Analysis of Scene Patterns Between U.S. and China Insurance Commercials

Zhe Pang (zp2249@columbia.edu)
Advisor: Professor John Kender (jrk@cs.columbia.edu)

Abstract

Advertising for insurance products is a highly competitive and rapidly growing market in both the United States and China. Insurance companies in both regions use various advertising strategies to attract customers, including emotional appeals, celebrity endorsements, and humor. However, cultural differences can significantly impact the effectiveness of these strategies in different regions due to differing attitudes toward the products.

This study focuses on analyzing the scene patterns in insurance commercial advertising in the United States and China. Scene patterns refer to the sequence and duration of scenes used in advertising and are a critical component in creating effective advertising campaigns. Data is collected from popular video-sharing platforms, including YouTube and Bilibili, and various tools and techniques are used, including scene detection using an open-source library, decision tree modeling, and Markov modeling, to process and analyze the data. In this way, we can provide insights into the cultural differences and similarities in advertising strategies for insurance products in the United States and China to help insurance companies better understand their target audience and create effective advertising campaigns in these two regions.

Introduction

The advertising industry is a competitive market in both the United States and China, and advertising strategies are key to attracting customers. However, cultural differences can significantly impact the effectiveness of these strategies in different regions. Previous studies have shown that advertising strategies that work in one region may not be effective in another due to cultural differences. Cultural values, attitudes, and beliefs are reflected in advertising, and understanding these differences is critical for companies seeking to expand into new markets.

In this context, insurance commercials were chosen as the target for this study, as they are a prime example of how cultural differences can impact advertising strategies and insurance products are often perceived differently in different regions due to cultural attitudes towards risk, security, and individual responsibility.

This report analyzes the scene patterns in insurance commercials advertising in the United States and China. The data collection for this study was done using the YouTube API and a web crawler for Bilibili. A random selection of 100 videos for insurance commercials in the United States and China was used. Pyscenedetect, an open-source library, was used for scene detection, and a decision tree model was employed to classify each scene into short, medium, or long. Markov models and 2-gram analysis were used to examine the transition probabilities between different scene states and scene combination probabilities.

The findings of this report provide insights into the cultural differences in advertising strategies for insurance products in the United States and China. By analyzing the scene patterns used in commercials in each region, we provide a detailed comparison of the advertising strategies used and identify key differences between the two regions. This study contributes to the literature on cultural differences in advertising and provides practical implications for companies seeking to create successful advertising campaigns in these two regions.

Methodology

Data Collection

The initial step of this project involves collecting insurance commercials from both the United States and China that meet our research requirements. We aim to gather a diverse range of commercials that can be representative of each region.

YouTube

YouTube is a widely recognized video-sharing platform that is particularly valuable for collecting data related to advertising. As a popular platform owned by Google, it boasts a vast user base and an extensive range of content, making it an invaluable resource for data collection. In the context of this study, YouTube serves as an excellent source for gathering data on insurance commercials in the United States.

Firstly, Many US-based companies have official YouTube accounts where they upload their commercial ads, while other users may also upload commercials from US companies onto the platform. Consequently, YouTube provides a rich and diverse collection of advertising content specifically from the US market, making it an ideal resource for analysis.

Additionally, YouTube offers the YouTube API, which is designed to facilitate easy retrieval of videos from the platform. This API allows developers to make queries and retrieve specific videos or collections of videos based on various criteria, such as keywords, channels, playlists, or categories.

In this study, we employed the Search API provided by YouTube to retrieve US insurance commercials specifically. By sending the query "US insurance commercials" through the API, we were able to obtain a list of videos that matched our criteria. To ensure that we retrieved individual commercials rather than compilations of multiple ads, we implemented a filter to only retrieve videos with a duration of up to 1 minute.

The Search API yielded a comprehensive list of videos along with their corresponding information, which was returned in a JSON file format. For each video in the list, we utilized the pytube library, an open-source Python package, to download the videos from YouTube. Pytube offers a user-friendly and efficient interface that allows developers to retrieve YouTube videos by providing the video ID. It not only facilitates the downloading of videos but also provides access to additional data such as audio files, titles, descriptions, and video lengths.

