Integrating Predictive Analytics and Social Media

Yafeng Lu, Robert Krüger, *Student Member, IEEE*, Dennis Thom, Feng Wang, Steffen Koch, *Member, IEEE*, Thomas Ertl, *Member, IEEE*, and Ross Maciejewski, *Member, IEEE*

Abstract— A key analytical task across many domains is model building and exploration for predictive analysis. Data is collected, parsed and analyzed for relationships, and features are selected and mapped to estimate the response of a system under exploration. As social media data has grown more abundant, data can be captured that may potentially represent behavioral patterns in society. In turn, this unstructured social media data can be parsed and integrated as a key factor for predictive intelligence. In this paper, we present a framework for the development of predictive models utilizing social media data. We combine feature selection mechanisms, similarity comparisons and model cross-validation through a variety of interactive visualizations to support analysts in model building and prediction. In order to explore how predictions might be performed in such a framework, we present results from a user study focusing on social media data as a predictor for movie box-office success.

Index Terms—Social Media, Predictive Analytics, Feature Selection

1 INTRODUCTION

Research on social media has intensified in the past few years as it is seen as a means of garnering insight into human behaviors. The unstructured nature of social media data also provides unique challenges and opportunities for researchers across a variety of disciplines. Businesses are tapping into social media as a rich source of information for product design, relations management and marketing. Scientists utilize social media data as a platform for developing new algorithms for text mining (e.g., [13]) and sentiment analysis (e.g., [45]) and focus on social media as a sensor network for natural experimentation for exploring social interactions and their implications (e.g.,[47]).

In using social media as a sensor network, researchers have developed methods that capture online chatters about real world events as a means of predictive model building. For example, work by Culotta [12] explored the use of Twitter for predicting seasonal influenza. Tumasjan et al. [43] found that the magnitude of Twitter messages was strongly correlated to German elections. Eysenbach [15] utilized regression modeling of Tweet counts to predict paper citations, and Zhang et al. [48] explored mining Twitter for emotions and predicting the opening-value of the stock market.

Currently, the visual analytics community has begun focusing on social media analytics with respect to developing tools and frameworks to collect, monitor, analyze and visualize social media data. Studies have ranged from geo-temporal anomaly detection (e.g., [9]) to topic extraction (e.g., [46]) to customer sentiment analysis (e.g., [33]). Such work focuses on capturing the incoming streams and enables the analysts to perform exploratory data analysis. However, little work has been done on developing tools for predictive analytics using social media. In 2013, the Visual Analytics Science and Technology (VAST) conference ran the VAST Box Office challenge using social media data to predict the opening weekend gross of movies. This particular contest served as an entry point to explore how users interact with visualization tools to develop predictions. Continuing from this contest, our work has focused on utilizing movie data from social media to explore the promises and pitfalls of visualization for predictive analytics. Unlike more specialized data sources (e.g., criminal incident reports, emergency department data, traffic data, etc.), movie data lends itself well to analyzing visual analytics modules as many casual users think of themselves as movie domain experts.

In this paper, we present a framework for social media integration, analysis and prediction. This framework consists of tools for extracting, analyzing and modeling trends across various social media platforms. In order to test our framework, we focus on the specific problem of predicting the opening weekend box-office gross of upcoming movies. This system integrates unstructured data from Twitter and YouTube with curated data from the Internet Movie Database (IMDB). Temporal trends and sentiment are extracted and visualized from social media, and IMDB features can be explored through parallel coordinate plots. Specifically, this tool was developed to support the exploration of predictive models while integrating user interaction to iteratively update the models, compare against past models, and explore similarities between movies. To demonstrate the efficacy of our system, we tested our framework with seven subjects and evaluated their prediction performance. We present lessons learned and future directions for improving the user in the loop workflow for predictive analytics.

2 RELATED WORK

This paper focuses on enabling analysts to explore, validate and filter social media data for predictive analytics. In this section, we discuss past work on current state-of-the-art in visual analytics surrounding both social media data and predictive model development.

2.1 Visual Analytics of Social Media Data

Recent visual analytics systems for social media analysis include Whisper [8], which focused on information propagation in Twitter, SensePlace2 [28], which focused on the analysis of geographically weighted Tweets, and TweetXplorer [31] which combined geographical visualization of Tweets along with their social networks. Other applications have explored the use of social media analytics for improving situational awareness in emergency response. Thom et al. [42] and Chae et al. [9] developed spatiotemporal visual analytics systems that integrated various social media data sources for anomaly event detection and disaster management. Our proposed framework takes cues from this previous work and is developed to integrate data from multiple sources, for our case study, we integrate Twitter, YouTube and IMDB data.

A wide variety of work also exists with regards to social media topic extraction and sentiment analysis of social media. Dou et al. [13] developed an algorithm for hierarchically organizing news content based on topic modeling. Hao et al. [18] applied topic based stream analysis techniques to detect sentiment in Tweets and created a sentiment cal-

• *Yafeng Lu, Feng Wang, and Ross Maciejewski, are with Arizona State University. E-mail: {lyafeng, fwang49, rmacieje,}@asu.edu.*
• *Robert Krueger, Dennis Thom, Steffen Koch, and Thomas Ertl are with University of Stuttgart, Germany. E-mail:{Robert.Krüger, Dennis.Thom, Steffen.Koch, thomas.ertl}@vis.uni-stuttgart.de.*

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

Fig. 1: Front Page of the Frozen Weekend. View (a) is the Tweet and Youtube comments line. The solid lines indicate the number of Tweets per day starting 14 days before the release (x-axis). The left y-axis indicates the number of Tweets.The dashed lines represents the number of Youtube comments per day using the right y-axis. Each color represents one movie. Clicking the legend highlights the corresponding trend line. View (b) is the opening weekend gross bar graph. The left bar indicates the real gross while the right bar indicates the baseline model's prediction. View (c) shows the Tweets and users.

endar and map. Nguyuen et al. [33] applied machine learning to Twitter to extract sentiment and compare dictionary based and machine-learning sentiment classifiers. Wang et al. [45] created a sentiment analysis and visualization system called SentiView to analyze public sentiment in Tweets and BlogPosts. Similar to previous work [24, 27], our framework also performs sentiment analysis on the ingested social media data. However, while previous work relies directly on automatic algorithms, we allow the users to interactively modify the sentiment of an item (e.g., a Tweet) as a means of correcting for classification errors. Overall, our framework builds upon prior visual analytics work with regards to social media analytics and expands this domain with regards to integrating predictive analysis and model building tools.

2.2 Predictive Analytics

It is important to note that our proposed framework is not the first to address predictive analytics. A variety of solutions exist for both novice and expert users (e.g., R [37], SAS [39], Weka [17], JMP [36], Excel). These software packages and tools provide a variety of machine learning algorithms that can be used for predictive analytics tasks, such as feature selection, parameter optimization and result validation. Many of these systems offer basic visualizations including residual plots, scatterplots and linecharts. However, most of their visualization are only used to display the final results and do not provide interactive means for manipulation, feature selection or model refinement; instead, these systems often opt to show baseline models or simple statistical measures for result validation, working as more of a black-box system. The goal of our framework is to directly integrate the analyst into the model building loop by enabling feature selection for model building and comparison. We include tools such as Parallel Coordinate Plots [21] and correlation rankings for quick comparison. Moreover, we have also created a variety of mechanisms for automatically suggesting similar instances within a dataset to enable the analyst to identify outliers and validate models based on the accuracy of prediction with regards to similar instances.

