FILM-SCRIPTER: OPEN SOURCE SCRIPT GENERATOR WITH A CHAR-RNN

Claire Goeckner-Wald

Caltech
MSC 318

ABSTRACT

This research attempts to recreate and validate the model of machine-generated short film script *Sunspring*, ideally to provide an open-source film script generator for filmmakers. First, we examine the methodology outlined by Ross Goodwin, creator of the *Sunspring* model. We then apply a "cumulative priming" method, with the goal of enhancing the script generation with plot continuity. It was found that the cumulative priming method was insufficient for significantly continuous plot.

Index Terms— char-rnns, film scripts, plot, priming strings, cumulative priming

Repository— `github.com/cgoecknerwald/film-scripter`

1. INTRODUCTION

Pew Research shows that young people are spending increasing amounts of time streaming entertainment content, and decreasing amounts of time watching traditional television [1]. Many entertainment providers, such as HBO, Hulu, Amazon Video and Netflix, offer television content to stream online, anytime. These four providers alone order large amounts of content to remain competitive with high-output free streaming services with less traditional entertainment, particularly low-budget independent creators. However, a single hour of serial television can cost between $2 and $3 million to produce [2].

Screenwriting is a mainstay of pre-production processes. High costs and slow turnaround for screenwriting can stymie productions. Because most scripts are semi-structured, and can be quantitatively analyzed, they are a target for artificial intelligence. Moreover, the same algorithm that generates the script could also assist in the identification of relevant locations, props, cast members, costumes, special effects, and visual effects. Semi-structured scripts have been shown to be quantitatively analyzable, allowing for a sufficiently intelligent algorithm to be able to recreate plot movements characteristic of episodic productions [2]. Natural language processing (NLP) has been enhanced to the point that screenwriting is now a valid option, provided we invest in long-term research. By investing in research for artificially created scripts, we can substantially enhance the entertainment industry with lower-cost, faster-time pre-production. This would allow for traditional studios, such as CBS Television Studios, to increase their productivity and throughput, in order to remain competitive with streaming services.

2. BACKGROUND

In 2016, Oscar Sharp, BAFTA-nominated filmmaker, teamed up with AI researcher Ross Goodwin to produce what some believe to be the first film with an artificially generated script [3]. The short film was produced for the 2016 48-hour Film Challenge of film festival Sci-Fi-London. The algorithm, which named itself Benjamin, was trained on science fiction scripts. Limiting the bot to science fiction scripts caused the generated scripts to have characteristics highly reminiscent of the genre – e.g., the line "he picks up a light screen and fights the security force of the particles of a transmission on his face" would likely not have been generated by a bot trained on the works of Jane Austen, who may never have used the phrase 'security force.' This genre restriction may have been a strategic move to limit the amount of garble produced. Rather than training the bot to produce character sequences from romance, western, sci-fi, horror, and more, the bot can only output sequences it has learned from common science fiction tropes, such as spaceships, stars, and black holes. Faith to the science fiction tropes would theoretically make for more sensical output when generating sentences that are otherwise random and independent from each other.

The *Sunspring* script was edited by the production team to fix some 'garbled script,' but it is unclear what type of editing was done because the bot's original output was not published. Furthermore, the director instructed the actors to interpret the script freely. With the added context provided by props, delivery, and direction, the final product of *Sunspring* was alternatingly "hilarious and intense". Critics called it a "novelty" and a "thought experiment", as well as "fascinatingly incoherent" [4][5]. By using a bot as a screenplay writer, script pre-production time was whittled down to essentially negligible editing. This, in turn, allowed nearly the full 48 hours of production to be spent on filming and post-production processes [6]. Thus, use of the algorithm expanded the working time for other, noninterchangeable processes.

The generated script, however, had a significant number of quirks and inconsistencies that earned it the labels 'incoherent' and 'neurotic' from critics [5]. Some of the inconsistencies are obvious to the viewer – particularly, the extremely short-term memory of the characters that cause conversations to seem more like a ping pong rally than an exchange of ideas. The dialogue also featured inconsistent over- and underreactions and a bizarre, rambling monologue. Other 'quirks' were not so visible to the audience, such as the stage instructions. One stage instruction reads "He is standing in the stars and sitting on the floor."

As demonstrated by these quirks, it is important that the script has context - both within itself, and to the film universe. For that reason, we seek to expand upon *Sunspring*'s achievements by enhancing the continuity through "priming" (also called "pre-seeding"). In theory, one could repeatedly feed a model its own output in order to generate output that bears stronger resemblance to the previous output, rather than generating each sequence independently.

