

Module 5

Data Visualization, and Exploration

Processing and Visualizing Data

Process

The process of understanding data begins with a set of numbers and a question. The following steps form a path to the answer:

Acquire

Obtain the data, whether from a file on a disk or a source over a network.

Parse

Provide some structure for the data's meaning, and order it into categories.

Filter

Remove all but the data of interest.

Mine

Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context.

Represent

Choose a basic visual model, such as a bar graph, list, or tree.

Refine

Improve the basic representation to make it clearer and more visually engaging.

Interact

Add methods for manipulating the data or controlling what features are visible.

Visualizing Data

- Scatter plot – this type of graph shows a correlation.
- Pie chart – this one shows proportions. However, make sure you don't use it when you have more than 5 slivers. In that case, use another graph.
- Line graph – use this graph to present trends and patterns.
- Table – this type of graph is perfect for presenting specific values.
- Bar chart – this one is excellent for showing comparisons.

For example, a pie chart is great for presenting how the budget for an ad campaign on social media is being used. This type of graph shows how much money is being spent and where.

<https://whatagraph.com/>

Influence Maximization

Influence maximization is a concept in network theory and social network analysis that **focuses on identifying the most influential individuals or nodes within a network in order to maximize the spread of information, ideas, or behaviors.**

It is particularly relevant in fields such as marketing, social media, epidemiology, and sociology.

Influence maximization problem

- The influence maximization problem is like a puzzle for marketers (but not just for them). Imagine you have a big network of people, like a social media platform. You want to make something popular, like a new product or an idea. However, you have a limited budget, so you can't reach out to everyone directly. Instead, you want to find a small group of people, called influencers, who can help you spread the word to lots of others.
- Think of these influencers as the cool kids in school who everyone listens to. You can find them using different methods, like looking at who has the most friends or who is in the center of social groups.
- Now, here's the interesting part: If you pick just one really good influencer, they can reach about half of the whole network! That's how powerful influencers can be.
- But here's the problem: Figuring out the best group of influencers to pick is really, really hard. It's like trying to solve a super tough puzzle. So, instead of trying every possibility, which would take a really long time, we use smart shortcuts (heuristics) to find a good group of influencers more quickly.
- To give you an idea, if we tried to find the best group of influencers in a big network like the one from the TV show Game of Thrones by checking every option, it would take a really long time and still might not be as good as using these shortcuts. So, that's why we use these clever methods to find the best influencers and make things popular faster.

Common heuristics and clever methods used in influence maximization:

- **Degree Centrality:** This heuristic selects nodes with the highest degree (number of connections) in the network as influencers. Nodes with many connections are more likely to reach a larger audience.
- **Betweenness Centrality:** Betweenness centrality identifies nodes that act as bridges or intermediaries between different parts of the network. These nodes control the flow of information and can be effective at spreading influence.
- **PageRank:** PageRank, originally developed by Google for ranking web pages, can be adapted to identify influential nodes in a network. It measures the importance of a node based on both the number and quality of its connections.
- **Closeness Centrality:** Closeness centrality selects nodes that are, on average, closer to other nodes in the network. These nodes can quickly transmit information to a wide audience.

Link Prediction

Link prediction has a ton of use in real-world applications. Here are some of the important use cases of link prediction:

- Predict which customers are likely to buy what products on online marketplaces like Amazon. It can help in making better product recommendations
- Suggest interactions or collaborations between employees in an organization
- Extract vital insights from terrorist networks

Link Prediction Problem

Refer Resource book

Collective Classification

Collective classification is a way to make predictions about things in a network, like people in a social network or items in an online store, by considering both the connections between them and their individual characteristics.

Collective classification methods use both structure and attributes, often by aggregating data from neighbors of a node and learning a model on the aggregated data.

In social networks, the degree distribution of nodes follows a power law where few nodes have many neighbors.

High degree nodes have incoming links from low degree nodes of different classes and many nodes have very few edges. Hence, using only local structure may lead to poor predictions.

Also, many social networks allow for different types of interactions (retweet, reply, like, etc.) that affect classification differently.

Applications in Advertising and Game Analytics (Use of tools like Unity30)

Applications in Advertising:

Interactive Ads: Unity allows advertisers to create interactive and immersive ad experiences. For example, you can develop mini-games or 3D product showcases within Unity to engage users more effectively than traditional static ads.

Virtual Product Demonstrations: Advertisers can use Unity to build virtual simulations or demos of their products or services. This provides potential customers with a realistic experience before making a purchase decision.

Augmented Reality (AR) Ads: Unity's AR capabilities enable advertisers to create AR ad campaigns. These ads can overlay digital content onto the real world through mobile devices, enhancing user engagement.

Advergames: Advergames are games created for advertising purposes. Unity is an excellent platform for developing advergames that promote brands or products while providing entertainment.