Recently, researchers in the visual analytics community have been developing methods for improving model building and predictive analytics. Berger et al. [5] used regression models for parameter space exploration. Choo et. al. [10] provided a classification system, iVisClassifier, using linear discriminant analysis to reduce dimensionality for improved data classification. Brown et al. [7] designed an interactive visual analysis system to improve clustering results by updating the distance function based on users' feedback to the display. We also integrate feature selection and sample filtering, but our system does not require users to be familiar with specific prediction algorithms. Instead, we focus on how much information and manipulation should be open to the user [2].

Most closely related to our work is that of Mühlbacher et al. [32] which developed an interactive visual framework for selecting subset features to improve regression models. They used R^2 to rank 1D features and 2D feature pairs, as well as a partition-based feature ranking. Their goal is to approximate the local distribution of a given target, and their visual analysis method helps to select subset features for regression models and validate the quality of models. Similar to their measure of selecting features, we also use a goodness-of-fit measure. Furthermore, we allow users to explore the correlation between features by using Parallel Coordinate Plots (PCP) because a good subset of features should also avoid multicorrelation [30]. Mühlbacher et al. also provides two general partitioning methods: domain-uniform and frequency-uniform. In our framework, local pattern detection is provided through brushing data items on any dimension from the PCP. We also allow users to choose to only train on brushed data items. Thus local patterns can be indicated by the goodness-of-fit of the model.

Since we enable users to select different features and training sets, we also allow for multiple model creation and comparison. This is akin to the Delphi method [34, 38] which has multiple experts forecast and modify their prediction iteratively by comparing to other experts' predictions before finalizing their results. In general, the Delphi method is used to obtain the most reliable consensus of group opinions. Our predictive analytics framework uses the concept from the Delphi method to allow users to make their prediction after building and exploring multiple models in multiple rounds. Similar to the Delphi method, in our system the user evaluates results, where each model represents one expert or one round of the expert's prediction.

3 Framework for Predictive Social Analytics

Our framework focuses on integrating multi-source data from social media for analysis and prediction. We combine trend analysis, sentiment analysis, similarity metrics and feature selection for model building, evaluation and prediction. In order to evaluate this framework, we deploy our tools to the problem of weekend box-office prediction. We combine data from IMDB, Twitter and YouTube and explore this data across a variety of visual analytics modalities. The system was built using D3 [6], JSON, R [37] and WEKA [17]. The use of R and WEKA allowed for direct integration of multivariate regression and support vector machines, while D3 was used to create charts and graphics for the interactive visualization. A client-server architecture was chosen in order to allow easy portability and testing of the system across platforms, and we also explored the use of Amazon cloud services. We used the Jersey RESTful web service [22] to enable the interface between the web interface and backend server. Preprocessing was done for sentiment analysis and word frequency counts and nearly-interactive rates are obtained for visualizing the data described below. By nearly-interactive, we mean that if the data is cached, the visualizations can be updated at greater than 10 frames per second (FPS), if the data is not cached then the user will see a wait symbol and typically experience a 5 second lag on the first query, after which the exploration of that movie's features will be at interactive frame rates.

3.1 Data Description

In data representation and exploration, we focused on views for social media data sources, such as Twitter and Youtube. As Twitter data is unstructured and dirty, it requires a deeper preprocessing and manipulation before extracting high quality features.

Twitter: We collected Tweets for 112 movies released since 2013. Tweets are collected based on the hashtag posted by a movie's official Twitter account. In all we have 2.5 million Tweets and each Tweet includes the posting time, retweet status, user profile information and Tweet text sentiment.

Youtube: We used a rule-based script to collect Youtube data which contains the total viewcount and timestamps. We then calculate a range of features such as comment volume and interpolated view counts prior to the opening weekend. Overall we were able to collect about 7 million YouTube comments for 1104 movies.

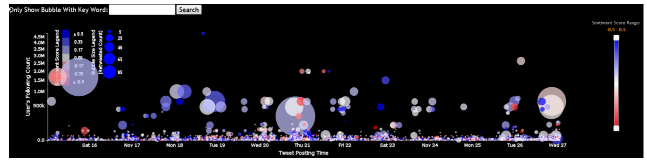
The Internet Movie Database: The Internet Movie Database (IMDB) has more than 2.8 million entries (Mar. 25, 2014) with each entry consisting of hundreds of features [1]. To deal with data noise and incompleteness, the available raw text IMDB data files were first converted into an SQL-database using JMDB [44]. The data then undergoes a data cleaning procedure.Challenges include the data sparseness and huge number of nominal values, such as cast names, which hamper machine learning. To overcome the data sparseness we calculated numeric values on a per-movie basis by aggregating gross incomes and ratings of previous movies that the cast of a new movie was involved in. Finally, we obtained a high quality movie data set of approximately 2000 movies with up to 72 features per movie.

3.2 Social Media Visual Analytics

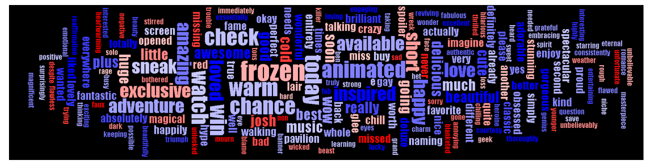
Our framework consists of a variety of views and analytical components. We provide an overview for quick trend analysis and exploration, detailed views for exploring tweet sentiment, and a similarity widget for overviews on related movies and their patterns. A core component of this framework is an iterative feature selection and model exploration module for analysis, model building and comparison.

3.2.1 Overview: Trend Analysis

When beginning analysis, users are initially presented with an overview of the data item they are trying to predict (in the case of our example, it is an overview of the movies being released in the upcoming weekend). Figure 1 shows the initial view in our web enabled system. Here, the weekend under exploration is from November 27, 2013. Figure 1(a) is a dual y-axis line chart showing the volume of Tweets and YouTube comments that have been collected relating to the movies. Users can highlight data elements by clicking on their corresponding legend entry. Key to this view is the fact that the multiple sources of data enable cross-validation. Due to the limits of the Twitter Streaming API, it is often the case that the Tweet stream will consist of missing data. However, there are many instances in which the YouTube comment traffic directly tracks that of the Twitter stream (just at different magnitudes as evidenced by the axis scales). In this manner, the analyst can quickly validate the accuracy of a source and determine what anomalies might be present.



(a) Bubble Plot



(b) Sentiment Wordle

Fig. 2: (a) A Tweet bubble plot where blue represents positive sentiment and red represents negative. The size of the bubble represents the number of times a Tweet has been retweeted, the x-axis is time, and the y-axis is the number of followers that Tweeter has. (b) A sentiment wordle where the word size represents the number of times it was used in a Tweet and color represents sentiment.

