MoVis - Movie Recommendation and Visualization

CPSC 547 Infomation Visualization Project

Ye Chen - clara.yechen@gmail.com

Yujie Yang - yangyujie.hust@gmail.com

Introduction

Nowadays, movies becomes a popular way for people to releax after one day's intense work, or spend spare time together with friends. And a good movie recommendation system could help users find a satisfying movie without much exploration and searching.

Tasks

1) Finding a target movie when the user has preferences for the rating, boxoffice, genre or year of movies. Current movie recommendation websites usually produce a list of movies after the user choose their preferences, as shown in Figure1. However, the users have to scroll up and down to find a target movie on the list without the overview of the selected movies. And therefore our aim is to help produce an overview of the movie data according to the user's preferences, as well as looking into the details of each movie.

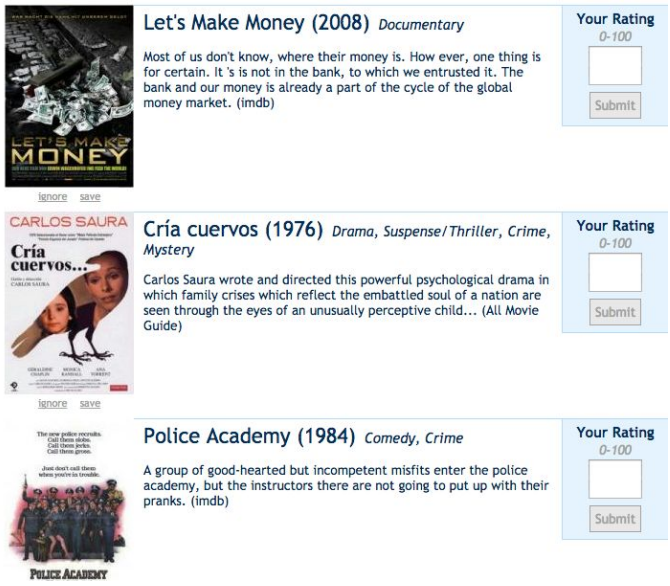

Figure 1. Criticker recommendation page

2) Exploring the movies casted by a specific actor/director and the other actors/directors. When the user is a fan of a actor/actress, he/she would be interested to see what movies

were casted by the actor/actress and also the networks of cooperations. Our aim is to provide an overview of related movies and a network graph of a specific actor/actress.

Dataset

Our dataset is from IMDB PRO, where they provide information about a movie's name, running time, release data, genre, boxoffice, rating, summary, list of directors and list of movie actors, as shown in Figure 2. The data we have scrawled is composed of the top 1597 movies from 2001 to 2011, as shown in Figure 3. In next step, we may scrawl the newest dataset from IMDB PRO with recently published movie datas. Since the data is mostly clean, we only have to do some simple data transformation from text to number like boxoffice; and text to network data like actors.