3. DATA

Screenplays are multi-faceted and store a variety of data types concurrently. Most film script databases have erratic storage methods and no standardization. Hand-cleaning is not effective, since we need "25-100 MB raw text, or 50-200 novels" for the long short-term memory (LSTM) model we will be replicating [3]. In addition, most film script databases are operating illegally, due to US and international copyright laws. Ross Goodwin was unable to release the dataset he trained the *Sunspring* bot on, due to concerns about liability, so we must build our own dataset from scratch.

In order to build the dataset, we used three resources: the Cornell Movie-Dialogues Corpus, Project Gutenberg, and the CMU Movie Summary Corpus. Since each of these resources contains different data types (dialogue, narration, and summaries, respectively), we trained three different models.

– *Cornell Movie-Dialogues Corpus:* contains "conversations extracted from raw movie scripts: 220,579 conversational exchanges. . . involving 9,035 characters from 617 movies" [8].

– *Project Gutenberg:* hosts "over 57,000 free eBooks. . . not protected by U.S. copyright law" [9].

– *CMU Movie Summary Corpus:* contains "42,306 movie plot summaries extracted from Wikipedia. . . " [10].

Thus, we trained: a model that could generate the plot summary; a model that could generate narration; and a model that could generate dialogue.

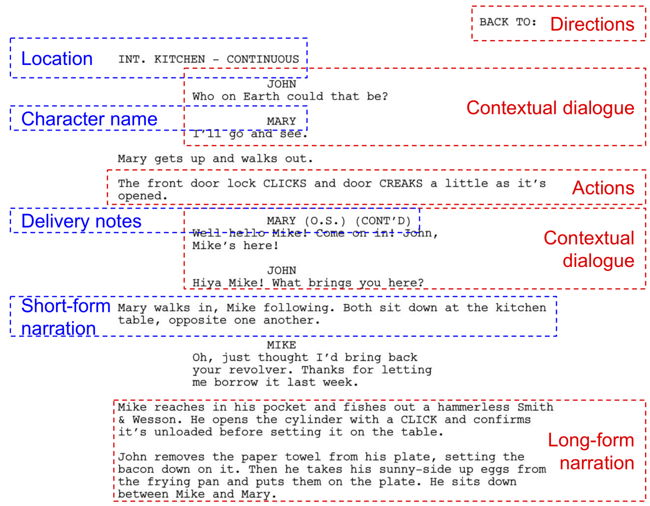

Fig. 1. Annotated example industry-standard screenplay [7]

THREEPIO Are you sure this things safe? EXT. TATOOINE - ANCHORHEAD SETTLEMENT - POWER STATION - DAY Heat waves radiate from the dozen or so bleached white buildings. Luke pilots his Landspeeder through the dusty empty street of the tiny settlement. An old woman runs to get out of the way of the speeding vehicle, shaking her fist at Luke as he flies past. WOMAN I've told you kids to slow down! INT. POWER STATION - DAY

PROMOTER ...Balboa!?Rocky raises his head. The promoter steps over. PROMOTER (continuing) ... Twenty bucks for the locker an' cornerman -- Two bucks for the towel an' shower, seven for tax -- The house owes ya, sixty-one dollars.The man peels off the money and departs... Rocky closes hislocker, nods to the defeated fighter, and leaves.INT. TROLLEY - NIGHTRocky is on the trolley heading to South Philly... His hair lookslike it has been shaped with hedge clippers. His name isMIKE. ROCKY Yo, Mike -- What's happenin' here?

EXT. GREENBOW, ALABAMA Mrs. Gump and young Forrest walk across the street. Forrest walks stiffly next to his mother. FORREST (V.O.) Now, when I was a baby, Momma named me after the great Civil War hero, General Nathan Bedford Forrest. EXT. RURAL ALABAMA A black and white photo of General Nathan Bedford Forrest. The General is in full Ku Klux Klan garb, including his horse. FORREST (V.O.) She said we was related to him in some way. And, what he did was, he started up this club called the Ku Klux Klan. They'd all dress up in their robes and their bedsheets and act like a bunch of ghosts or spooks or something.

Fig. 2. Example scraped screenplays: *Rocky*, *Forrest Gump*, and *Star Wars: A New Hope*, respectively.