In Figure 1(b), the user can also get an overview of a baseline linear regression model prediction for that weekend. Since data for the opening box office gross has already been collected for historical weekends, the user is also shown the actual box office value. In this manner the analyst can quickly gain insight into the limitations of a proposed model. The buttons beneath the bar charts allow the user to directly navigate to a detailed view of the movie where visualizations showing word frequency, retweet count, Twitter followers and Tweet sentiment can be explored.

Figure 2 shows two of the detailed views, a temporal bubble plot and a sentiment wordle view. The bubbles in Figure 2(a) represent an individual Tweet and are colored based on the mined sentiment, where each Tweet has been processed using a dictionary based sentiment analysis, SentiWordNet [3]. This assigns each word in the Tweet to a score ranging from -1 to 1, negative to positive sentiment respectively. Each Tweet is then assigned an overall sentiment score by summing the sentiment of all words in the Tweet and then normalizing the sum. A blue color indicates positive sentiment while red indicates negative. The size of the bubble represents the number of times a Tweet was retweeted while the height on the y-axis indicates the number of followers the Twitter account has. The x-axis represents time. Similarly, all the Tweets related to a movie are converted into a wordle (Figure 2(b)), where the size of each word represents the number of times the word appears in the movie data set and the color represents the sentiment of the word. From this view, users can quickly filter for Tweets with particular keywords and they can modify the sentiment value in cases where the dictionary matching is wrong (for example, cases where the Tweet says "I want to see Frozen so bad!" will be a negative Tweet when in reality the sentiment is positive). Future work will deploy more machine learning techniques to allow for interactive

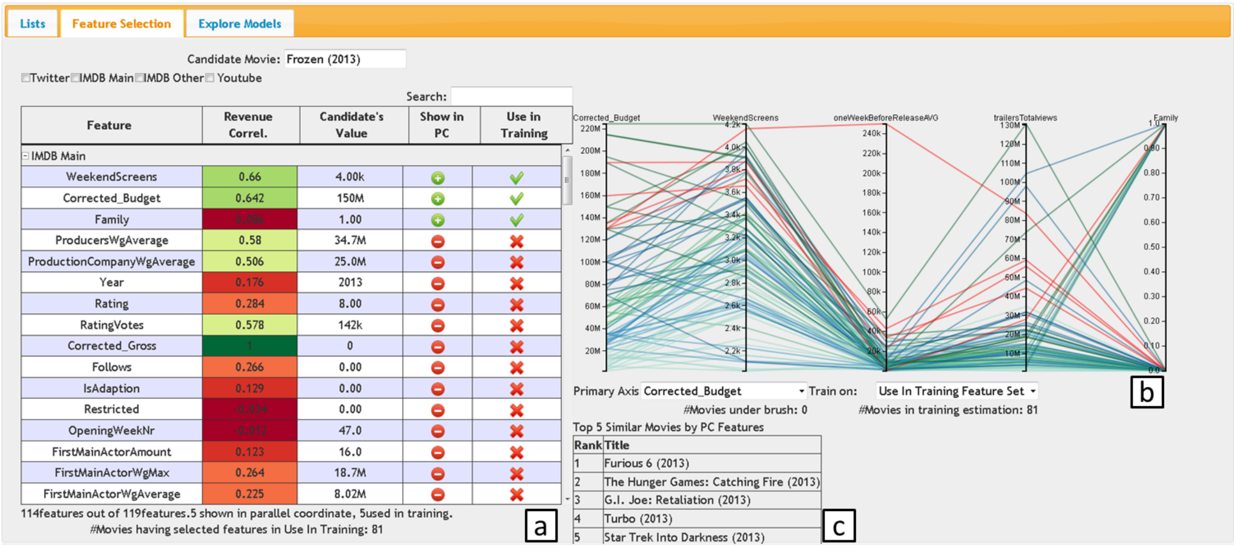


Fig. 3: Feature Selection page with Frozen as an example. View (a) is a Feature Selection table having four groups of features with their correlation to revenue mapped to a divergent color scheme. View (b) is a Parallel Coordinate view with the five most similar movies highlighted in red. View (c) lists the five most similar movies suggested by the system based on features in the PCP.

Tweet labeling for advanced sentiment classification and analysis.

3.2.2 Feature Analysis and Selection

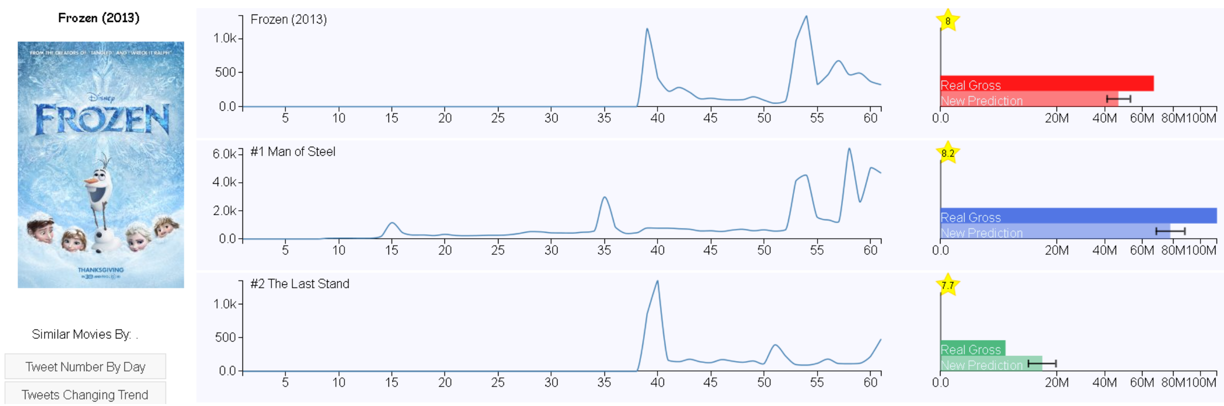
While the overview and detail visualizations enable exploratory data analysis, the key contribution of our work is the interactive modeling and prediction components. Feature values of movies can give insights and hints about their box office success. Moreover, they can be used as predictors for a movie's opening weekend revenue. Using Twitter, Youtube and IMDB data sources, we extracted four groups of features for model building with 119 features listed in the Feature Selection Table (Figure 3). Given the large number of features, it is necessary to provide the users with a suitable starting point for analysis. As such, we utilized known predictive features for movie analysis from previous work [41] (e.g., budget, number of screens the movie opens on, etc.). Thus, when the users begin their exploration process, they are presented with a baseline model to compare against. Other options would include integrating automatic feature selection as an entry point for analysis (e.g., [26, 49]).

Our goal was to augment model building by adding tools for a user to modify and explore various features. In order to quickly enable this exploration, our Feature Selection Table (Figure 3(a)) utilizes a variety of interactions and visual overlays. First, for the candidate movie being predicted (in this case Frozen), features which are not available are grayed out. Second, each of the columns in the feature selection table provides the details of a movie. The first three columns include information on the feature's name, the correlation to the revenue, and the candidate movie's value. These columns can be automatically sorted from high to low or low to high simply by clicking on the column header. The Revenue Correlation column is also color coded to directly highlight correlated features. A myriad of work has been done in feature selection [29, 35, 40] and correlation is traditionally used as one of the major factors in feature selection. A high correlation of a feature to the response variable (in our case the movie revenue) indicates that this feature could greatly impact the model. We use a green to red divergent color scale [19] where green represents a high absolute value of correlation and red represents a low value of correlation, with .5 being the midpoint value. Although correlation here is univariate (meaning we do not show correlations between multiple features) and non-linear dependencies are not taken into account, it still provides important information to users for feature detection and analysis.

