The Same Thing – Only Different: Classification of Movies by their Story Types

Chang Liu Armin Shmilovici Mark Last

Ben-Gurion University of the Negev (Israel)
P.O.B. 653 Beer-Sheva, Israel
liuc@post.bgu.ac.il {armin, mlast}@bgu.ac.il

Abstract

Story types depict the development of movie stories in terms of the protagonist's character traits and the motivations that drive him in facing his challenges. We define a novel task of story type classification of movies and propose a lightweight machine learning solution. A crowdsourcing experiment was performed to label 45 movies for their perceived story types. We extract movie features that indicate different aspects of the movie characters and apply Decision Tree and Naive Bayes classifiers as classification algorithms. Although the labeled dataset is relatively small, the story type classification accuracy is significantly above the baseline with the F1 measure in the range of [0.63-0.77]. The preliminary results suggest that simple movie features can be used by machine learning algorithms to detect the abstract concepts of story types.

Keywords: *Story Analytics; Computational Narrative; Movie Under-standing; Story Type Classification*

1. Introduction

Background. While storytelling is a form of human artistic expression, we conceive of story writing and, in particular, script-writing for TV and movies as being built upon a few fairly known generative principles – a structure. Literature researchers have identified structural similarities between different stories. They claim that most stories can be attributed to a fairly *small* set of *unique* plots [3,12] about a few *archetypal* characters [28]. The Hollywood cliché is of the studio executive telling the script-writer: "Give me the same thing... only different!"

We chose to build our research upon the 10 story types of [9] due to their detailed and clear definitions with plenty of examples. Revealing the high-level structure of movies (i.e., the story type and its main elements) is a major aspect of understanding the movie plots. Humans can identify and understand most of the elements of a story, such as the characters and their motivations, events and their consequences etc., and can categorize the story into one of the story types. However, current movie analytics technologies are able to detect only relatively primitive story elements, such as the human actors and some low-level actions [13, 22]. Researchers in the computational narrative understanding community have recently made progress in understanding narratives in text (books and movie scripts) [25, 24, 14, 36, 2]. To the best of our knowledge, we are the first to study story type classification for movies. Another major and common challenge faced by narrative understanding community is the collection of large-scale, reliable labeled datasets [1, 19,25], especially in the movies domain.

Our first objective is to provide a labeled dataset of movies [10] to facilitate the use of supervised machine learning algorithms for the problem of story types classification. A crowd-sourcing experiment is used for constructing the dataset. The collected labels are analyzed to verify the following two hypotheses: (1) most movies adhere fairly well to a general structure, described by the screenwriting book [9] and (2) even non-experts can identify those story types after watching a movie.

Our second objective is to provide a lightweight solution for the challenging task of story type classification, with the use of relatively simple methodology and features.

The original contributions of our paper to the domain of computational narrative understanding in movies are two-fold: a) We introduce the first benchmark dataset for the problem of story type classification that will be released to the research community; and b), We demonstrate that the story type of a movie can be automatically detected using some relatively simple movie features.

Typology. Following, we describe the 10 story types defined by [9]. The most important point to mention is that there are three abstract elements characterizing each story type, and those core elements can be used to distinguish between the different story types. From the story type we can infer the existence of some abstract concepts in the movie plot, such as a monster character or a detective character.

1. **Monster in the House:** A hero is forced to save a trapped group of people from being killed by a monster he inadvertently unleashed. Examples: Jurassic World, Jaws.
2. **Golden Fleece:** A driven hero must lead a group of

allies to retrieve a prized possession through a perilous journey that was not what the hero expected. Examples: Avengers, Infinity War, Back to the Future.

3. **Out of the Bottle:** A greedy hero must learn to undo a spell he initiated before it turns into a curse he cannot undo. Examples: Liar Liar, Big.

4. **Dude with a Problem:** An unsuspecting hero must survive at all costs when he is dragged into a life or death situation he did not see coming and cannot escape. Examples: 1917, the Martian.

5. **Rites of Passage:** A troubled hero's only way to overcome a growing life crisis is to defeat his worst enemy – himself. Examples: Brooklyn, Inside Out.

6. **Buddy Love:** An inadequate hero must rise above an extremely difficult situation to be with a uniquely unlikely partner who is the only one capable of bringing him peace. Examples: E.T., Zootopia.

7. **Whydunit:** A devoted hero must find the truth behind an intriguing mystery before he is swallowed by the darkness he desperately seeks to expose. Examples: Captain Marvel, Bladerunner, The Silence of the Lambs.