By passing the video ID obtained from the Search API's results to pytube, we were able to seamlessly download the US insurance commercials.

Bilibili

Bilibili is a highly popular video-sharing platform in China, known for its large user base and diverse content. It has gained significant popularity among the Chinese population and has become a go-to platform for sharing and consuming videos, including advertisements.

In the context of this study, Bilibili serves as a valuable source for obtaining insurance company video data from China.

One of the challenges in collecting video data from Chinese insurance companies is the limited presence of official accounts on video-sharing platforms. Unlike social media platforms like Weibo, where insurance companies predominantly promote their products through text and pictures, finding official accounts on video-sharing platforms that upload insurance ads is challenging. However, Bilibili provides a solution as many users upload insurance ads on the platform. This user-generated content creates a substantial collection of insurance commercials that can be accessed and analyzed, simplifying the data collection process.

Unlike YouTube, Bilibili does not provide an official API for data retrieval. So we applied a web scraping process for the data collection and utilized an open library downloader.

First, we initiated a search request on the Bilibili platform using the website's search entry and the keyword "Chinese insurance commercials" in Chinese. This search query enabled us to obtain a list of relevant videos for analysis.

After obtaining the search results, we parsed through the HTML file to extract the necessary information. Specifically, we retrieved a fake URL for each video, which contained the bv ID (video ID in the Bilibili website API). This fake URL served as a reference for accessing the videos.

We then utilized the bilibili-downloader package to facilitate data retrieval. By passing the fake URL to the bilibili-downloader, it converted the fake URL into the actual URL of the video. This conversion process allowed us to download the videos directly to our local storage. In addition to video downloads, the bilibili-downloader can also retrieve the video title and duration, which enables us to apply the same filter as we had in the YouTube data collection flow in order to prevent multiple ads in one video.

Dataset

Using the streamlined data collection process described above, we were able to collect a total of 50 random US insurance commercials and 50 random Chinese insurance commercials for this project. It is important to note that the sample size is relatively small, even though we employed automated data collection methods. This is primarily due to the nature of Chinese insurance commercials on Bilibili, which are predominantly user-generated rather than official uploads. Consequently, there is a higher likelihood of duplicate commercials appearing in the final results.

Since the videos on Bilibili are uploaded by regular users, relying solely on the video descriptions and titles may not be sufficient to identify duplicate content accurately. As a result, a manual review of the collected data was necessary to ensure that no duplicates were included in the final dataset. This manual review process allowed us to eliminate any redundant commercials and ensure the uniqueness of each included video.

Although the sample size of 100 videos may seem relatively small, it still provides a representative dataset for our project. By carefully selecting a diverse range of US and Chinese insurance commercials, we aim to capture key differences and similarities in scene patterns and analyze the underlying advertising strategies used in both regions.

Video Preprocessing

Once the video dataset for this study has been collected, the next crucial step is to preprocess the videos and extract scenes for further analysis. Scene extraction allows us to break down the videos into meaningful segments that can be studied and compared to uncover patterns and insights.

Scene Detection

The first step in preprocessing the video dataset for analysis is to split the entire video into multiple scenes. A scene is defined as a continuous segment of video footage that captures a specific event, location, or action without any significant changes in content or context. Each scene can be seen as a coherent and meaningful unit within the video, representing a distinct part of the overall narrative or visual composition.

In this project, we employed a visual-based approach to detect and separate different scenes in the videos. To achieve this, we utilized the Pyscenedetect library, which is a Python library specifically designed for detecting shot changes in videos and automatically splitting them into separate clips.

The scene detection algorithm we employed, called "detect-adaptive," is particularly effective in identifying jump cuts in the input video. It analyzes the frames of the video and identifies areas where the difference between two subsequent frames exceeds a threshold value. This threshold value is determined by using a rolling average of adjacent frame changes.

To calculate the difference between two adjacent frames, the algorithm converts the colorspace of each decoded frame from RGB to HSV. By taking the average difference across all channels from frame to frame, the algorithm can effectively detect scene changes.