The final two columns in the Feature Selection Table are associated with the Parallel Coordinate plot visualization and the model training data selection. The "Show in PC" column, when selected, will add that feature as an axis of the Parallel Coordinate Plot. The "Use in Training" column, when selected, will add all data elements that contain all of the features selected into the training set. To quickly see what features have been selected, the analyst can sort the features by clicking the column header. When features are selected, the footer information about the Feature Selection Table will update and tell the user how many features have been added to the training set, as well as the amount of movies that exist having all of these features. In this manner, the analyst can determine how many data elements can be used to train a model and they can quickly make decisions about the tradeoff between the use of more features or more training samples. For example, if a user chooses to select a Twitter feature, only 112 movies in our data set have associated Twitter data. Thus, the number of elements in the training set decreases. However, Twitter data may have a high correlation to the opening weekend gross. As such, the analyst can actually build multiple models with multiple features for training and analysis.

Another way to select the training data is through interaction with the parallel coordinate plot view. Let us consider the case in which a user has sorted the features by correlation to revenue, selected some features with higher correlation to the gross, and selected features that he/she suspects are important. These selected features can now be further explored in the PCP view (Figure 3(b)) by simply activating the "Show in PC" cell in the corresponding table row. Referring to the candidate movie's value, shown in the fourth column, the user can further filter out movies far away from this value in the PCP view. Figure 3(b) shows features of the movie "Frozen" with highly correlated features in different group and the movie's genre, "Family". Pairwise correlations between features are explored in the PCP view. For example, the WeekendScreens (the number of screens in which a movie was released during its opening weekend) and the oneWeekBeforeReleeaseAVG (the daily average number of Tweets that are related to a movie one week before its opening) variables are correlated. These axes can be dragged and dropped to explore more pairwise dimension correlations so that an analyst can choose features with low multi-correlation in order to improve the model performance. Users can then interactively select ranges on each axis to filter the data and can select an option to train the model using only the selected data.

(a) Similarity by Sentiment Wordle



(b) Similarity by Youtube Trailer Trend

Fig. 4: Similarity Widget View with Frozen. (a) is the top two most similar movies' wordle view with the Tweet sentiment wordle as the similarity criteria. Each wordle consists of the top 200 sentiment words. (b) is the top two most similar movies' line chart view with the 1 year Youtube Trailer view count trend as the similarity criteria.

The PCP view can also be used to generate insight into the data. For example, by brushing and selecting only Family movies using the Boolean genre feature "Family," one can define the training set to be only those movies that are considered to be "Family" movies. Moreover, the PCP view allows the analyst to select a primary axis, this selection defines the feature on which we base the PCP line color scheme. For example, if we color the lines based on the genre axis "Family" we can see that family movies rarely obtain a very high gross. From there, the user could train the model for only Family movies or could look for genre crossover movies such as Family and Animation.

The final item in our Feature Analysis and Selection widget is the "Top 5 Similar Movies by PCP Features" view, Figure 3(c). Given the feature vector corresponding to the features selected in the parallel coordinate plot, our system automatically calculates a Euclidean distance metric between the candidate movie and all other movies that appear in the PCP view. The five movies with the smallest Euclidean distance are then summarized in a tabular view.

3.2.3 Similarity Widget

While the Feature Analysis and Selection Tools show the top 5 most similar movies, we have also developed a series of tools for enabling users to explore temporal and sentiment similarities with regards to social media trends and specific feature similarities such as genre and ratings. Figure 4 shows our similarity widget page. Items in this similarity view focus primarily on similarity across social media (as opposed to the previous widget which used a Euclidean distance metric across many features, this view is a pairwise feature similarity). The left side of Figure 4 shows the various similarity options provided while the center view displays line charts or wordles depending on the selection. We have ten predefined metrics and one "Make Your Own Similarity" option. The rightmost area shows the model predictions and the actual weekend gross for similar movies via a bar graph.

This widget enables analysts to quickly find and compare the accuracy of predictions based on various criteria of similarity, and to perceive if the given prediction model typically underestimates, overestimates or is relatively accurate with regards to movies that the analyst deems to be similar. In this manner, a user can further refine their final prediction value. In this work, we have defined ten similarity criteria with distance calculation methods focusing on matching temporal trends through sequential normalization or Euclidean distance metrics for magnitude comparisons. In all similarity matches, we show the top five most similar movies. These views allow users to directly compare Tweet trends and sentiment words between movies deemed to be similar in a category. Figure 4 contains snapshots from Frozen's similarity page cropped to the top two most similar movies by Sentiment Wordle and Youtube Trailer Comments.

Though similarity metrics used in this page are not directly transformed into modeling features, by providing an analyst with insight into these secondary variables, coupled with the model performance with similar movies included in the training set, further refinement of the prediction is made possible. For example, an analyst may compare the absolute difference between Tweets/Youtube comments of two movies, or they can inspect the trend of the Tweets through line chart comparison using the Tweets Changing Trend similarity metric. This tool also allows users to quickly compare the current movies un-

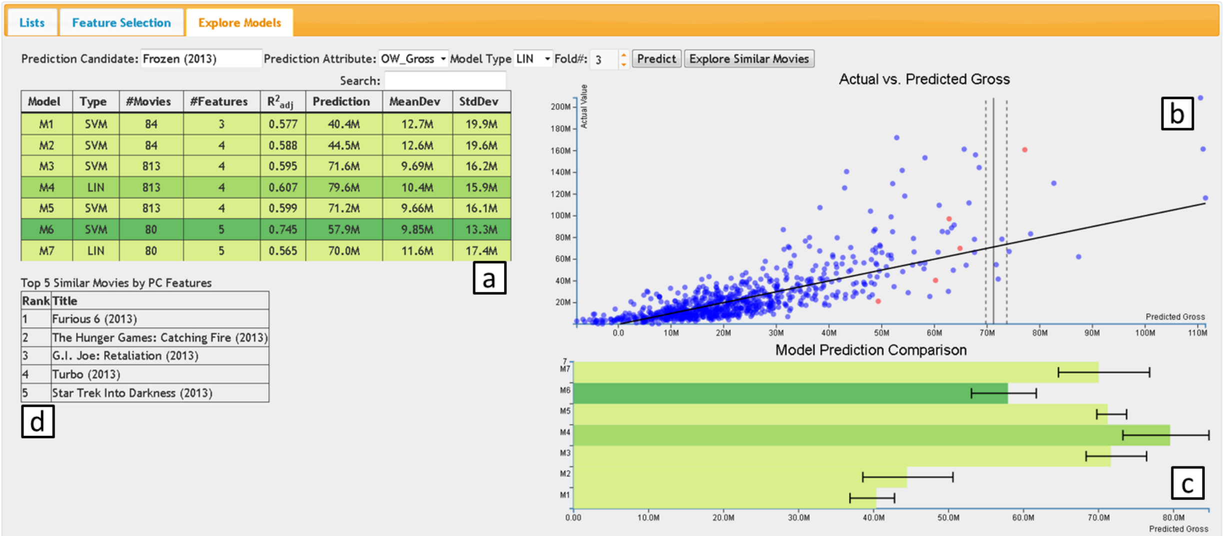


Fig. 5: Multiple Method Modeling with Frozen as the candidate movie. View (a) is a trackable Model History Table recording each model the user built. View (b) is the scatterplot of the Actual vs. Predicted Gross showing a model's prediction for each movie in the training set and the prediction result together with a stable range for the candidate movie. View (c) is the bar graph of the Model Prediction Comparison having each model's prediction stacked. View (d) lists the five most similar movies as was done in View (c) of the Feature Selection page.