8. **The Fool Triumphant:** An innocent hero whose only way to defeat the prejudices of a group is to change himself without losing what made him the group's target of contempt in the first place – his uniqueness. Example: Moneyball.

9. **Institutionalized:** An outsider whose only way to save his individuality is by going against the many who wish to make him like them. Examples: American Sniper.

10. **Superhero:** A uniquely special hero must defeat an opponent with stronger capabilities by using the same powers that disconnect him from the people he hopes to save. Examples: Iron Man, Taken.

For example, the movie *Die Hard* belongs to the *Dude with a Problem* type, where an "innocent hero" (a bullheaded policeman) visiting his wife is accidentally involved in a "sudden event" (terrorists crash an office party and take his wife as a hostage) and he has to use his brains rather than his might in a "life or death battle" to subdue the high-tech terrorists and save his wife. Sometime a story-type is also written for the antagonist character to make it more realistic. For example, from the terrorists' point of view, the movie *Die Hard* belongs to the *Monster in the House* type: in a *secluded* office building, their *sin* of greed has awakened a mighty and vengeful *monster* (the hidden policeman) who kills them one by one until they figure out his weakness (his hostage wife) and use her to set up a trap before a final direct confrontation.

Some of the story types are typically used for producing some of the "standard" movie genres. For example, most horror movies are of the *Monster in the House* type; most detective movies are of the *Whydunit* type; and most romantic comedies are of the *Buddy Love* type. The same story can be told using different story types and genres. For example, a superficial story about a fictitious British secret agent who fights an evil secret organization that threatens the safety of the world can be told with different story type twists for producing a) action movies: in the *James Bond* movies (typically of the *Superhero* type) the character has superhero-like fighting skills; or b) comedy movies: the *Johnny English* movies are of the *Fool Triumphant* type. The character is innocent about his clumsiness and ridiculed by the establishment in which he operates.

2. Dataset Preparation

Movie Selection. We limited our movie selection to the 51 movies used in the *MovieGraphs* dataset [15,27]. This dataset contains 7,637 manually calibrated scene boundaries. The description of each scene contains detailed manually provided information such as the characters' appearance, the characters' relations (e.g., parents), character interactions (e.g., greeting), characters' emotion etc. The reason we chose to use this dataset is that it is rather *complete* in terms of the available information that is crucial for narrative understanding. We had to discard 6 movies, either because we could not obtain the exact version of the movies used in the *MovieGraphs* dataset, or because of an irrelevant movie genre from point of view of the plot (e.g. Biographical movie), and eventually labeled 45 movies.

Story-Type Annotation. For the story type labeling, we selected 119 human annotators (out of 180 applicants, all senior undergraduate students in the Software and Information Systems Engineering Department), based on their English proficiency level and their level of interest in watching movies. During the annotation process, we ensured that: (1) Each annotator was assigned at least 3 movies (including one of the 9 movies with "gold-standard types"); and (2) Each movie was annotated by at least 5 different annotators (in practice, except for one, all movies were assigned to 6 or more annotators). The annotators were provided with the guidelines that described in detail the background concepts, definitions of story types. They were asked to choose at least one story type for a given movie out of the 10 main story types described above, while *there was no limit on the maximum number of types they could choose*. To evaluate the annotator's attention during the task, ten simple quizzes about the movie plot were embedded in each movie.

Data Analytics. Our annotated dataset consists of 45 movies, with each movie labeled by 1 to 3 story type labels. There are 17 movies with only one label, 16 movies with two labels and 12 movies with three labels. On average, each annotator selected 1.5 story types per movie, and for 40 out of 45 movies, at least half of the annotators agreed on the same story type. The above chance agreement between annotators on so many movies confirm our first hypothesis that *most movies adhere fairly well to a general structure described by the screenwriting guidebook* [9]. All 10 story types assume that a movie has a single storyline. Cases of poor annotator's agreement were more common for the few movies with multiple storylines, (e.g., *Crash* (2004)).

To evaluate the quality of the annotations, we compared our collected story type labels with the "gold standard" labels for the 9 movies for which we have "gold-standard story types" (labeled by professional scriptwriters who are proficient in the *Save the Cat!* Theory [4,6]). After weighting the students' annotations, the agreement between the experts' labels and weighted students' labels increased to 8 out of 9 "gold standard" movies, indicating that the weighting strategy correctly strengthens the annotators who understood the concepts better and weakens those who did not. The agreement between the collected labels and the gold ones confirms our second hypothesis that *even non-experts can identify the story types after watching a movie and reading the guidelines*.