By passing a list of decoded frames to the scene detector, we were able to obtain a comprehensive list of scenes detected in the video dataset. We can then calculate the scene duration with frame timestamps and retrieve the frame at which the scene change occurs.

Scene #	Duration (sec)	Preview
1	00:00:02.700	
2	00:00:01.400	
3	00:00:01.300	
4	00:00:02.700	
5	00:00:01.333	

Fig1: Examples of Scene Detection Results

Scene Statistics

For each video, we extracted features from the dataset by calculating basic statistics of the extracted scenes from the list. Specifically, we focused on the scene durations and computed the mean and standard deviation for each video. These statistics provide insights into the overall duration patterns of scenes within each video and contribute to the subsequent classification step comparing US and China insurance commercials.

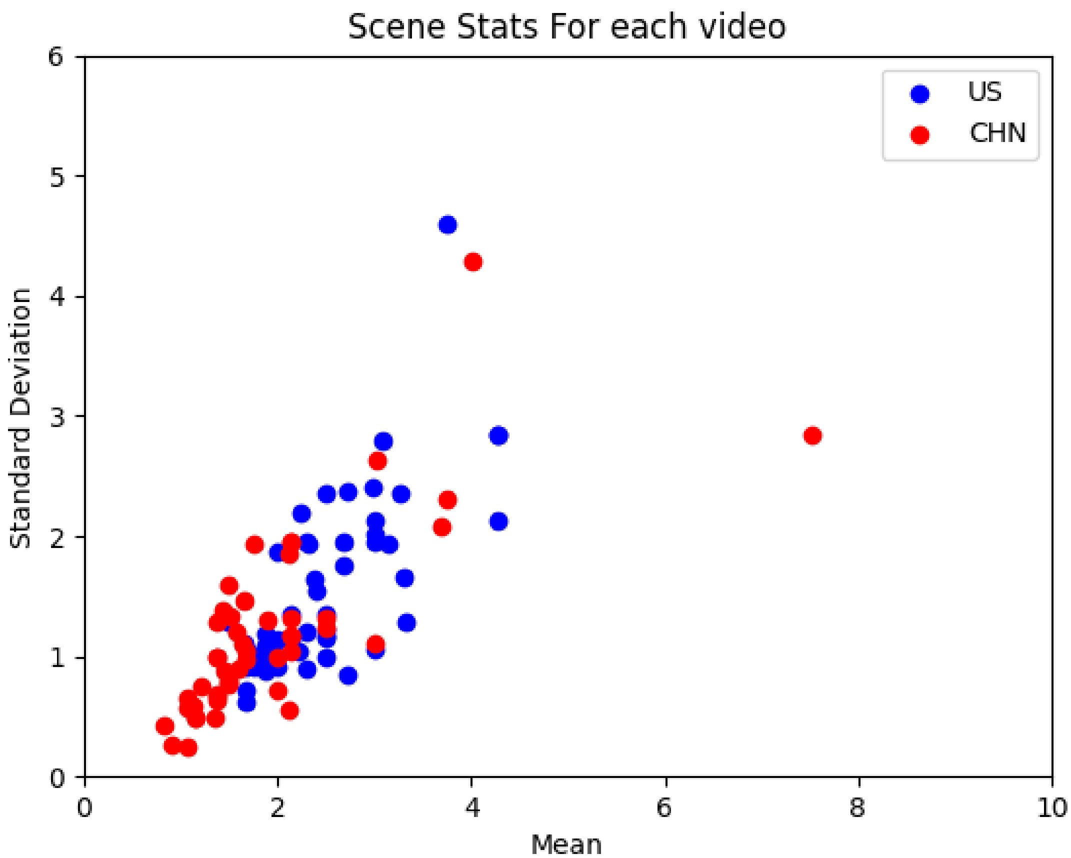

Fig 2: Scatter Plot of scene stats for each video

Video Classification

After preprocessing the videos and extracting relevant features, we can leverage the resulting data to perform classification between US and China insurance commercials. This classification step allows us to categorize each video according to its country of origin, providing us with valuable insights into the differences between videos from these two regions.

