Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

StarTroper, a film trope rating optimizer using Deep Learning and Evolutionary Algorithms

Rubén Héctor García-Ortega[1] | **Pablo García-Sánchez**[2] | **Juan Julián Merelo-Guervós**[3]

[1]Development Department, Badger Maps, Granada, Spain
[2]Department of Computer Science and Engineering, University of Cádiz, Cádiz, Spain
[3]Department of Computer Architecture and Technology, University of Granada, Spain

Correspondence
*Rubén Héctor García Ortega, Badger Maps.
Email: raiben@gmail.com

Present Address
Badger Maps. Granada, Spain.

Abstract

Designing a story is widely considered a crafty yet critical task that requires deep specific human knowledge in order to reach a minimum quality and originality. This includes designing at a high level different elements of the film; these high-level elements are called tropes when they become patterns. The present paper proposes and evaluates a methodology to automatically synthesise sets of tropes in a way that they maximize the potential rating of a film that conforms to them. We use deep learning to create a surrogate model mapping film ratings from tropes, trained with the data extracted and processed from huge film databases in Internet, and then we use a Genetic Algorithm that uses that surrogate model as evaluator to optimize the combination of tropes in a film. In order to evaluate the methodology, we analyse the nature of the tropes and their distributions in existing films, the performance of the models and the quality of the sets of tropes synthesised. The results of this proof of concept show that the methodology works and is able to build sets of tropes that maximize the rating and that these sets are genuine. The work has revealed that the methodology and tools developed are directly suitable for assisting in the plots generation as an authoring tool and, ultimately, for supporting the automatic generation of stories, for example, in massively populated videogames.

KEYWORDS:
Content Generation; Tropes; Computational Narrative; Deep Learning; Genetic Algorithms

1 | INTRODUCTION

Crafting film scripts is quite challenging because of the plot complexities and the *multiplicative production function of entertainment* (Hennig-Thurau & Houston 2019), that promulgate that the elements involved in the development of a media product need to work together and a single failing one may provoke a disaster in cascade. In fact, the concept of narrative itself can be seen as a complex adaptive system where interactions between their elements or events make the story emerge (Sack 2014). In order to tackle this complexity in the current research we are going to to describe the elements of the film and how well they combine and interact; our candidates for both mechanisms are the tropes that have been discovered in the films, and the massive human-evaluated ratings, respectively.

A trope is as a recurring narrative device or pattern, according to the definition by Baldick (2015); it could be a technique, a motif, an archetype or a *cliché*, used by the script writers, producers and directors to achieve specific effects that might vary from interest-increasing to surprising through recall familiarity or entertaining, in their creative works, such as books, films, comics or videogames. Some tropes are broadly adopted and academically studied such as the *Three-act Structure* formulated by Field (1982), the *hero's journey* studied by Vogler (2007), the *McGuffin* popularized by Hitchcock, according to ois Truffaut, Hitchcock, and Scott (1985), and the *Chekhov's Gun* formulated by the Russian writer with the eponymous name, according to Bitsilli (1983); however, there are thousands of not-so-widely used tropes as well, discovered and catalogued everyday by professionals and enthusiastic of the storytelling; their study is organic, dynamic and extensive, according to García-Ortega, Merelo-Guervós, Sánchez,

and Pitaru (2018). In fact, their emergence, self-organization and pattern formation are difficult to model, implying they can be approached as a complex system (Juarrero 2000). In general, we might say that the set of tropes defines the overall narrative architecture over which the narrative layer is eventually set. Tropes do not define plot univocally, but constrain it in a number of ways. Thus, we can roughly characterize a film by using the set of tropes that we can find in it.

Along this paper, we are going to use the analogy of the *Film DNA* for describing its set of tropes (and genres; hereafter when we say tropes we will also include genres, since they are a type of meta-tropes); we will define it as the set of tropes that are found in a film and define things such as its structure, characters, events, mood, settings and narration. As tropes are *living concepts*, whose number grow as they are discovered as common patterns in other stories, the *Film DNA* is, by definition, incomplete and evolving, yet it is still interesting to define stories, categorize them and model them from a mathematical perspective. The challenge of our research is to build original synthetic Film DNAs based on a huge corpus of film-tropes, and through computational intelligence, in a way that they have an intrinsic potential quality when reflected together in a film.

At the same time, we need to be able to associate a measure of quality to the *Film DNA* and it needs to summarize many factors, at last, perceived and evaluated by humans. Luckily we have access to databases with films' information that includes the genres of the films and their human evaluated ratings, provided by the community of fans. If we are able to construct a knowledge base of films, including the Film DNAs that we mentioned previously, the genres and the rating, in a huge *extended dataset*, we would be able to process them in order to make suggestions of tropes that optimize the predicted rating; however, even though intuitively the *Film DNA* is a profound way to describe a story from many different perspectives, following the analogy of the DNA, there are epigenetic factors that could deeply affect the performance of the story as well. This method does not guarantee the quality of a film, as a film that develops from a *Film DNA* may implement them in infinite ways with very different results in terms of quality; however, it can be used as an indicator of the *potential of the story* or the most probable implementation based on the universe of currently analyzed films. A synthetic film DNA might also find points in the film landscape that have not been explored so far, making new blockbusters emerge in an arena that is increasingly dominated by data analytics.