User Engagement Analytics: Unity's analytics tools allow advertisers to track user interactions within their ad experiences. Advertisers can measure user engagement, interaction duration, and other metrics to assess ad effectiveness.

A/B Testing: Unity can be used to create variations of ads or ad experiences for A/B testing. Advertisers can compare user engagement and conversion rates between different versions to optimize ad campaigns.

Applications in Advertising and Game Analytics (Use of tools like Unity30)

Applications in Game Analytics:

Player Behavior Analysis: Unity's analytics features provide insights into how players interact with games. You can analyze player behavior, progression, and retention rates to optimize game design.

Monetization Strategies: Unity's ad monetization platform can be integrated into games to earn revenue through in-game ads. Analytics can help measure ad performance and its impact on player engagement and revenue.

Level Design and Balancing: Game developers can use analytics to fine-tune level design, character balance, and difficulty progression. Data-driven decisions ensure a better gaming experience.

Player Segmentation: Unity analytics can segment players based on their behavior. Developers can tailor in-game experiences, offers, or content to different player groups.

Predictive Analytics: Advanced analytics can predict player behavior, such as churn prediction (identifying players likely to stop playing) or forecasting in-game item purchases.

Multiplayer Game Analytics: Unity supports multiplayer games, and analytics can be used to track player interactions, match outcomes, and overall game performance in a multiplayer environment.

Performance Optimization: Unity analytics can monitor game performance metrics like frame rate, memory usage, and crash reports. Developers can identify and resolve performance issues for a smoother gaming experience.

Applications in Advertising and Game Analytics (Use of tools likePyCharm)

Applications in Advertising:

Data Analysis and Visualization: PyCharm can be used to write Python scripts for data analysis and visualization. Advertisers can use PyCharm to process advertising data, create visualizations, and gain insights from advertising campaign performance.

Machine Learning and Predictive Analytics: PyCharm provides a convenient environment for developing machine learning models. Advertisers can use Python libraries like scikit-learn, TensorFlow, or PyTorch to build predictive models for customer behavior, ad click-through rates, or ad conversion rates.

A/B Testing: PyCharm can be used to write code for running A/B tests. Advertisers can design experiments, collect and analyze data, and make data-driven decisions about the effectiveness of different ad strategies.

Ad Campaign Automation: Python scripts developed in PyCharm can automate various aspects of ad campaign management, including bid optimization, ad budget allocation, and ad content generation.

Social Media Data Analysis: PyCharm can be employed to collect and analyze social media data relevant to advertising campaigns. Python libraries like Tweepy or praw can be used to access social media APIs.

Applications in Advertising and Game Analytics (Use of tools like PyCharm)

Applications in Game Analytics:

Game Data Analysis: PyCharm can be used to write Python scripts for analyzing game-related data. Game developers can use it to process player telemetry data, gameplay logs, and other game-related metrics.

Player Behavior Analysis: Python scripts can be developed to analyze player behavior within games. PyCharm allows developers to explore player actions, progression, and engagement patterns.

Content Management: Game developers can use PyCharm to manage game content, including scripting game events, dialogue, and character behaviors. This is crucial for storytelling and gameplay design.

Modding and Scripting: PyCharm can be used for scripting and modding games. Game developers and modders can create custom game features, assets, or modifications using Python scripts.

Game AI: Python is often used to develop game AI components. PyCharm provides a coding environment for designing and implementing AI algorithms and behaviors.

Player Feedback Analysis: Python scripts developed in PyCharm can process and analyze player feedback from forums, reviews, and surveys, helping game developers make improvements.

Collecting and analyzing social media data

1. Define Your Objectives:

Determine the goals of your analysis. What questions do you want to answer? Are you looking to understand user sentiment, track trends, or analyze user interactions?

2. Choose Social Media Platforms

3. Data Collection:

- a. API Access
- b. Web Scraping
- c. Third-Party Tools

4. Data Storage

5. Data Preprocessing

6. Sentiment Analysis

7. Text Analysis

8. Visualization

9. Trend Analysis

10. User Engagement Analysis

11. Reporting and Insights

12. Continuous Monitoring

Visualization approaches

1. Slopegraphs
2. Parallel Coordinates
3. Alluvial Diagrams
4. Sunbursts
5. Circle Packing
6. Horizon Charts
7. Streamgraph

1. Why is influence maximization a complex computational challenge in social networks?
2. How can link prediction methods be employed to discover potential user connections within a social network?
3. In the realm of game analytics, how can Unity be utilized to analyze player behavior and enhance game design?
4. What is the significance of virality in the context of influence maximization, and how does it relate to the spread of influence in social networks?
5. Can you provide an instance where collective classification improves predictions in a real-world application?
6. Could you illustrate an interactive experience using Unity and offer an example to showcase its implementation?
7. What is the rationale behind data preprocessing in the analysis of social media data, and why is it essential in deriving meaningful insights?