Approaches explored

For the classification tasks based on the scene mean and standard deviation of each video, we employed six different approaches: linear SGD classifier, logistic regression, SVM, KNN, random forest, and decision tree.

The linear, logistic, and SVM classifiers yielded accuracy rates of approximately 50% to 55%. Through hyperparameter fine-tuning, the KNN and random forest classifiers can improve accuracy to up to 75%. Notably, the decision tree classifier outperformed the other models, attaining an accuracy of 85% following fine-tuning.

Considering the satisfactory performance of these models, we decided to exclusively utilize them for our classification tasks and not use any neural-network-based models, since these models are easy to interpret and simplify insight extraction.

Classifiers	Accuracy	Precision	Recall
Linear SGD	55%	62.5%	45.45%
Logistic Regression	50%	57.14%	36.37%
SVM	50%	57.14%	36.37%
KNN (with fine-tuned num_neighbors = 9)	75%	75%	81.82%
Random Forest (with fine-tuned max_depth = 14 and n_estimators = 412)	75%	80%	72.73%
Decision Tree	85%	83.34%	90.09%

Fig 3: Results of Different Classifiers

Decision Tree

After assessing the performance of various classifiers, we determined that the decision tree classifier exhibited the highest accuracy. Consequently, we chose this model to uncover key differences in scene distributions between US and China insurance commercials.

The decision tree is a supervised machine learning algorithm. It is a flowchart-like structure where each internal node represents a decision based on one or more features, each branch represents the outcome of the decision, and each leaf node represents a class label or a predicted value. The decision tree algorithm learns from training data by recursively splitting the data based on the features that provide the most information gain or the best separation of classes. The splitting process aims to create homogeneous subsets of data at each node, where the instances within each subset share similar characteristics. This process continues until a stopping criterion is met, such as reaching a maximum tree depth or when further splitting does not improve the classification accuracy significantly.

To ensure interpretability, we fine-tuned the hyperparameters of the decision tree classifier. The specific hyperparameter we focused on was max_depth, which denotes the maximum depth of the decision tree. Through experimentation, we observed that increasing the max_depth beyond 3 did not yield any improvements in accuracy. Hence, we concluded that a decision tree with a max_depth of 3 adequately captures the essential disparities between US and China insurance commercials in terms of scene distribution. Further exploration into deeper decision trees was unnecessary.

Decision Tree max_depth	Accuracy	Precision	Recall
1	80%	73.33%	100%
2	65%	75%	54.54%
3	85%	83.34%	90.9%
4	85%	83.34%	90.9%
5	85%	90%	81.82%

Fig 4: Results of Decision Tree with different depths

Scene Classification

After experiments on multiple classifiers on videos which used the mean and standard deviation of scenes in each video, we can then use the best-performing decision tree model to get insights on the difference in scene distribution in videos from the US and China, and classify each scene based on their durations for the next analysis step using Markov Model and 2-gram.

Short, Medium, and Long

The decision tree model with a max_depth of 3 successfully captures all the essential features from the dataset. Upon examining its structure, we can observe that it makes a total of 6 decisions, with 3 decisions based on the mean of scene durations and 3 decisions based on the standard deviation. The decisions related to the mean of scene durations primarily divide the scene duration time into three distinct intervals: 0 - 1.648s, 1.648 - 2.143s, and above 2.143s. Along with the information from the standard deviation, these intervals are used to determine the class of the video.

As a result, we can utilize these three intervals as a classification for scenes: 0 - 1.648s represents short scenes, 1.648 - 2.143s represents medium scenes, and above 2.143s represents long scenes. By encoding each scene into a single state based on its duration, we can effectively analyze the sequence of scenes in the videos.