der analysis to recently released movies with the same MPAA rating and genre. When the user builds a model involving Twitter features, the top 5 most similar movies listed in the Feature Selection and the Explore Models page can be compared in the similarity page.

3.3 Model Building, Analysis and Verification

Based on recent literature and the general use of prediction models, we support the creation of three different types of models: Support Vector Machine (SVM) [11], Linear Regression (LIN) [30] and Multilayer Perceptron (MLP) [20]. Using the linear regression model with the budget and the average number of daily Tweet (TBD)s for a movie as regressors and the opening weekend gross as response, the system provides users with a baseline prediction result together with a 95% confidence interval for each movie. The baseline model results are shown in both the front page (see Figure 1(b)) and the similarity page's right-hand bar graphs (Figure 4).

Besides exploring the baseline model, the user can build a more complex model, bringing in domain knowledge and analytic insights. For instance, the user is allowed to interactively set up parameters and build models with different feature sets, training instances (movies) and model types. We use several error measures to give the analyst feedback about the quality of fit and the prediction stability. By using the interactive Feature Selection and Explore Models pages, the user can iteratively change the features, training sets and model types to improve a model's quality. We measure the model's accuracy using the adjusted R^2, denoted R^2_{adj}. Using R^2_{adj} has the following advantage: R^2 never decreases when a regressor (feature) is added to the model, regardless of the value of the contribution of that variable; however, R^2_{adj} will only increase when adding a variable to the model if the addition of the variable reduces the residual mean square. Otherwise R^2_{adj} decreases when adding terms that are not helpful [30]. With a feature set of size p and a number of instances (movies) n, R^2_{adj} is defined as:

$$R^2_{adj} = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)} \tag{1}$$

where SS_{Res} is the sum of squares of the residual, and SS_T is the sum of squares of total.

3.3.1 Base Line Model

We used the model proposed in our VAST Boxoffice Challenge 2013 submission [27] as our base line model, which is described as follows:

$$OW = \beta_0 + \beta_1 TBD + \beta_2 Budget + \varepsilon \tag{2}$$

With all 110 movies in the training set, the estimation of parameters in Equation 2 are $OW = 6.878 \times 10^6 + 1303 \times TBD + 0.26 \times Budget$ with $R^2_{adj} \approx 0.6$ and $P \ll 0.05$.

3.3.2 Advanced Models

As most of the attributes are proportional to the box office success (e.g. the more budget, the higher weekend gross potential) we can even achieve good results using linear regression model. More advanced models can be built using a Support Vector Machine (SVM) or a Neural Network, i.e. Multilayer Perceptron (MLP). To achieve good results, these algorithms have to be finely configured by setting input parameters based on the input data. We ran a grid search (parameter optimization method) to find out the best parameter settings. For SVM we use a linear kernel and a nu-parameter of 0.4, which constrains the influence of a single instance (movie) to the model. Considering the relatively small number of movies when compared to the large feature space we also tested an RBF kernel. However this did not achieve better R^2_{adj} results than with the linear kernel. For MLP we use the backpropagation learning rule and use a learning rate of 0.3, 200 training epochs and a momentum rate of 0.85 to achieve good results.

3.3.3 Multiple Methods Modeling

Predictive models help to reveal relationships between the predictors and the response variable, but no matter how good the prediction is, no cause-effect relationship can be implied. Also, the accuracy of one prediction can hardly be generalized to all other predictions. In statistical analysis, experts usually explore residual distributions, outliers, influential points, and model stability. In our system, besides using statistical methods, we apply visual analysis methods for exploring the residual distribution.

In the page "Explore Models" the user can select which algorithm to use, set the number of folds for the stability test, train models to predict the movie's revenue, and compare between models. The Explore Models view is shown in Figure 5. For model building, the feature and training set configurations from the Feature Selection page are applied. After the prediction is executed, the analyst can use the Actual vs. Predicted Gross view (Figure 5(b)) to obtain an overview of the residuals, as was presented in [24]. A diagonal referential line indicating "the perfect prediction" is also drawn. This means, the closer the data points lie to the referential line, the better the overall fit of the model. The top 5 most similar movies are highlighted in red to quickly guide comparison and analysis. The user can change these similar movies based on adding/removing features in the Parallel Coordinates view (Feature Selection page). To submit a good prediction for a particular movie, it may be more important that the model fits for similar movies than fits the overall training set. In other words, if the model predicts well for similar movies this may be an indicator that it also gives good results for the prediction candidate.

Our tools also enable the exploration of influential points. An influential point is an outlier in both the predictor and the response domain, and these points are known to have a noticeable impact on the model coefficients [30]. If an influential point is removed from the training set, the fit of the model will change by a relatively large degree and usually fit other points better. This fact can be used to improve prediction results. Instead of using statistic diagnostics, such as Cook's D and DFFITS [4], we allow the user to directly remove such instances and only train on selected movies. In this way, influential points can be implicitly removed via exploring differences between different models.

Finally, the Model History Table (Figure 5(a)) enables the comparison of multiple models so that the analyst can review the predictions by re-investigating their scatterplots. In combination with the Model Comparison view (Figure 5(c)), the user can also get an overview of the prediction deviations, review the increase or decrease of prediction precision and select his/her final prediction. Our goal is to build a model which can help the analyst to better predict the upcoming movie's opening weekend gross, not to build an adequate model that fits all the training data very well.

To estimate the performance and to test the model's stability, we provide an n-fold cross-validation [16, 23]. For the cross-validation we partition the data into n folds. Each fold includes num_{movies}/n instances. The movies of each fold are predicted once, using the other folds for training. This way, we ensure that the model generalizes and is not overfit to the training data. For the prediction candidate, every fold is used once to predict the outcome. Thus, for each prediction we get n results. The dashed vertical line in the scatterplot shows the range of these results. A smaller range indicates that the model is stable. This range is also shown in the bar graph below the scatter plot, where all predictions can be compared.

