S., & Park, S. H.

- Using the U.S. home video game industry as a case study
- Identifies six generations of technological changes in video game consoles

Conclusion:

- The change of competitive strategies ✕
- The success in a game industry requires building a **network of complementary products**, which causes an increase in **switching costs** and **loyalty**

2 Literature Review

General Methods

What have the previous research done?

Email marketing:

- The development of **Universal Emotional Detector** by utilizing recipient characteristics, headline characteristics, and sending time.
- Demonstrate the effectiveness of **Sentiment Analysis** in Text

Research Gap

Only use the number of opens, number of clicks, open hour, and click hour , no **feedback** or **whether recommend**

2 Literature Review

Knowledge-Based Recommendation System

1. Contribution:

- Utilize CNN and BLSTM-RNN for the detection of depressive and stressful content in users' sentences
- Enhance emotional well-being in online social network (OSN) users

2. Limitation:

Adaptability to diverse user behaviors

Novel self-attention-based Sequential Model

1. Contribution:

Being an order of magnitude faster than comparable CNN/RNN-based models

2. Limitation:

Extending the model to diverse information

- Tags
- Prices
- Reviews

2 Literature Review

Specific Methods for STEAM

What have the previous research done?

Factorization Machines (FM), Deep Neural Networks (DNN), and a hybrid model combining both (DeepFM) — are tested for their ability to handle multiple inputs and various input variable types

Future Directions

Parameter analysis, user studies, and leveraging updated datasets for more extensive text analysis.

How to Fill the Research Gaps

- When making sentiment analysis, we need to consider not only the customer's past purchase behavior but also the customer's **preferences**, **purchasing power,** and **preference about items**
- The ratings might not be reliable, because they are easily adulterated under manipulations. Compared with rating, **reviews** and **whether to recommend** may also be a credential worth using

Methodology

03

Dataset Description

Australian User Reviews

- Comments data from 25799 user of 3682 games.

Why we use this dataset?

- Large number of samples for statistical significance
- Highly resembled audience (Australian young gamers, English speakers).
- Only comments, no ratings, which is suitable for sentiment analysis.
- Plenty of game options for recommendation with points of parity and difference.

Users' Review

user_id	de_Butl3r
posted	October 13, 2014
helpful	0.38 (3 of 8 people found this review helpful)
recommend	True
Item_id	209870
reviews	It's a very fun game i recommend as its nearly like.....

Games' Info

develper	Kotoshiro
genres	['Action', 'Casual', 'Indie', 'Simulation', 'Strategy'],
item_id	761140
title	'Lost Summoner Kitty'
release_data	'2018-01-04'
tags	['Strategy', 'Action', 'Indie', 'Casual', 'Simulation']
price	4.99
discount_price	4.49
early_access	False

# of users: 25799

# of items: 3682

Sentimental Analysis

- **Goal:** Estimate the user's attitudes towards items
- **Input:** dataset "users' reviews"
- **Model:** Transformer models BERT
  - DistilBERT, a streamlined version of BERT
  - Smaller and faster
- **Output**: (-1, 1)
  - Set Threshold at 0
  - If > 0, Positive attitude
  - If < 0, Negative attitude

Reviews	Label	Score
would 100% recommend this to my friends but it also has its down sides like when you accept for a game then 2 minutes later get a ban, duping and many other problems i can be bothered typing.	POSITIVE	0.690941
So much fun! All the classes, weapons all suit that class and they are also quite well balanced.Much fun to play!	POSITIVE	0.999879
Good game. The VAC errors have gotten increasingly bad though, hard to play now.	NEGATIVE	-0.989987
看见SAVANT就进来了, 实在是太吊了!	POSITIVE	0.960474
This game is amazing, and it still is only in early access but yes it does get bit dull aft er a while but hey if we help support the g ame and give feedback then together, we can build this game from a shitty house to a villa in Italy	NEGATIVE	- 0.873343

Neural Collaborative Filtering

FM

DeepFM

- **Goal:** Predict users' ratings of other items and offer recommendation
- **Reason:** Features are correlated with each other
  (E.g. User with positive attitude more likely have "Recommend=True")
- **Input:**
  - User-item interactions + Other features (recommend, helpful, etc.)
  - Each categorical field is represented as a vector of one-hot encoding
  - Each continuous field is represented as the value itself
  - Process them into a sparse matrix
- **Object Function:** MSE
- **Output**:

$$y_{FM} = \langle w, x \rangle + \sum_{j_1=1}^{d} \sum_{j_2=j_1+1}^{d} \langle V_i, V_j \rangle x_{j_1} \cdot x_{j_2}$$