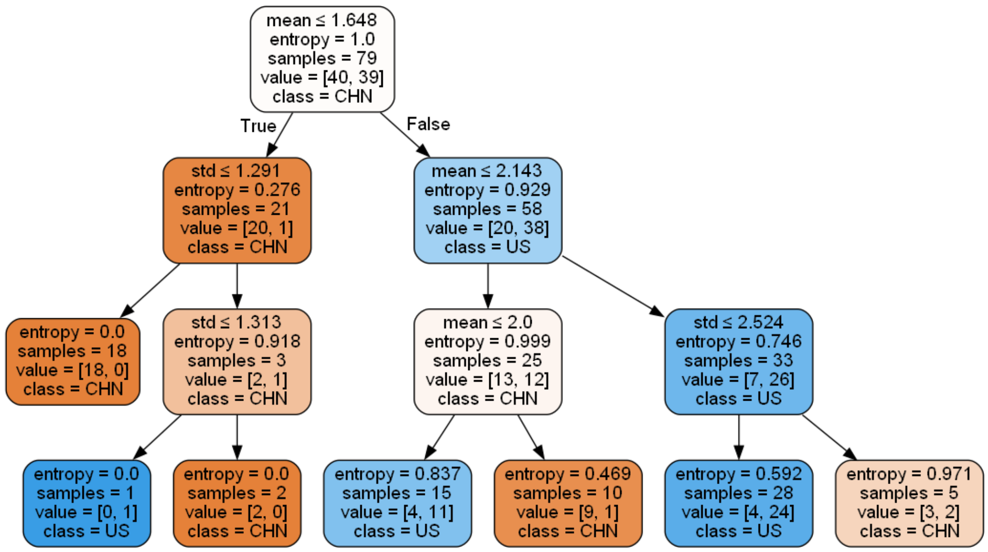

Fig 5: Structure of Decision Tree with max_depth 3

Markov Model and 2-gram Analysis

After successfully classifying each scene into one of the three states (short, medium, or long) based on the decision tree model, we can proceed to analyze the patterns in the sequence of these states. This analysis will provide insights into the key differences between US and China insurance commercials. To accomplish this, we can employ the Markov model and 2-gram approach to find possible patterns in the sequence of different scenes.

Encoding

To facilitate the analysis of scenes using the Markov and 2-gram models, we need to encode each scene into a corresponding state based on the intervals obtained from the decision tree model. For a given list of scenes, we begin by adding a start token "x" at the beginning. Then, for each subsequent scene in the list, we encode it as either "s" (short), "m" (medium), or "l" (long) depending on its duration. Finally, we add an end token "e" to mark the conclusion of the scene sequence. By employing this encoding scheme, we transform the list of scenes into a sequential representation of tags, allowing us to apply the Markov and 2-gram models for further analysis. For the 2-gram model, an additional start token would be needed at the start.

Markov Model and 2-gram

A Markov model, also known as a Markov chain, is a mathematical model that represents a sequence of events or states in which the probability of transitioning from one state to another depends only on the current state and is independent of previous states. It is a memoryless model that assumes the current state encapsulates all relevant information for predicting future states.

In the context of this study, each scene type has enough information to transition into the next scene type. By calculating these transition probabilities based on the observed scene sequence, we can gain insights into the likelihood of different scene transitions and understand patterns or trends within the commercials.

A 2-gram, or bigram, model is a specific type of Markov model. It assumes that the current state or event depends only on the immediately preceding state or event. In the case of scene sequence analysis, the 2-gram model considers the current scene type (state) as dependent on the previous scene type (state), and the transition probabilities are estimated based on the observed frequencies of these state-to-state transitions.

By utilizing the 2-gram model, we can capture more localized dependencies between consecutive scenes and further analyze the sequential patterns within the insurance commercials.

Results

After calculating the transition probabilities in the 2-gram model, several key findings emerged from the analysis of scene sequences in US and China insurance commercials.

	xx	xl	xs	xm	ll	ls	lm	ml	ms	mm	sl	le	me	xe
China	0.5	0.07	0.36	0.06	0.182	0.627	0.118	0.2	0.579	0.179	1	0.074	0.042	0.01
Us	0.5	0.25	0.1	0.15	0.326	0.433	0.171	0.321	0.44	0.22	1	0.069	0.018	na

Fig 6: Results of the 2-gram model

Firstly, it was observed that both US and China commercials exhibit a consistent pattern where a short scene is almost always followed by a long scene. This pattern is commonly observed in video clips featuring conversations or transitions between camera positions and settings. It indicates a shared understanding of video shooting practices in both cultures.