3.3.4 Auxiliary Analysis

Instead of depending totally on an automatic model, most industry predictions also utilize an expert's domain knowledge. For example, if a movie is released next to an expected blockbuster, its performance could be also impacted. With our system, analysts can query any movie by its title to investigate features. Users can also go to previous weekends to see how much money those movies made. A user can also investigate the Twitter and Youtube data to explore the advertising campaign and public sentiment. Usually a successful movie has either an effective advertisement campaign, positive public reactions, or both. From the bubble plot shown in Figure 2(a), large bubbles usually are Tweets from the movie production company and the bubble size indicates the spread power. If the large bubbles separate along the time line, it is likely that the company has continued advertising its movie.

4 Case Study: Predicting Disney's Frozen

This section demonstrates how an analyst would use our system to predict Frozen's opening weekend gross. This process consists of multiple steps, which can be iteratively traversed in different ways. However, we suggest the following procedure. First, the user gets an overview of the Twitter and Youtube comments using the dual-y-axis line chart to compare movies released together. Second, details can be investigated using the detail pages of the candidate movie. Third, the user can explore similar movies and compare their gross, as well as how well the baseline model performed for them. After having a general impression of the expected revenue, the user can navigate to the Feature Selection tab to explore and select features or filter movies to create a model. Finally the user can build and explore different models and their prediction ranges in the Explore Models view. Step 4 and step 5 can be iteratively applied until the user feels they can make a confident prediction.

To illustrate these 5 steps, we will take Frozen as an example. Starting on the overview page, the line chart in Figure 1 (a) indicates that there are 4 movies released on the same weekend (Frozen, Black Nativity, Homefront, and Oldboy).We quickly see that online chatter (Tweet and YouTube comment volume) about Frozen is not dominating the other weekend movies, in fact it is trending similarly to the movie Black Nativity. This phenomenon indicates that it is unlikely that Frozen will obtain an anomalously large gross as the market will be shared by competitors.

In the second step, using the detailed view of Frozen (see Figure 2) the Tweet sentiment is analyzed. One can see frequent Tweet keywords and the sentiment polarity. Also, the retweet volumes provides information about users' interest in the movie and the advertisement campaign. For example in Figure 2 we can see that Frozen does not have a large Tweet and retweet volume compared to other blockbusters; however it does have a very positive sentiment (blueish dots). The movie sentiment score for Frozen is approximately 0.8 which is very high among all 112 movies having Twitter data.

In the third step the similarity widget is explored (see Figure 4). This reveals that movies similar to Frozen were under-predicted with the baseline model, which predicts about $44M for Frozen. The fourth step focuses on the analysis and selection of the movies features (see Figure 3). There are two main views for feature selection: the correlation view showing relationships between a feature and the revenue, and; the relationship among features depicted in the PCP view. From our baseline model we select the number of opening screens, the budget and the weekly average of Tweet counts as an initial feature selection. This gives us a model with $R^2_{adj} \approx 0.58$ (M1 in Figure 5). To further improve the model, we add another feature, view counts of the movie's YouTube trailers, and built both an SVM and LIN model. R^2_{adj} improved to approximately 0.6 while the prediction deviations from the different folds decreased. Next, using our background knowledge, we explore the genre of this movie (in this case the genre is "Family"). While adding the Family feature to the Parallel Coordinates, we find that the gross distribution for Family movies is significantly different to most non-Family genres. Thus, for our last prediction iteration, we add the family feature to the model. We obtained an R^2_{adj} score of 0.745. Finally, we review the Model Prediction Comparison graph and decided to finalize our prediction between $60M to $70M based on the best performing models.

5 Evaluation

In order to evaluate the effectiveness of this framework for predictive analytics, we performed a user study. On March 20th, 2014 we enlisted seven graduate students from China, India, the United States and Germany and asked them to predict the results of four different movies. The first two movies predicted were to provide them with baseline training, the next two movies were to be released on March 21st, thus having them do an actual future prediction. The movies we had them predict included Disney's Frozen (2013) and The Hunger Games: Catching Fire (2013) (which were used for training) and Divergent (March 21, 2014) and Muppets Most Wanted (March 21, 2014) (which were the movies to be predicted). For Frozen and the Hunger Games, their weekend box office data was removed for the training exercise in order to simulate the prediction process.

Of the seven participants, six were male, one was female and all were PhD students. Prior to participation, we surveyed them about their cinema affinity and data visualization knowledge on a scale from 1-5 (with 1 being the lowest). From the seven participants four claimed to be visualization experts. Five subjects rated their movie affinity

Table 1: Results for Frozen and Hunger Games. The opening weekend gross for Frozen is $67M and for the Hunger Games it is $158M.

subject	user1	user2	user3	user4	user5	user6	user7	BoxOffice.com	BoxofficeMojo
Prediction(Frozen)	55.9	59	50	60	57.7	62.5	58	47	44.7
Abs Error	11.1	8	17	7	9.3	4.5	9	20	22.3
Prediction(Hunger Games)	71.1	135	NA	100	95.9	86	75	166	167
Abs Error	86.9	23	NA	58	62.1	72	83	8	9

Table 2: Results for Divergent and Muppets. The opening weekend gross for Divergent is $56M and for the Muppets it is $16.5M.

subject	user1	user2	user3	user4	user5	user6	user7	BoxOffice.com	BoxofficeMojo
Prediction(Divergent)	54.1	53	40	50	30.1	47.5	48	66	51
Abs Error	1.9	3	16	6	25.9	8.5	8	10	5
Prediction(Muppets)	50.6	21.5	28	15	35	21.4	20	25	22
Abs Error	34.1	5	11.5	1.5	18.5	4.9	3.5	8.5	5.5

as low (1-2), and two rated medium (3-4). Their machine learning knowledge was mostly low, with only two participants claiming a basic knowledge of machine learning and prediction related tasks (these students had all taken regression analysis and/or data mining courses, as such we feel that they can be considered to have a relatively high level of expertise in the modeling and analysis process). The two subjects that rated their movie affinity as low were those that rated their machine learning and predictive analytics knowledge as high. Thus, we have three subjects that were casual users with limited domain knowledge and limited analytics experience, two subjects that had some domain knowledge and limited analytics experience, and two subjects that had expertise in data mining and predictive analytics but limited domain knowledge.

To introduce the system, we walked through an example analysis of the movie After Earth and explained our proposed analytics process (similar to the case study in section 4). Subjects were then asked to predict Frozen and The Hunger Games. During the analysis and prediction process of these two movies, they were open to ask any questions, such as the meaning of a feature, how to use a special function of the system, and what information could help to choose proper features and improve the model performance. After they submitted their final prediction about a movie, we told them the real gross so that they could make a comparison and adjust their strategy for the next movie. After practicing with these two movies, they used the system (unaided) to predict the new movies Divergent and Muppets Most Wanted.

To get a deeper understanding of the users analysis processes this study was carried out as talk-aloud [14] session. The users were asked to speak their thoughts out loud explaining their actions. We recorded voice and system interaction by video. After the study we summarized the key results and classified them into System Usability, Social Media Exploration, Feature Selection and Model Comparison.