4. SCRIPT-GENERATION

Given the three datasets available, we theorized that one could use priming strings to recreate plot structure without the excessive dialogue-tagging necessitated by Murtagh's proposal [2]. By first generating a plot summary, one could prime the output generated by next model (either narration or dialogue), in order to target the narration and dialogue to the plot summary. While it isn't guaranteed that the model adheres to the strings it has been primed with, it is in theory far more likely to generate similar strings.

(In theory, this could be expanded. One could generate a plot summary, manually section it into the traditional 3 acts, then train one narration and one dialogue model for each act's plot summary.)

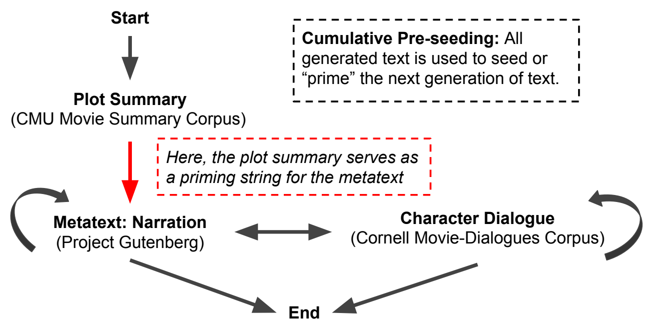

Fig. 3. Proposed script generation via cumulative priming. This generation pattern would create a script with the sequential pattern "Plot Summary → Narration or Dialogue → . . . → Narration or Dialogue → The End," which was chosen because it is an approximation of the actual patterns present in film scripts.

5. CHAR-RNN

PyTorch, the open-source machine learning library for Python, was chosen as a base technology for training due to its ease of use and large community. The code was based off of Sean Robertson's `char-rnn.pytorch`, which in turn was borrowed from the Practical PyTorch series [11]. Training each model took 11 - 16 hours, using Nvidia's CUDA on an Nvidia GTX 1070. The following training stats were applied:

– LSTM (Long Short-Term Memory);

– 2^{16} (65,536) iterations;

– 2 or 3 layers;

– 512 hidden states;

– 0.01 learning rate;

– 256-char sequence length;

– 128 batch size;

– 0.25 dropout.

No significant difference was noticed between 2 or 3 layers, besides increased training time. The repository for the code of this model is available at:

`github.com/cgoecknerwald/film-scripter`.

6. MODEL PERFORMANCE

The following sample dialogue demonstrates the abilities of the dialogue generation trained on the Cornell Dialogue Corpus. The model was primed with "Where," and demonstrates increasing temperatures.

Temp.	Output
`0.2`	Where are you going to do to you? \n What are you doing to see the street? \n What do you mean? What are you talking to you? \n What are you doing?
`0.4`	Where are you so change? \n What do you think? \n So what are you talking to you a lot of beat the British of you are going to do to you?
`0.6`	Where are you talking to you to see the same? \n What did you get his starship? \n What do you want out? \n Oh, no. I like you.
`0.8`	Where are you? \n Yeah, I could say nothing. \n What? \n Spercoursex, sure. I saw you the time you got hype. \n What is your pattern and do something? You know that, someone up? Saint suicides Lieutenable, Mother. Frank.
`1.0`	Where are, bene? \n I thought he love afters six man with degene, Jertain... I saw me. \n I dunnoch - his direct travel, to see you. \n So–

Fig. 4. With increasing temperature ('Temp.'), the model has more 'confidence,' and will generate wilder character sequences, even going so far as to make new words and new grammatical patterns. Newlines are represented as \n. Each generation was primed with the string "Where", and each sample listed was the first sample output, to avoid cherry-picking results.

7. CONCLUSION

The cumulative priming technique explored was insufficient to improve upon Ross Goodwin's prior work. The models were largely unable to capture the 'spirit' of sequences they were primed with. However, because the models were not trained particularly well due to the lack of computational resource, further research is necessary before entirely discounting cumulative priming.

Furthermore, the models used, though fast learners, were too erractic at high temperatures. Rather than abstractly explore new concepts and connections, they simply made new words and new grammatical patterns that throw the script off-track.

Finally, there is a lack of comprehensive film script training sets. This research therefore provides its open-source tools to work around erratic literature from Project Gutenberg (see `gutenberg_parser.py` and others).