Order-2 feature interaction

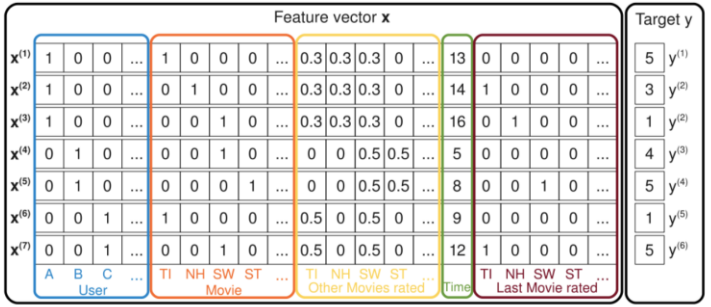


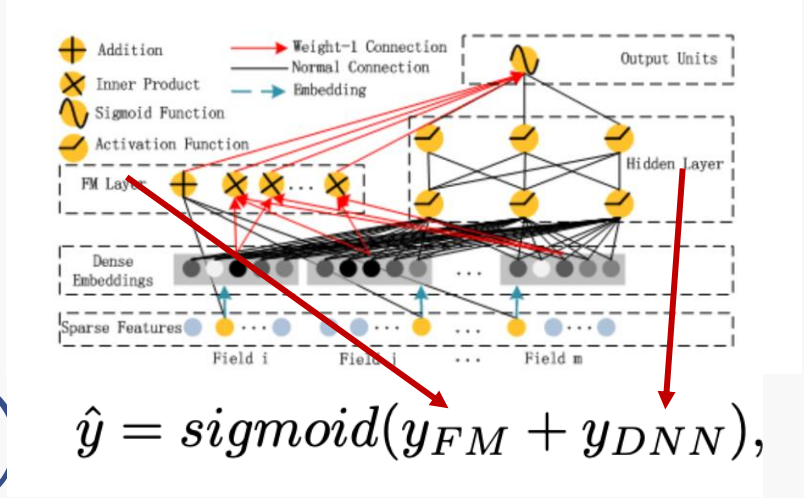
Figure 5: Example of FM's input and output data format.

Neural Collaborative Filtering

FM

DeepFM

- **Advantages:**
  - Allow non-linear relationships between features.
  - Capture more complex high- order feature interactions.
- **Input & Output:**



Comparision of Accuracy

	FM	DeepFM
RMSE	1.156429745	< 1.156429745

More likely to have Decrease
in MSE from FM to DeepFM

Neural Collaborative Filtering

FM

User-item interaction

Other features

Problem

Cold Start

Can't be used on
new items & new users

DeepFM

FM Component
DNN Component

Hybrid Model

What is Hybrid Model?

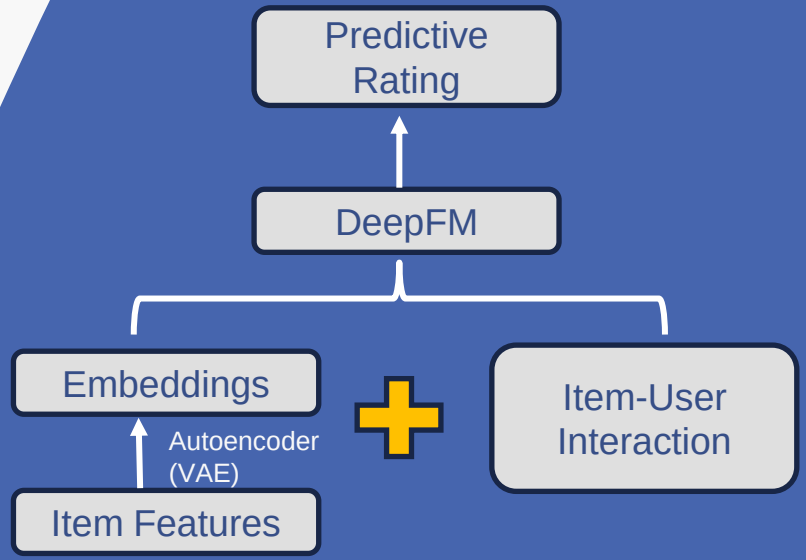
- Combine content-based method with DeepFM

How to build Hybrid Model?

- **Step 1:** Include items' information in training data (genres, tags and descriptions)
- **Step 2:** Use Autoencoder (VAE) to extract embeddings for items; Combine them into the sparse matrix; Train a new DeepFM to predict ratings
- **Step 3:** With descriptions, genres, tags, we can extract their embeddings by autoencoder. Use trained DeepFM to predict ratings

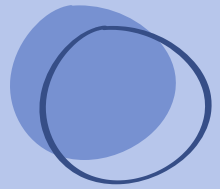
Cold Start for Item SOVLED!!