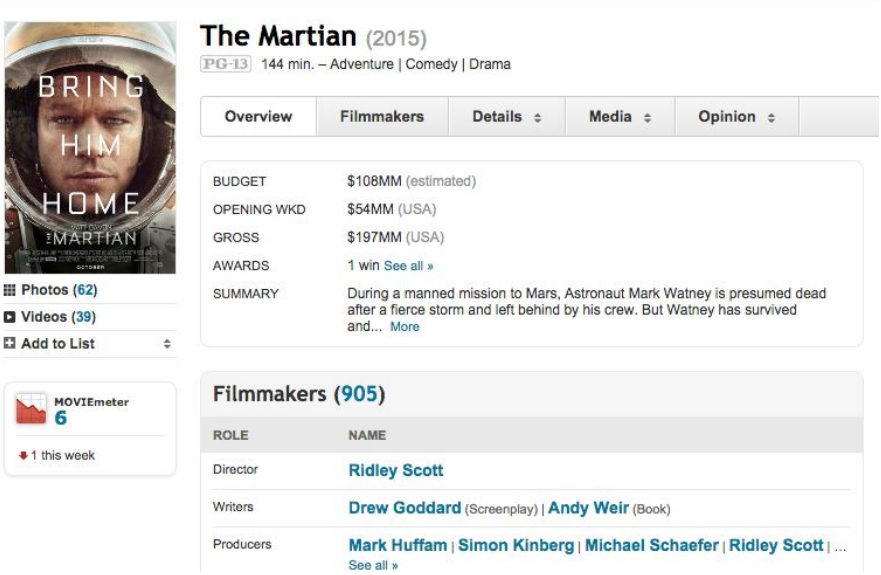

Figure 2. IMDB PRO page

id	imdb	name	year	rate	runtime(mi	certificat	boxoffice	budget	opening_th	release_da	budget_US	budget_oth	other_type	Release Da	gener	gener	gener	gener
1	tt0499549	Avatar (20	2009	7.9	162	PG_13	761000000	$237,000,0	3,452	18 Decembe	237000000			16-Dec-09	Action	Adventure	Fantasy	Sci-Fi
4	tt0468569	The Dark Ki	2008	9	152	PG_13	533000000	$185,000,0	4,366	18 July 200	185000000			14-Jul-08	Action	Crime	Drama	
9	tt0383574	Pirates of	2006	7.3	151	PG_13	423000000	$225,000,0	4,133	7 July 2006	225000000			24-Jun-06	Action	Adventure	Fantasy	
10	tt0435761	Toy Story 3	2010	8.4	103	G	415000000	$200,000,0	4,028	18 June 20	200000000			17-Jun-10	Animation	Adventure	Comedy	Family
14	tt1055369	Transforme	2009	6	150	PG_13	402000000	$200,000,0	4,234	24 June 200	200000000			22-Jun-09	Action	Adventure	Sci-Fi	
16	tt1201607	Harry Potte	2011	8.1	130	PG_13	381000000	$125,000,0	4,375	15 July 20	125000000			11-Jul-11	Adventure	Fantasy	Mystery	
18	tt0121766	Star Wars:	2005	7.7	140	PG_13	380000000	$113,000,0	3,661	19 May 2005	113000000			19-May-05	Action	Adventure	Fantasy	Sci-Fi
22	tt1399103	Transforme	2011	6.3	154	PG_13	352000000	$195,000,0	4,011	29 June 20	195000000			28-Jun-11	Action	Adventure	Sci-Fi	
26	tt0413300	Spider-Man	2007	6.2	139	PG_13	337000000	$258,000,0	4,252	4 May 2007	258000000			30-Apr-07	Action	Adventure		
27	tt1014759	Alice in W	2010	6.5	108	PG	334000000	$200.000.0	3.728	5 March 20	200000000			05-Mar-10	Adventure	Family	Fantasy	

Figure 3. Example of dataset

Previous work

FilmFinder[1] features tightly coupled interactive filtering, where the result of moving sliders and pressing buttons is immediately reflected in the visual encoding.

Co-author graph in Microsoft academic search(Figure 4) show us the top 30 co-authors of the main researcher. The distance of two nodes has a negative correlation with cooperation times of the two people. It allows users drag the nodes to explore and change the main researcher in the middle.

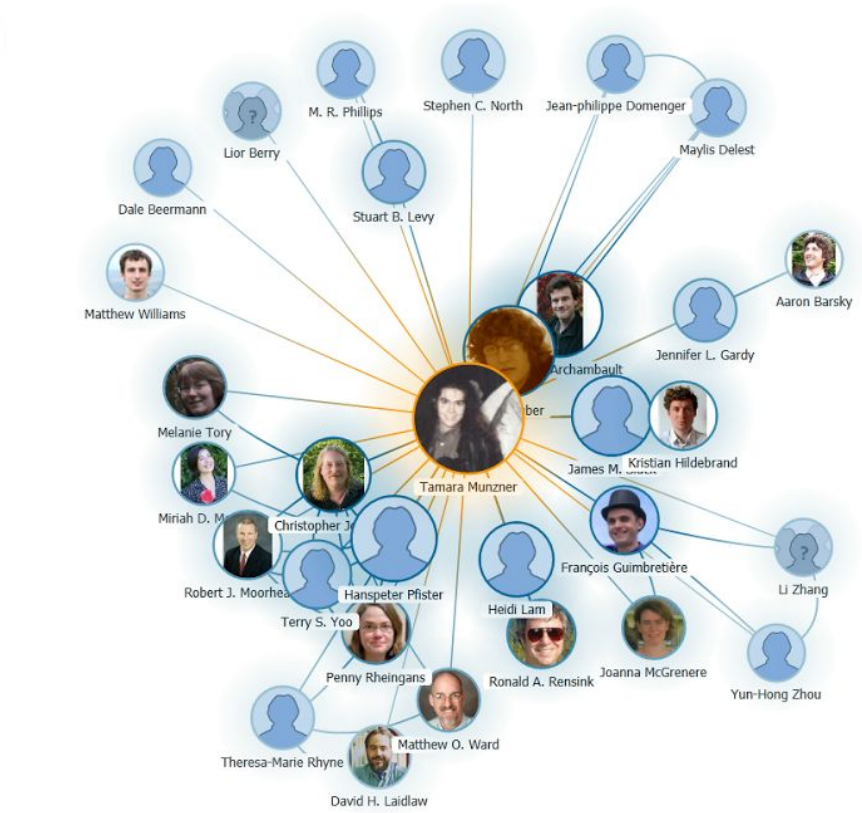

Figure 4. Co-author graph of Tamara on Microsoft academix research[1]