Additionally, there is a notable difference in the sequence patterns between the two regions. In the case of long scenes, Chinese videos have a higher probability of being followed by a short scene compared to US videos, which have nearly equal probabilities of being followed by either a short or long scene. As a result, the most common sequence observed in both regions is "slslsl...". However, US videos tend to have more occurrences of additional long and medium scenes interspersed within the "sl" pairs, leading to patterns such as "sllslllsl," "slmsllsl," "slmlslsl," and so on.

Furthermore, when given a medium scene, US videos have roughly equal probabilities of transitioning to a long or short scene, and with a higher tendency to remain in the medium state than China videos do. In contrast, China videos are more likely to switch to a short scene in the same situation. In addition, the possibility of having an end token followed is much higher in China than in the US. This can be caused by the common duration of China commercials which are mostly within 15 seconds, while the common duration of US videos is 30 seconds. As a result, Chinese insurance companies have a shorter time frame to showcase their brands and contact information compared to their US counterparts. This discrepancy in scene durations reflects the different advertising strategies employed by insurance companies in the two regions.

Finally, it was observed that when given a start token, the scene following it in China videos tends to be a short scene, while in US videos, the scene following the start token is more likely to be a long or medium scene rather than a short scene. This difference indicates a variation in the initial scene selection and pacing between US and China insurance commercials, with China videos favoring a quicker transition to shorter scenes and US videos often opting for longer or medium scenes to begin the video.

These observations indicate that while both US and China insurance commercials are predominantly characterized by a sequence of short and long scenes, US videos are more likely to incorporate additional long and medium scenes compared to their Chinese counterparts.

The Markov model analysis confirms the findings observed in the 2-gram model. By examining the transition probabilities between states, we observe consistent patterns in both US and China insurance commercials. The initial scene patterns remain consistent and "sl" is a common sequence in both regions. Furthermore, US videos exhibit a tendency to include medium and long scenes inserted between the "sl" pairs. When given a medium scene, China videos have a higher likelihood of reaching an end state compared to US videos.