5.1 System Usability

Key findings here indicated more details on system design. All subjects reported ease of use and interaction with the system. Furthermore, the length of the user study demonstrated the subjects' engagement. No instructions were given on the time needed to make a prediction; however, subjects spent over 1 hour on average tuning system parameters and exploring the data. Subjects also were excited to compare their results Monday and indicated they wanted to try this again. Design issues they faced were that they wanted even more transparency in the data. As no subject was a self-rated expert in cinema (most indicating they had seen less than two movies in the past 6 months) many of the subjects wanted more information about the movie features. They suggested direct links to the IMDB pages for the movies to allow even greater detail views. Overall, the most used views were the similarity page and the feature selection page.

Subjects all started their analysis on the overview page, exploring time series trends and comparing how they felt the movies on the weekend would fare when compared to others. They typically looked at the Twitter and YouTube volumes and sentiment data. At the beginning they found it difficult to interpret those visualizations as they were unfamiliar to a user; however, by the end of the study the users were requesting more features, wanting to create difference maps of the movies to look for keyword differences in the sentiment analysis and also to identify what was being discussed differently between YouTube and Twitter. As such, it is clear that more text analysis is needed for further insight generation. A clear example of gaining insight was shown during the analysis of the movie, Divergent. No subject in our group had heard of this movie; however, when inspecting the data they saw that The Hunger Games was often referred to in context with this movie. This grounding gave them the contextual clues which they needed in order to analyze Divergent.

Negative comments focused on the disconnect between the similar movies and the users' thought process. In the Feature Selection page, users are presented with the five most similar movies with respect to the selected PCP features. This is calculated as a Euclidean distance metric, and the calculation is a black-box to the user. As such, analysts were often wary of these movies and preferred to use the "create your own similarity" option on the similarity widget page. However, this again required more domain knowledge than some users had, with many again requesting details about what genre, rating, etc. a particular movie had. Future work should include better views for multidimensional similarity matches and more transparency in the similarity metrics. Yet, what the process highlights is that all subjects, even those with little self-proclaimed movie knowledge, are able to bring some background knowledge into the prediction process, which could be used to add value when compared to a purely automated prediction process.

5.2 Feature Selection

All users worked with the Feature Selection table to determine which data was available for a movie and remarked on how they felt the prediction was more reliable when they knew that the data existed. Again, this indicates that transparency in the model training can improve an analyst's confidence. During the feature selection process, most users started with the baseline settings, inspected the results and then iteratively chose more features with high correlations, reinspected and then iterated again. Other users again applied their domain expertise and chose features that seemed interesting to them. For example, the user that had seen 10 movies in the theater in the past six months used his domain knowledge to select features which are not obviously highly correlated to the revenue but these features considerably improved that subject's model.

Participants who decided to add Twitter related features typically based this choice on the genre of the movie, stating that Twitter users would be interested in Divergent but not in the Muppets. One user, with a basic background knowledge in prediction tasks commented on how the Parallel Coordinate view enabled her to choose features that were independent (i.e., not multi-correlated). Other users engaged the PCP view to filter out movies to create models based on genre or movie ratings. Overall, they spent a large amount of time exploring features and discussing what they felt these features meant. They also found it extremely helpful to see how the selection of different features impacted the amount of movies available for training.

Negative comments revolved around users' frustration in feature selection, noting that there should be a way to provide more details on what is likely to be a good feature. For the inspection of correlations, one user noted that it was hard to use the PCP view and had difficulty distinguishing the highlights. However, the users all liked the design of the framework, and commented on how it would be useful to change the domain to look at other specific problems of interest. For future work, we plan to explore how to improve the presentation of features. Obviously showing all features (in this case 116) is a huge amount of information overload; however, we also want to involve the user and allow him/her to use domain knowledge to guide the modeling and prediction process. We plan to explore several methods of automatic feature selection as a means of organizing information for visual presentation and exploration and performing user studies across various feature set visualizations in order to explore this area.

5.3 Model Comparison

As for the Feature Selection view, participants found the model comparison features extremely useful. Starting with some initial predictions, they tried to improve the model to reduce the errors. Users often focused on prequel movies (particularly during the Hunger Games prediction) and focused on developing a model that was a good fit for known prequels or known movies within a genre. One user repeated the feature inspection, selection and modeling until he was able to create a model that strongly fit to the prequel (in the case of the Hunger Games). Others tried to inspect all outliers and then made decisions based on their domain expertise regarding movie similarity. This would lead to an iterative model building and refinement loop. Users also inspected the scatterplot and would then access the similarity comparison tools to explore the impact of Twitter on the model prediction. Users noted that Twitter seemed to have an impact depending on the type of movies, and many came to the conclusion that Twitter was relevant when predicting Science Fiction movies (such as Divergent) but less relevant when predicting Family movies (such as the Muppets). Again, subjects indicated a desire for even further transparency of the inner workings of the model prediction.

5.4 Prediction Results

Table 1 and 2 show the results of our user study in both the training trial and the actual prediction trial. For the training results (Table 1), subjects were found to have a lower error than that of the experts for Frozen; however, for the Hunger Games, subjects found this very difficult to model. It is important to note that we went through the example of After Earth, Frozen and the Hunger Games for training in order to give subjects examples of a low outlier, a good fit, and a high outlier respectively. In this way they can explore all possible scenarios prior to the actual prediction task.

For the actual results (Table 2), 5 of our 7 subjects were able to best BoxOffice.com predictions for Divergent and 2 of our 7 subjects were able to best both expert prediction websites. Only two subjects erred on the far low end of the spectrum for this movie (subjects 3 and 5). For the Muppets, 4 of our 7 subjects were able to best the experts, with one subject (subject 4) accurately predicting this would be a box office failure. Again, subject 5 was an outlier, and subject 1 predicted that the Muppets would be an outlier on the positive end of the spectrum.

Overall, the results of our study are quite positive. Given our subjects self-reported lack of movie knowledge, it is clear that the integration of social media and visual analytics for model building and prediction can quickly generate insight at a near professional prediction level. Subjects 2 and 7 had the highest self-reported domain knowledge and (as seen in Table 2) outperformed experts from BoxOffice.com (and Subject 2 outperformed the BoxofficeMojo results as well). The machine learning and regression experts were subjects 4 and 6 and they also outperformed the experts. The remaining subjects can all be considered more casual users and had a higher variability. In both future prediction cases, over half the subjects were able to best the experts over the course of a one hour training session. Furthermore, such work indicates that visual analytics can have a direct impact on the modeling and prediction process. As noted by Lazer et al. [25], there is a need for tools that can improve insight into large data analytics and an increased transparency can potentially lead to improved model efficacy. Future work will look at doing a more formal evaluation where a larger subject pool is recruited and more analysis between the three groups is performed.

6 Conclusions

This paper presents an interactive framework integrating social media and predictive analytics, and the presentation of a talk aloud study that discusses design successes, pitfalls and potential future directions. Analysts can utilize the system to explore and combine information, and underlying mechanisms for similarity matching and data filtering can help a user quickly engage in exploratory data analysis as part of the model building process. We allow for the quick integration of structured and unstructured data sources, focusing on box office predictions as our example domain. In comparison to state-of-the-art in visual analytics, we have worked towards improving a user's understanding of the modeling and training process. Our results were validated through case studies and user studies. We have demonstrated that such a tool can quickly enable non-domain experts to be competitive with domain experts in a given area. This seems to stem from a combination of a user's (in our case limited) domain knowledge with the interactive visualization interface. While with our system semi-professionals are not always able to beat the expert models from boxoffice.com and boxofficemojo, respectable results were obtained across a majority of users. As the industry's models and predictive practices are not available, it is difficult to comment on their workflow. However, talking with experts from SAS and JMP, they recognize a need for integrating more interactive visuals in the model building process. Overall, we believe that such a framework could be applied to a wide range of social media data in which analysts want to locally extract information from social media and use trend values and other metrics as input to their modeling process. We believe that predictive analytics in general can be improved upon by integrating human knowledge into the workflow and can add more transparency to the oftentimes black-box model that encompasses many of the current prediction methods (e.g., SVM).