MoVis Description

Main Page

To give users a general taste of all the movies, we design this main page Figure 5, with a Time X Rating scatterplot as the main view. In the main view, each point stands for a movie item, and the users could zoom in/out to explore the dense part. We encode genre into color hue, and show the distribution of genre in a histogram below the main view. Meanwhile, the histogram also acts as a genre filter. We divide boxoffice into three classes and encode them into size. And runtime could be used to filter movies by change the min and max of the range on the time axis. All the attributes and the corresponding channels used in the main view are listed in Table 1.

Table 1. Attributes and the corresponding channels used in the main view

Attribute	**Channels**	**Notes**
Release Time	Spatial region(x axis)	Scatterplot
Rating	Spatial region(y axis)	Scatterplot

[1] From http://academic.research.microsoft.com/VisualExplorer#878415

Genre	Color hue	Histogram; Filter
Runtime	Position on common scale/Length	Filter
Boxoffice class	Size	Divided into three classes
Movie count per genre	Height of the bins	Histogram

We create a new feature called detail finder. The users could choose one of three sizes, and drag the finder window to explore the main view. All the points in the finder window will be listed in the detail list below. And each row includes the link to the movie's iMDB page. When the mouse moves to a specific row, the corresponding point in the main view will be highlighted.

Table 2. What-Why-How analysis of main page

System	MoVis Main page
What: Data	Table: four ordered attributes, one categorical attribute(genre)
What: Derived	Derived table: one ordered attribute(boxoffice class), one quantitative value attribue(item count per genre)
Why: Tasks	Find items by users' choice of filters
How: Encode	Scatterplot; histogram; lists
How: Facet	Multiform; overview-detail; linked highlighting
How: Reduce	Item filtering
Scale	Items: 1,000-10,000