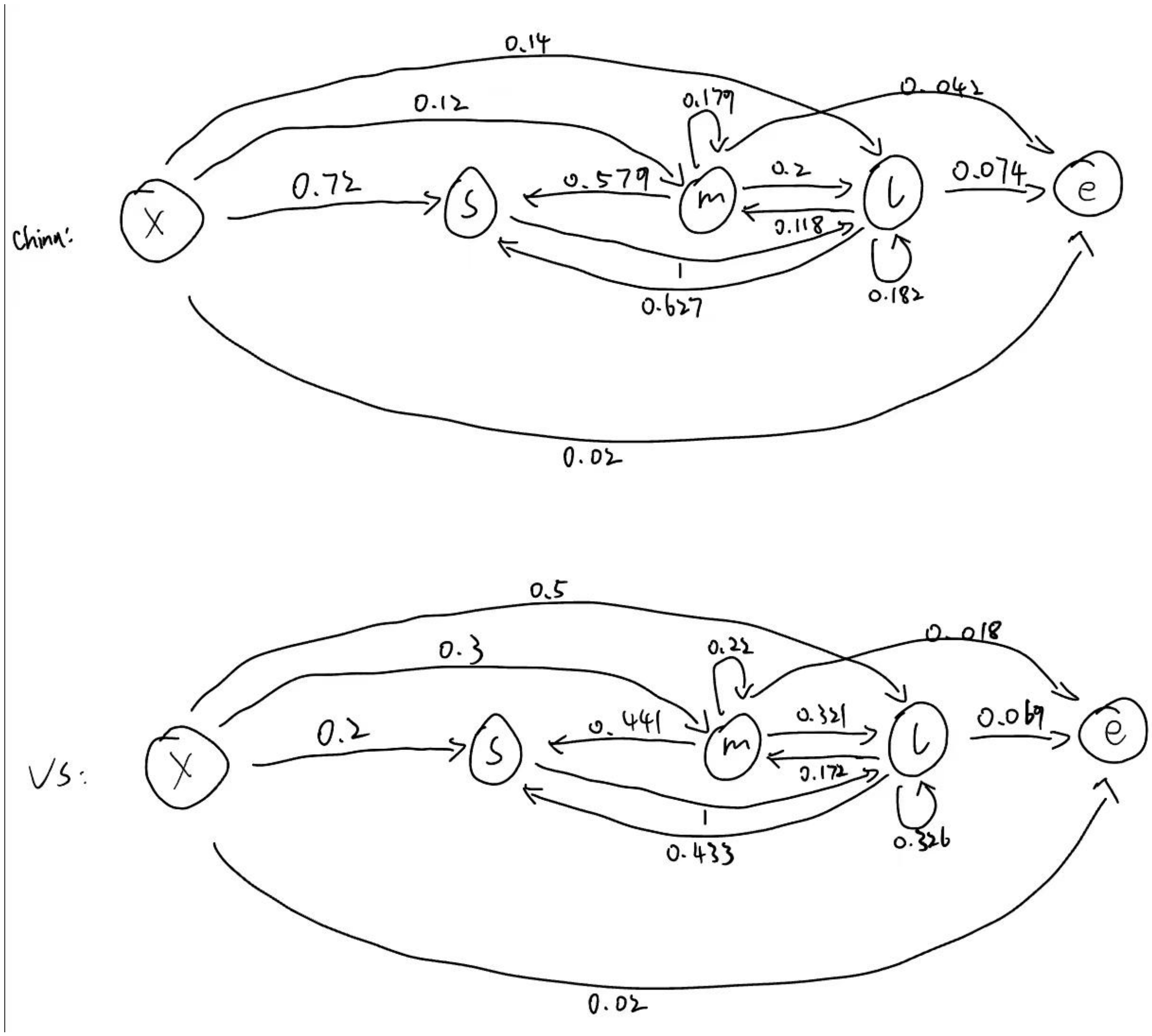

Fig 7: Results of the Markov Model

Conclusion

In conclusion, this research report focused on analyzing and comparing insurance commercials from the United States and China. Through a series of preprocessing steps, including scene detection and classification, we were able to gain valuable insights into the key differences between these two regions.

Using machine learning techniques, such as decision tree classification and Markov modeling, we uncovered interesting findings about the scene distributions and sequences in the videos. Both the decision tree and Markov models (including the 2-gram model) provided valuable information on the scene durations and transitions, shedding light on the underlying patterns and structures of the commercials.

Our analysis revealed that both US and China insurance commercials exhibit a common pattern of short scenes followed by long scenes. However, US videos displayed a higher likelihood of incorporating medium and long scenes in between the "sl" pairs, while China videos tended to have a higher probability of reaching an end state.

These findings suggest that while there are similarities in scene sequences between the two regions, there are also notable differences in the inclusion of additional scenes and the likelihood of ending the sequence. Such variations may reflect cultural differences and video production practices in the respective regions.

In addition to the current research findings, there are several potential avenues for future work that can further enhance our understanding of insurance commercials in the US and China.

Firstly, addressing the issue of data duplicates from Bilibili could significantly increase the dataset size and provide a more robust analysis. By collecting a larger and more diverse set of videos, we can verify if the observed patterns and findings still hold and generalize across a wider range of samples.

Furthermore, incorporating the image frames when scene changes occur can add more detailed visual information to the analysis. By extracting visual features from each scene, we can explore the relationship between visual elements and scene classifications. This can provide deeper insights into the visual composition and aesthetic choices made in insurance commercials.

Another interesting direction is to explore the application of neural network-based approaches for the classification task. Utilizing deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can potentially improve the accuracy of the classification and offer a more nuanced understanding of the underlying features and patterns in the videos. Additionally, interpreting the trained neural network models can provide insights into the important visual and temporal cues that contribute to the classification decisions and find key differences between insurance commercials in the two regions.