The three most frequent story types in the dataset are *Buddy Love*, *Dude with a Problem*, and *Rites of Passage*. On the other hand, story types such as *Superhero*, *Out of the Bottle*, *Monster in the House* and *the Fool Triumphant* are rarely selected. The type imbalance has become more severe due to our limited selection of movies. As common in such cases of severe imbalance [9], we kept only the three most frequent story types and replaced the rest with an additional *"Other"* label. The infrequent labels aggregation may not be necessary in a larger, less imbalanced dataset.

3. Methodology

Story Type Classification. In our evaluation experiments, we use two popular classification algorithms, C4.5 Decision Tree and Naive Bayes, with 5-fold cross validation. Considering the limited size of the particular dataset we use, instead of performing multi- label classification, we classify movies with respect to each story type separately, through binary classification. This limitation can be resolved once a larger scale dataset is available.

Feature Extraction. The 10 story types contain subjective and highly abstract concepts (such as "monster" or "fool") which are, apparently, difficult to infer from a few numerical features. According to the definitions, all stories are about the hero and the hero's surroundings. Characters are the essence of a story, and the attributes of the characters in a movie determine how the story develops: their activities, their emotions, the changes they undergo, their relationships, etc. Therefore, we mainly concentrate on the character-centric features, which measure the character importance and relationships.

Basic Features: Scenes are the basic narrative units of a movie. A typical scene consists of multiple shots. We use the *MovieGraphs* dataset to extract the scenes and its shots information about each movie. How rapid the scenes/shots change across the entire movie might, to some extent, reflect the complexity of the story. Moreover, the average number of shots per scene could be the clue to the intensity of the scene events, because a larger number of shot changes implies more event details, such as dialogue exchanges during conversations, moves during a fight, or even different memories during a flashback etc.

Character Network Features In order to capture the relationships between the characters in a given movie, we build a static character network for each movie, applying the CoCharNet algorithm [23] to the *MovieGraphs* dataset. The character network is a weighted graph consisting of nodes representing the characters and edges representing the co-occurrences between each character pair. Nodes are weighted by the appearance duration of the movie characters, while edges are weighted by the co-occurrence duration of the pair of characters they connect.

From the story type definitions it can be inferred that some types contain latent character relationships that can be discovered from their character network. E.g., a *Fool Triumphant* story is expected to have a *single hero* (the "fool") while a *Buddy Love* story is expected to have *two key characters*, the hero and the buddy.

Based on the character co-occurrence networks constructed for each movie, we extract the features, which reflects information about the character's social relations in the movies. The average (avg.) and standard deviation (std.) of the node degree represent how strongly "connected" the characters in the movie are and the features of edge weights are expected to show how dominant the corresponding social connections are.

Temporal Key Character Features It is a reasonable assumption that an important character appears in almost all the scenes, and therefore has an almost uniform appearance distribution. Therefore, by computing the similarity between the characters' appearance distributions and the uniform distribution using KL divergence, we can estimate the importance of each character. The smaller the KL divergence, the more important the character is.

4. Experiments

Experiment Design. As described above, we replaced each of the 7 least frequent story type labels by the *"Other"* label. Together with the three most frequent story types, we now have four story categories for the classification task. Our experiments were run on an annotated dataset of 45 movies, including 26 movies of the *Buddy Love* type, 19 of the *Dude with a Problem* type, 13 of the *Rites of Passage* type, and 22 assigned with the *Other* label. Considering the small size of the dataset, instead of applying multi-label classification algorithms, we attempted to identify each story type separately, i.e., we conducted four binary classification experiments on the four categories.

We evaluated the performance of the classifiers by stratified 5-fold cross validation, i.e., 36 movies for training and 9 movies for testing in each fold, and the fold splitting was done separately for each story type. Moreover, in order to further minimize the influence of data imbalance, we weighted the samples with the inverse frequency of their label. We evaluated the following classification algorithms available from the Scikit-Learn Library [26]: Logistic Regression, Naive Bayes, Support Vector Machine with linear kernel, and C4.5 Decision Tree.

We performed feature selection by computing the mutual information between each feature and every class label; then, for each label, we selected the top three, four and five features with the highest information gain (no matter which set they were from) for the classification task. The features were all normalized with L1 normalization.

There were two hyper-parameters in the experiment: We limited the maximum depth for the decision tree classifier to 3, to avoid overfitting. The KL divergence threshold was set to 1.0 for determining the number of key characters.