8. FURTHER WORK

In the screenwriting world, scripts are broken down into acts, sequences, scenes, and beats. Each term describes a unit of change within the film. Traditionally, the *plot* features a *climax* and has 2 - 3 *acts* (forming the macrostructure). Each act is composed of *sequences*; each sequences is composed of *scenes*; a *beat* is a unit of action or behavior (forming the plot's microstructure).

$$\text{plot} \supset \text{acts} \supset \text{sequences} \supset \text{scenes} \supset \text{beats}$$

However, it is currently difficult to capture these qualitative subtleties with current machine learning techniques.

In 2009, Murtagh et. al. attempted to quanititatively analyze film scripts [2]. As shown in Figure 5, they were able to quantify the script dialogue such that scenes could be clustered by relationship to the overall plot. This is important for script generation because it allows us to quantify the amount of change occurring in each scene. Therefore, we could theoretically apply some important machine learning concepts to the relatively abstract quality of plot, thereby enhancing and maintaining the overall plot arc(s).

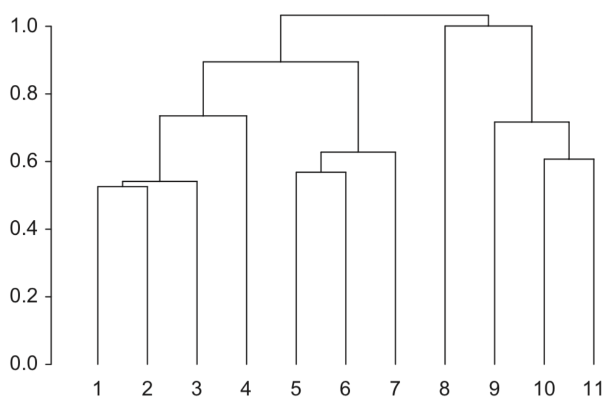

Fig. 5. Hierarchical Clustering of 11 beats from *Casablanca* [2]. The clustering of these beats represents the contribution of each beat to the total 'scene'. Beats that are clustered together are beats that are conceptually closely related. For example, beats 5, 6, and 7 are each of main characters Rick and Laszlo expressing rapprochement towards one another. Thus, they are closely conceptually related.

However, no legal dataset currently exists that can provide a sufficiently large amount of tagged and ordered dialogue to train using Murtagh's algorithm. In general, there is a great need for curated datasets before progression in this field.

9. ACKNOWLEDGEMENTS

The author wishes to thank Ross Goodwin at Interactive Telecommunications Program NYU and BAFTA-nominated filmmaker Oscar Sharp for inspiration via *Sunspring*; Sean Robertson (`spro`) of Prontotype for the MIT-licensed basis for the char-RNN repository in PyTorch; Cristian Danescu-Niculescu-Mizil of the Department of Information Science, Cornell University, for the Cornell Movie-Dialogs Corpus; Bamman, O'Connor, and Smith at the Language Technologies Institute and Machine Learning Department at Carnegie Mellon University for the CMU Movie Summary Corpus; and the countless authors whose works in the public domain were invaluable to the meta-text training set (via the Gutenberg Project).

10. REFERENCES

[1] Lee Rainie, "About 6 in 10 Young Adults in U.S. Primarily Use Online Streaming to Watch TV," 2017.

[2] Adam Ganz Murtagh, Fionn and Stewart McKie, "The Structure of Narrative: The Case of Film Scripts," *Pattern Recognition*, vol. 42, no. 2, pp. 302–312, 2009.

[3] Ross Goodwin, "Adventures in Narrated Reality: New forms & interfaces for written language, enabled by machine intelligence," 2016.

[4] Dyllan Furness, "Sunspring is an Absurd Short Film Written by AI | Digital Trends," June 2016.

[5] Annalee Newitz, "Movie written by algorithm turns out to be hilarious and intense," June 2016.

[6] HAL 90210, "This is what happens when an AI-written screenplay is made into a film," *The Guardian*, June 2016.

[7] Mendaliv, "Sample from a screenplay, showing dialogue and action descriptions," https://commons.wikimedia.org/w/index.php?curid=4970002, 2008, Online; accessed December 2018.

[8] Cristian Danescu-Niculescu-Mizil and Lillian Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs," in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.

[9] Michael Hart, "Project Gutenberg: Free eBooks," http://www.gutenberg.org/, 1971, Online; accessed August - December 2018.

[10] Brendan O'Connor David Bamman and Noah A. Smith, "Learning Latent Personas of Film Characters," in *Association for Computational Linguistics 2013: Sofia, Bulgaria*, 2013.

[11] Sean Robertson, "char-rnn.pytorch," https://github.com/spro/char-rnn.pytorch, 2018, GitHub repository; accessed December 2018.