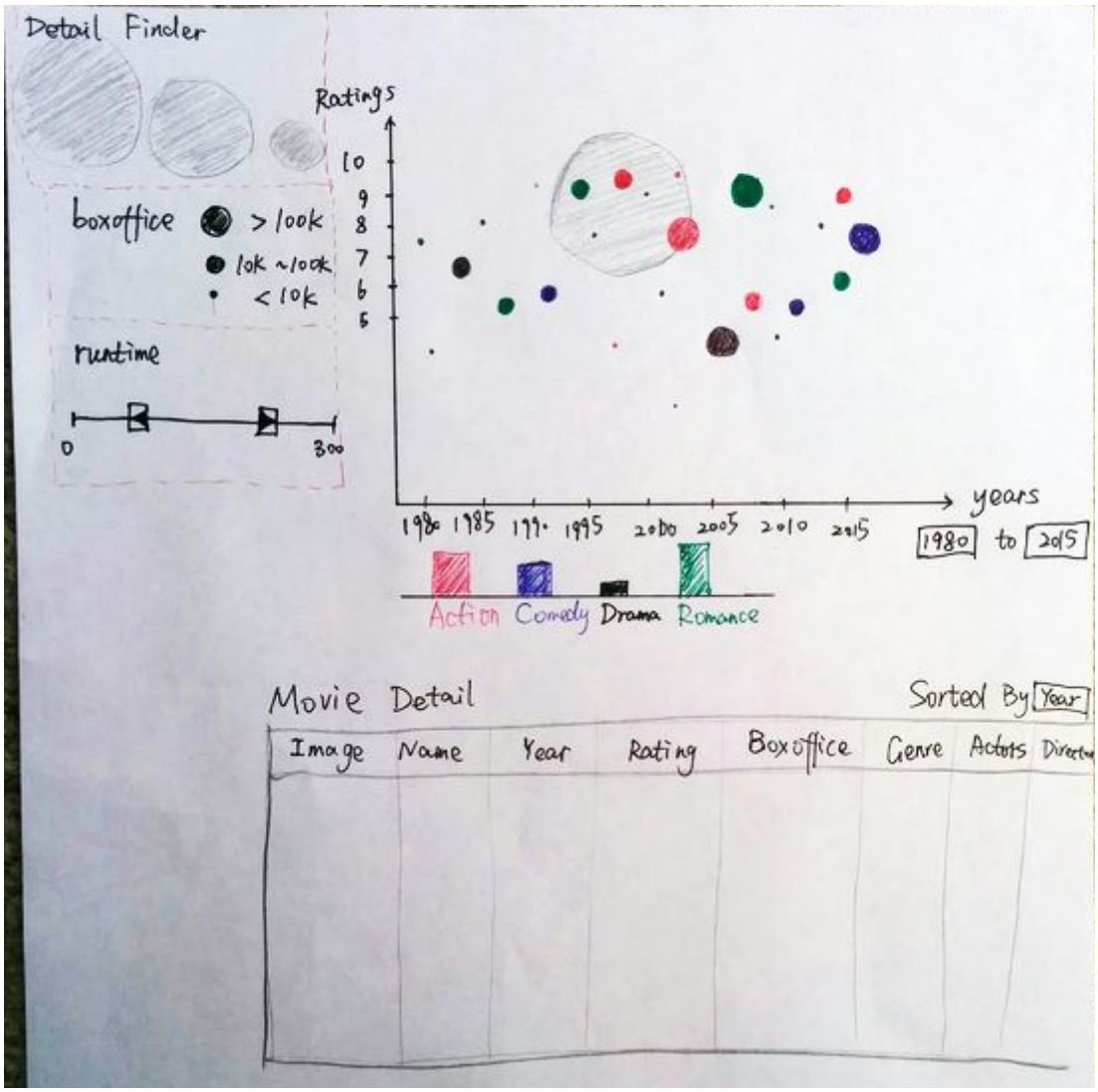

Figure 5. Overview of the main page

Actor/Director Page

Each actor/director in our data has a personal page showing their products and relationship. Each page has two views(Figure 6): the upper one is similar with the one in main page(Figure 5), but we encodes each movie into image/text. If you click on any image, a detail window of the movie will pop up. And the lower one shows the relationship in node-link diagram with the main actor in the middle of the graph. This idea comes from the co-author graph in Microsoft academic search(Figure 4). People who cooperate with the main actor are encoded into nodes and the relationship into links. The distance of two nodes has a negative correlation with the cooperation times. Table 3 shows the What-Why-How analysis of actor/director relationship vis. When you click on any of the nodes, it will directly go into the corresponding actor/director's personal page.

Table 3. What-Why-How analysis of actor/director relationship vis

Vis	MoVis actor/derector relationship vis
What: Data	Table

What: Derived	Networks(actors/directors as nodes; cooperation of actors/directors in the same movie as links)
Why: Tasks	Explore relavant nodes
How: Encode	Node-Link diagram; encoding connection degree into length of links(short length means strong connection)
Scale	Nodes: dozens; links: dozens