Overall, this research provides valuable insights into the characteristics of insurance commercials in the US and China, contributing to a better understanding of the nuances in advertising approaches and storytelling techniques employed by insurance companies in these markets. The findings can serve as a foundation for further studies in cross-cultural advertising research and provide practical implications for marketers seeking to create effective insurance commercials in these regions.

Appendix

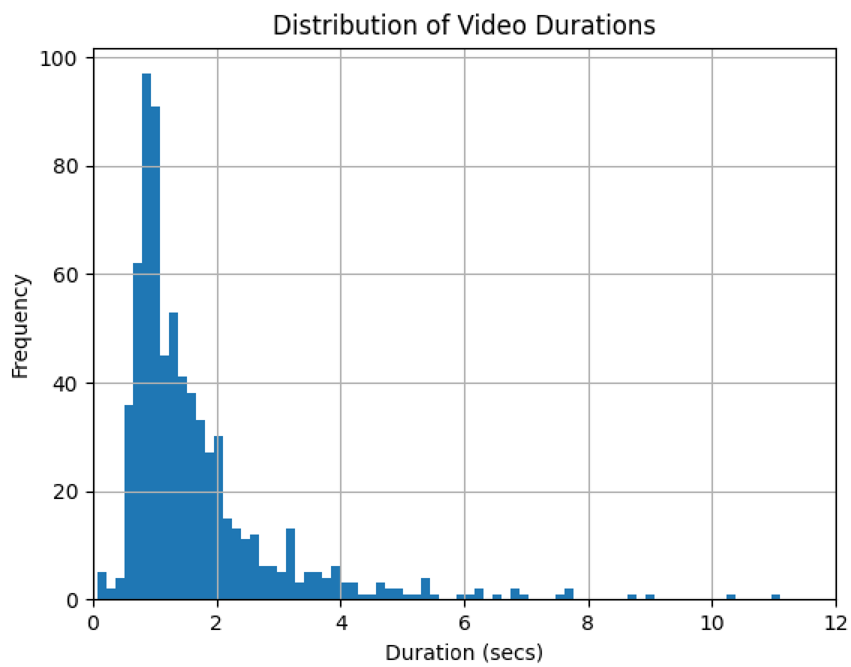

Fig 8: Distribution of Scene Duration in Chinese Videos

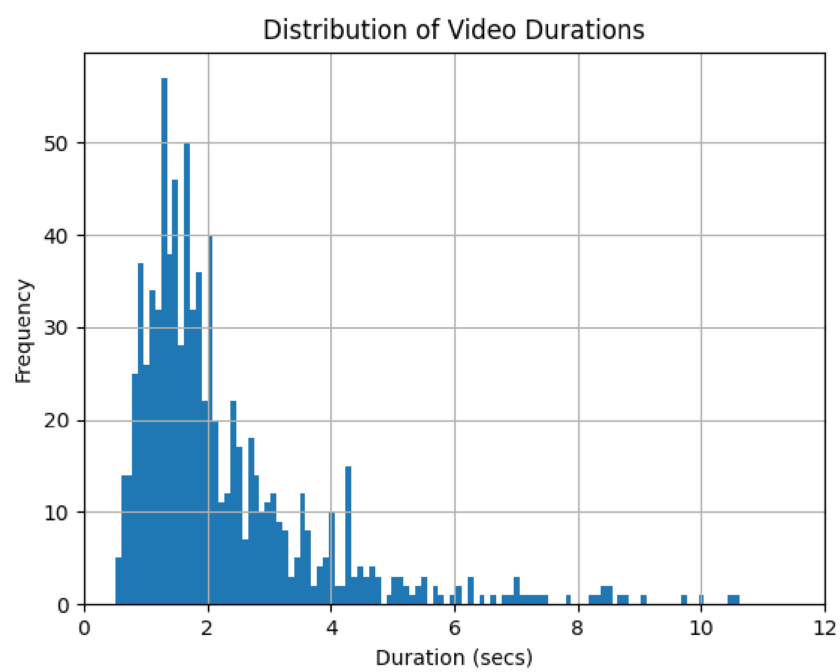

Fig 9: Distribution of Scene Duration in US Videos