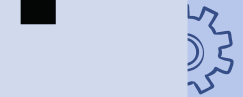
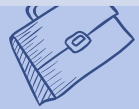
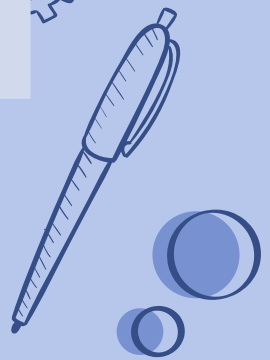
Pipeline



Result & Conclusion

04

Numerical Results

- **Sparse Matrics Result (FM+DNN)**
  - Root Mean Squared Error: 1.156429745
  - Hybrid training set AUC: 0.61551553

User id:sergarino

	recommended_item	score
0	4000	1.785538
1	48240	1.453563
2	359320	1.415809
3	4700	1.228740
4	383080	1.129833
5	333600	0.952720
6	299360	0.911937
7	620	0.790226
8	218620	0.752030
9	239030	0.748293

User id:76561198076909484

	recommended_item	score
0	730	6.060867
1	351570	4.817703
2	277430	4.622171
3	4700	3.964309
4	304410	3.776902
5	383080	3.746215
6	17410	3.691403
7	204300	3.625409
8	113400	3.512536
9	265930	3.327490

User id:76561198090715178

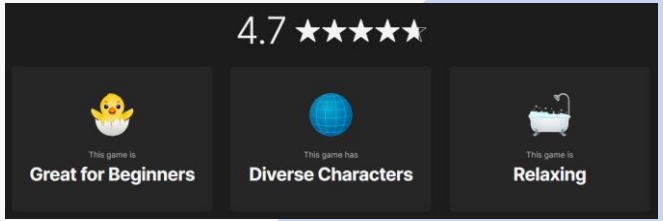
	recommended_item	score
0	304930	5.767410
1	211820	4.800005
2	220200	4.243279
3	440	3.381788
4	4000	3.314588
5	333930	3.299770
6	49520	2.652835
7	570	2.570523
8	31280	2.398552
9	287980	2.362482

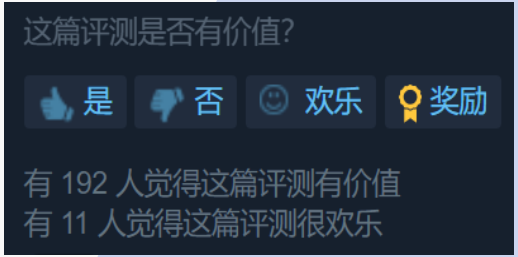
- **Sentimental Result (Label + Score)**

Reviews	Label	Score
It's a very fun game i recomend as its nearly like TITANFALL but its FREE!Play this game now	POSITIVE	0.981100
So much fun! All the classes, weapons all suit that class and they are also quite well balanced.Much fun to play!	POSITIVE	0.999879
Tower defence with rather standard level design and tower design. Game seems like a bad console port and as I was on my last 2 achievements the game crashed and reset my progress.	NEGATIVE	-0.999742
看见SAVANT就进来了，实在是太吊了!	POSITIVE	0.960474
This game is amazing, and it still is only in early access but yes it does get bit dull after a while but hey if we help support the game and give feedback then together, we can build this game from a shitty house to a villa in Italy	NEGATIVE	- 0.873343

Managerial Insights

- **Importance** of **reviews** for recommendation system
  - √ Check the score and click
  - √ √ Pay attention to the reviews (Show the attitude of users in a more natural way)
- Collection of **diverse information**
  - √ User feedback
  - √ √ Feedback from other users (Whether other users find the comment helpful)
- **Sentiment analysis** on **user reviews**
  - Capture customer characteristics
  - Design more effective campaign/marketing strategies





Contribution & Future Work

Contribution

- Predict **degree of satisfaction** of Steam users
- Assist the capture of **users' preference**
  - Better recommend game user preferred products to **potential customers**
  - Increase **potential revenue**
- **Generalize** to other entertainment platform or service

Limitation

- Only applicable to old users, not new users.
- Unsure about authenticity of reviews

Future Work

- Difficult to recommend for new users → Collect initial preferences when new users open the game → Input in the model
- Comment **MAY NOT** accurately reflects user characteristics → Expand the scope of textual/sentimental analysis