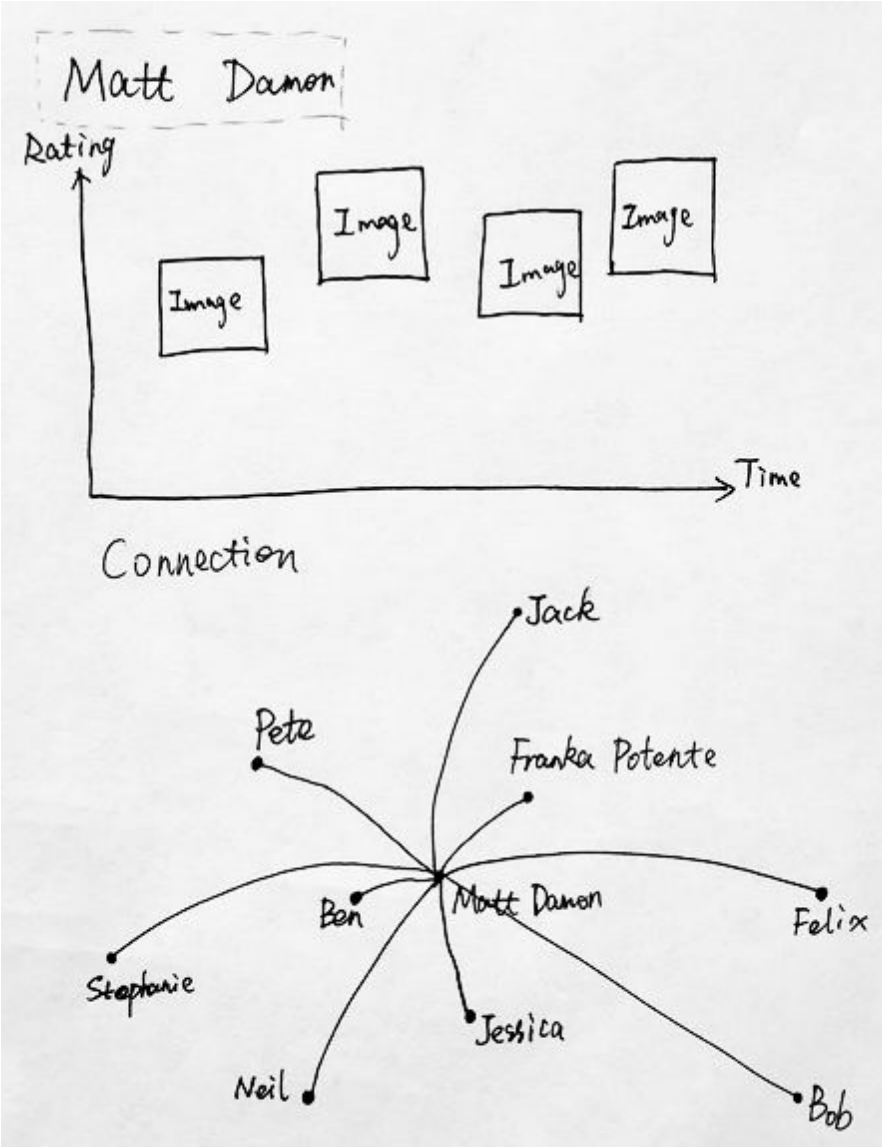

Figure 6. Overview of the actor/director page

Scenario

Scenario I

Stephanie is a computer science student at UBC. In one Saturday night, she finished all the projects just before deadline and wanted to watch a movie to relax at home. She opened the MoVis website, chose her favourate genre animation movies. She zoomed in the scatterplot and draged the finder tool to select several movies with scores higher than 8 in recent 10 years. Then a list of the chosen movies came to eyes. She sorted the list by boxoffice, and suddenly found the first one UP was made by her favourate director Pete Docter. Therefore, she decided to watch this lovely movie and was deeply moved by the romantic story.

Scinario II

Felix is a big fan of Matt Damon. One day, he wanted to watch a movie after working. He went to the personal page of Matt on MoVis website. He found he'd watched all the movies shown on Matt's timeline. When he scrolled down, he saw the network of Matt. He found that Matt has a strong connection with Franka Potente, who is actually the actress of his favourate movie The Bourne Identity. He clicked Franka's name, entered her personal page, and found a movie Blow with a high rating. So he finally chose this one for his night and enjoyed it a lot.

Implementation approach

We will build the vis system using HTML, D3 and CSS. We'll not use any pre-existing software.

Personal expertise

Ye:

I am in the programme of Mangement Information Systems. I had some experiences with web developing and played with Tableau before, but I have not used any VIS programming tool before. I personally like movie a lot, and it would be interesting to play with movie data and make recommendation vis out of it.

Yujie:

I did my bachelor in Automation and transfered to Computer Science now. I've done a few programming projects but I have neither web development skills nor experience with visualization. However, during the learning of vis, I found it interesting to visually "read" and think of data. Although I might not go further deep in this area, it is still neccessary and useful for me to implement those vis concepts into a realistic application. That's bacislly why we decide to do a movie recommendation vis.

Milestones and schedule

Dates	Milestones
Nov 8	Data scrawling from iMDB Pro and agoodmovietowatch.com and data cleansing
Nov 14	Explore D3 and create the scatterplot with filters of the main view
Nov 20	Enable the interaction in the main page (zoom-in and zoom-out function, linked highlighting for other small plots, detail finder function)
Nov 23	Polish the main page
Nov 30	Derive the connection data of actors/directors from the original table and create the network for a specific actors/directors in limited screen
Dec 8	Fulfill the actor/director page and network vis
Dec 14	Finalize the vis and write the report

Reference

[1] Ahlberg C, Shneiderman B. Visual information seeking using the filmfinder[C]//Conference companion on Human factors in computing systems. ACM, 1994: 433-434.