5. Results and Discussion

Table 1 presents the results of our classification experiments. It shows baseline results, best results (classifier + selected features), and the features used for obtaining the best results. The baseline results were computed by labeling all movies with the positive type in each binary classification experiment, respectively, leading to the universal 1.00 recall. Although we sometimes failed to significantly improve the F1 measure, as for the *Buddy Love* story type and *Other* story types shown in Table 1, the obtained precision and accuracy were encouraging, and passed the t-test (p value less than 0.05) for all story types compared to the baseline.

As we expected, the importance of some features is story type dependent. For example, the edge weights in the character networks play an important role for identifying the *Buddy Love* movies, because in those movies the relation (edge weight) between the two main characters (the hero and his/her buddy) is expected to be much stronger than the other features, which can somehow be reflected by the average and standard error of the edge weights. The number of characters in the third quarter of the movie is the most important feature for identifying the *Dude with a Problem* type. Considering the movie deconstruction, the third quarter of the movie is likely to include the "confrontation" part of story, as well as the "development", in which the protagonist (the dude) experiences his/her essential change. During this period, it is likely that either his companions or his enemies appear. Another important feature is the number of edges, which reveals how many connections the characters have, instead of how strongly they are connected to each other. In a typical *Dude with a Problem* movie, almost every character is connected to the protagonist, leading to a high number of edges in the character network. The most important feature of the *Rites of Passage* type is the average degree of the graph, which reflects on average, how connected the characters are, and may give a hint of the social network in the story. Regarding the *Other* class, although it is a mixture of seven less frequent story types, the four most dominant features are all character-related, showing the characters' importance within the stories. Finally, basic features (e.g. scenes per minute, average scene duration, movie duration) also played important roles in identifying the three most frequent story types.

Table 1. Story type binary classification results. For each story type, we present its baseline result and the best combination of features and classifiers (DT for decision tree and NB for Naive Bayes). The features are all normalized with L1 normalization.

Method	Prec.	Recall	F1	Acc.	Selected Features
			Buddy Love		
Baseline	0.58	1.00	0.73	0.58	1. avg. edge weights
DT/NB+L1	**0.67***	0.92	0.77	**0.69***	2. Std edge weights 3. scenes/min
			Dude with a Problem		
Baseline	0.42	1.00	0.45	0.29	1. # key char (3^{rd} quat)
NB+L1	**0.58***	0.75	0.63	**0.67***	2. scenes/min 3. number of edges
			Rites of Passage		
Baseline	0.29	1.00	0.45	0.29	1. avg. degree 2. avg. scene duration
DT+L1	**0.75***	0.63	0.63	**0.80***	3. scenes/min 4. duration
			Other		
Baseline	0.49	1.00	0.66	0.49	1. # key char (full) 2. # key char (3^{rd} quat)
NB+L1	**0.63***	0.81	0.70	**0.67***	3. std. edge weights 4. avg. edge weights

With a small and imbalanced movie dataset (the main limitation of this research) and with simple character-related features, we were able to identify the three most frequent story types with accuracy significantly higher than the baseline (majority vote). Further progress can be achieved with more annotated movies and more extracted features. With a sufficiently large dataset, one could attempt to classify the 36 dramatic situations of [7].

The selected features are highly related to movie stories and can be used as one of the main information sources for narrative understanding. The high level abstract concepts associated with each story type, e.g., *an innocent hero*, *a sudden event*, and *a life or death battle* for the *Dude with a Problem* story type, can provide some automatic understanding about the protagonist's character traits and the motivations that drive him in facing his challenges.

A practical application of movies story type detection may be the construction of a better movie recommendation system [5,8] assuming that each movie lover has a strong preference towards certain story types.

References

[1] Avgerinos, C., Nikolaidis, N., Mygdalis, V., Pitas, I.: Feature extrac- tion and statistical analysis of videos for cinemetric applications. In: 2016 Digital Media Industry Academic Forum (DMIAF). pp. 172–175 (2016). https://doi.org/10.1109/DMIAF.2016.7574926

[2] Barros, C., Lloret, E., Saquete, E., Navarro-Colorado, B.: Natsum: Narrative abstractive summarization through cross-document timeline generation. Information Processing & Management **56**(5), 1775–1793 (2019). https://doi.org/https://doi.org/10.1016/j.ipm.2019.02.010

[3] Booker, C.: The seven basic plots: Why we tell stories. A&C Black (2004)

[4] The "gold-standard types" are provided in the manner of movie deconstruction articles, an example can be found here: https://savethecat.com/beat-sheets/the-silver-linings-playbook-beat-sheet

[5] Choi, S.M., Ko, S.K., Han, Y.S.: A movie recommendation algorithm based on genre correlations. Expert Systems with Applications **39**(9), 8079–8085 (2012). https://doi.org/https://doi.org/10.1016/j.eswa.2012.01.132

[6] Snyder, B.: Save the Cat! Goes to the Movies: The Screenwriter's Guide to Every Story Ever Told (2007).