7 Future work

For future work we want to improve a users understanding of a feature's impact on a model. We also want to develop methods to explore and select features according to multivariate dependencies and feature engineering. Visualization can explain results and reveal complex dependencies. To find such dependencies we want to integrate and orchestrate even more data sources, such as news media and other social media sources like bitly and Facebook, as well as weather and seasonal information such as holidays. Moreover we expect dependencies between past and concurrent weekend releases to be highly important. We also want to focus on the machine learning aspect of prediction. As our models makes structure assumptions, for example, the linear regression model only covered linear relationships, we think we can further improve predictions by investigating the domain data more deeply and use these insights to help analysts choose the right algorithms and options.

8 Acknowledgments

This work was supported by the U.S. Department of Homeland Securitys VACCINE Center under Award Number 2009-ST-061-CI0001. We also wish to acknowledge to the cooperative graduate program Digital Media of the University of Stuttgart, the University of Tuebingen, and the Stuttgart Media University (HdM), as well as the German Federal Ministry of Education and Research (BMBF) in context of the VASA project and the Stuttgart Vast Challenge Team 2013.

REFERENCES

[1] IMDb Database Statistics, 2014. http://www.imdb.com/stats.

[2] R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *IEEE Symposium on Information Visualization*, pages 143–150. IEEE, 2004.

[3] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation*, 2010.

[4] D. A. Belsley, E. Kuh, and R. E. Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons, 2005.

[5] W. Berger, H. Piringer, P. Filzmoser, and E. Gröller. Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction. In *Computer Graphics Forum*, volume 30, pages 911–920. Wiley Online Library, 2011.

[6] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec. 2011.

[7] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *IEEE Conference on Visual Analytics Science and Technology*, pages 83–92. IEEE, 2012.

[8] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, 2012.

[9] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE Conference on Visual Analytics Science and Technology*, pages 143–152. IEEE, 2012.

[10] J. Choo, H. Lee, J. Kihm, and H. Park. ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 27–34. IEEE, 2010.

[11] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[12] A. Culotta. Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the First Workshop on Social Media Analytics*, pages 115–122, New York, NY, USA, 2010. ACM.

[13] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, 2013.

[14] K. Ericsson and H. Simon. *Protocol Analysis: Verbal Reports as Data*. MIT Press, Cambridge, MA, 1993.

[15] G. Eysenbach. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, 13(4), 2011.

[16] S. Geisser. *Predictive inference*. Chapman & Hall, New York, 1993.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations Newsletter*, 11(1):10–18, Nov. 2009.

[18] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Keim, L. Haug, and M.-C. Hsu. Visual sentiment analysis on twitter data streams. In *IEEE Conference on Visual Analytics Science and Technology*, pages 277–278. IEEE, 2011.

[19] M. Harrower and C. A. Brewer. Colorbrewer. org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[20] S. Haykin. *Neural networks: A comprehensive foundation*. Prentice Hall PTR, 1994.

[21] A. Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2009.

[22] Jersey Team. Jersey: Restful web services in java. `https://jersey.java.net/`, 2014.

[23] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1137–1143. Morgan Kaufmann, 1995.

[24] R. Krüger, H. Bosch, D. Thom, E. Püttmann, Q. Han, S. Koch, F. Heimerl, and T. Ertl. Prolix - visual prediction analysis for box office success. In *IEEE Conference on Visual Analytics Science and Technology*, 2013.

[25] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of google flu: Traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.

[26] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502, April 2005.

[27] Y. Lu, F. Wang, and R. Maciejewski. VAST 2013 Mini-Challenge 1: Box Office VAST-Team VADER. In *IEEE Conference on Visual Analytics Science and Technology*, 2013.

[28] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *IEEE Conference on Visual Analytics Science and Technology*, pages 181–190. IEEE, 2011.

[29] T. May, A. Bannach, J. Davey, T. Ruppert, and J. Kohlhammer. Guiding feature subset selection with an interactive visualization. In *IEEE Conference on Visual Analytics Science and Technology*, pages 111–120. IEEE, 2011.

[30] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. Wiley, 2012.

[31] F. Morstatter, S. Kumar, H. Liu, and R. Maciejewski. Understanding twitter data with tweetxplorer. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1482–1485. ACM, 2013.

[32] T. Muhlbacher and H. Piringer. A partition-based framework for building and validating regression models. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):1962–1971, 2013.

[33] V. D. Nguyen, B. Varghese, and A. Barker. The royal birth of 2013: Analysing and visualising public sentiment in the UK using Twitter. In *IEEE International Conference on Big Data*, pages 46–54. IEEE, 2013.

[34] C. Okoli and S. D. Pawlowski. The delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1):15–29, 2004.

[35] H. Piringer, W. Berger, and H. Hauser. Quantifying and comparing features in high-dimensional datasets. In *12th International Conference on Information Visualisation*, pages 240–245. IEEE, 2008.

[36] S. Publishing et al. *JMP 10 Modeling and Multivariate Methods*. SAS Institute, 2012.

[37] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[38] G. Rowe and G. Wright. The delphi technique as a forecasting tool: issues and analysis. *International journal of forecasting*, 15(4):353–375, 1999.

[39] SAS Institute Inc. *SAS/STAT Software, Version 9.3*. Cary, NC, 2011.

[40] J. Seo and B. Shneiderman. A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In *IEEE Symposium on Information Visualization*, pages 65–72. IEEE, 2004.

[41] J. S. Simonoff and I. R. Sparrow. Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance*, 13(3):15–24, 2000.

[42] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 41–48. IEEE, 2012.

[43] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *International AAAI Conference on Weblogs and Social Media*, 10:178–185, 2010.

[44] J. Ulbts. JMDB: Java Movie Database. `http://www.jmdb.de/`, 2014.

[45] C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang. Sentiview: Sentiment analysis and visualization for internet popular topics. *IEEE Transactions on Human-Machine Systems*, 43(6):620–630, 2013.

[46] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. J. Zhu, and H. Qu. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2012–2021, 2013.

[47] D. Zeng, H. Chen, R. Lusch, and S.-H. Li. Social media analytics and intelligence. *Intelligent Systems, IEEE*, 25(6):13–16, 2010.

[48] X. Zhang, H. Fuehres, and P. A. Gloor. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, 26:55–62, 2011.

[49] Z. Zhao, L. Wang, and H. Liu. Efficient spectral feature selection with minimum redundancy. In *AAAI Conference on Artificial Intelligence*, 2010.