[7] Figgis, M.: The Thirty-Six Dramatic Situations. Faber & Faber (2017)

[8] Frémal, S., Lecron, F.: Weighting strategies for a recommender system using item clustering based on genres. Expert Systems with Applications **77**, 105–113 (2017). https://doi.org/https://doi.org/10.1016/j.eswa.2017.01.031

[9] Snyder, B.: Save the Cat!: The Last Book on Screenwriting You'll Ever Need. Cinema/Writing, M. Wiese Productions (2005)

[10] Dataset link in GitHub

[11] Tapaswi, M., Bauml, M., Stiefelhagen, R.: Storygraphs: visualizing character interactions as a timeline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 827–834 (2014)

[12] Tobias, R.B.: 20 MASTER Plots: and how to build them. Penguin (2011)

[13] Ji, J., Krishna, R., Fei-Fei, L., Niebles, J.C.: Action genome: Actions as composi- tions of spatio-temporal scene graphs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)

[14] Jorge, A.M., Campos, R., Jatowt, A., Nunes, S.: Information processing & management journal special issue on narrative extraction from texts (text2story): Preface. Information Processing & Management **56**(5), 1771–1774 (2019). https://doi.org/https://doi.org/10.1016/j.ipm.2019.05.004

[15] Kukleva, A., Tapaswi, M., Laptev, I.: Learning interactions and relationships between movie characters. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9849–9858 (2020)

[16] Lee, O.J., Jung, J.J.: Integrating character networks for extracting narratives from multimodal data. Information Processing & Management **56**(5), 1894–1923 (2019). https://doi.org/https://doi.org/10.1016/j.ipm.2019.02.005

[17] Lee, O.J., Jung, J.J.: Modeling affective character network for story analytics. Future GenerationComputer Systems **92**, 458–478 (2019). https://doi.org/https://doi.org/10.1016/j.future.2018.01.030

[18] Lee, O.J., Jung, J.J., Kim, J.T.: Learning hierarchical representations of stories by using multi-layered structures in narrative multimedia. Sensors (Switzerland) **20**(7), 1–27 (2020). https://doi.org/https://doi.org/10.3390/s20071978

[19] Lee, O.J., Kim, J.T.: Measuring narrative fluency by analyzing dynamic interaction networks in textual narratives. In: Text2Story@ ECIR. pp. 15–22(2020)

[20] Liu, C., Last, M., Shmilovici, A.: Identifying turning points in animated cartoons. Expert Systems with Applications **123**, 246–255 (jun 2019). https://doi.org/https://doi.org/10.1016/j.eswa.2019.01.003

[21] Liu, C., Shmilovici, A., Last, M.: Towards story-based classification of movie scenes. PLoS ONE (2020). https://doi.org/https://doi.org/10.1371/journal.pone.0228579

[22] Materzynska, J., Xiao, T., Herzig, R., Xu, H., Wang, X., Darrell, T.: Something- else: Compositional action recognition with spatial-temporal interaction networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)

[23] Tran, Q.D., Jung, J.E.: Cocharnet: Extracting social networks using character co-occurrence in movies. J. Univers. Comput. Sci. **21**(6), 796–815 (2015)

[24] Vargas, J.V.: Narrative information extraction with non-linear natural language processing pipelines. Drexel University (2017)

[25] Papalampidi, P., Keller, F., Lapata, M.: Movie plot analysis via turning point identification. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 1707–1717 (Nov 2019). https://doi.org/https://doi.org/10.18653/v1/D19-1180

[26] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

[27] Vicol, P., Tapaswi, M., Castrejón, L., Fidler, S.: Moviegraphs: Towards understand- ing human-centric situations from videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)

[28] Schmidt, V.L.: 45 Master Characters, Revised Edition: Mythic Models for Creating Original Characters. Penguin (2011)

[29] Zhang, H., Boons, F., Batista-Navarro, R.: Whose story is it anyway? automatic extraction of accounts from news articles. Information Processing & Management **56**(5), 1837–1848 (2019). https://doi.org/https://doi.org/10.1016/j.ipm.2019.02.012