The main objective of this paper is to demonstrate how computational intelligence can be used to generate and improve sets of tropes that maximize the potential rating of the films that conforms them, in the context of authoring tools and Content Generation. Our approach extracts 11846 *Film DNAs* that contain in sum 26246 different tropes from external Data Sources and maps it to a database of film ratings and genres, dealing with disambiguation heuristics in order to build what we have called the *extended dataset*. However, submitting a set of tropes to the box office is impossible, which is why we use Deep Learning (LeCun, Bengio, & Hinton 2015) to create surrogate models that are able to infer the rating from any combination of tropes. We perform different analysis in order to determine the quality of the predictions and the parameters that could affect them. Later on, a Genetic Algorithm (GA) (Whitley 1994) and their operators are defined in a way that the trope combinations, formerly Film DNAs, are evolved relying in the surrogate model to maximize the rating.

The remaining of this work is organized as follows: in the Section 2 we explore the current state of the art in plot generation based on tropes, in the Section 3 we deepen the methodology presented above, in the Section 4 we describe the experiments carried out to evaluate the methodology and discuss the results, and in the Section 6 we summarize the outcomes and future work.

2 | STATE OF THE ART

Film-makers, researchers and software developers are using known narrative patterns to build compelling stories, whether through an authoring tool or an Automatic Content Generator. Some of these patterns are being studied for decades, for example, the Propp's formalism (Propp 2010), first edited in 1928 and based on seven different roles, every one with a list of actions that can take over the course of a story, in a fixed sequence of 31 functions. Propp's formalisms are currently used by the computer scientists to build systems that generates instances of Russian folk tales (Gervás 2013) or stories and discourses with characters/places/objects relating to Iwate prefecture in Japan through micro/macro story techniques (Imabuchi, Akimoto, Ono, & Ogata 2012). Other example of a popular narrative pattern widely used today is the *hero's journey*, proposed by Campbell (2008), first edited in 1949, and later on, reviewed from the perspective of the film industry by Vogler (2007). It divides the story in 3 acts (*departure*, *initiation* and *return*), with 17 non-mandatory stages that are the universal scaffolding of ancient myths as well as modern day adventures. The *hero's journey* has been used for interactive storytelling in video games (Delmas, Champagnat, & Augeraud 2007), and massive backstories generation (García-Ortega, García-Sánchez, Merelo Guervós, Ginés, & Cabezas 2016).

The Propp's formalism and the *hero's journey*, as well as many other widely used narrative patterns, are included in the definition of trope. According to Mellina and Svetlichnaya (2011), "a trope is a unit of literary currency, recurring in works over time and gaining meaning through audience recognition of its connotations and associations". Their work, is one of the inspiration sources of our current research, as they use the tropes from TV Tropes, a wiki that collects and documents descriptions and examples of tropes, together with other features from IMDb, an online database of information related to multimedia content, including the rating. They use the *Jaccard coefficient* to measure the similarity between sets of tropes and discovered that this similarity moderately predicts external measures of film acclaim. As in their study, in our research we will use the two data

sources, TV Tropes and IMDb, and we will use the Jaccard coefficient as trope similarity as well. However, in spite of having a similar base, their focus is the community detection and ours is to use the dataset to predict the rating of Film DNAs.

Other authors are using the tropes from TV Tropes in their research as well. In the work of Thompson, Padget, and Battle (2018), a system of agents relies on tropes to obtain a consistent narrative, to describe the social norms that model the world in which they live. The authors, as in this work, use tropes available on TV Tropes as a base and translate them into logic statements that express duty or obligation, using TropICAL language, which are the input for a logic programming solver; however, they do not use all the tropes, but a small set chosen by hand, which means that the range of resulting stories is going to be limited. Guarneri et al. (2017) built an authoring tool that proposes the use of tropes according to a narration structure, and takes the objects from the card Game 'Once upon a time'. They selected 94 elements out of 176 present in the Periodic Table of Tropes, a subset of the most famous tropes from TV Tropes.

Nevertheless, it is very complicated to evaluate the content generated by an automatic generator, not only because of its non-deterministic behaviour that makes it difficult to predict its outputs, but also because of the subjective, diverse and stochastic nature of the audience, as stated by Yannakakis and Togelius (2018). To evaluate a generator one can use directly the opinion of the designer, or indirectly from the audience, for example, using surveys, as in the work by Guarneri et al. (2017). In our research, we are also interested in a smart use of the tropes within a story but, unlike Guarneri et al. (2017), we are relying in Artificial Intelligence techniques by simulating and estimating the quality of the content via some metrics, while benefiting from the wisdom of collective opinions (**?**)surowiecki2005wisdom) in IMDb for the rating. In their work, Hsu, Shen, and Xie (2014) used different prediction models for the film rating from features of the films in IMDb and, according to their evaluations, the Neural Network gave the best results, in comparison with linear combinations and multiple linear regressions. In our current research, we will apply the same idea of using a Neural Network as predictor, but fed from the film's tropes instead of from the film's features such as the director, actors or writers. In fact, through the use of user-generated data, it is possible to obtain a large corpus of examples to be used in computational narrative, as in the work of Guzdial, Harrison, Li, and Riedl (2015). Moreover, it is possible to extract information about review sites, according to Pang and Lee (2007), such as MetaCritic or IMDb, to be the input of a model like the one we propose in this article.

Some authors as Bui, Abbbass, and Bender (2010) have addressed the problem of the optimization of stories by applying evolutionary computation. In their work, they developed a regular grammar model with causal relationships and they evolve it, demonstrating that evolutionary computation can potentially contribute significantly to story generations. Our previous work also relies in Evolutionary Computation as a mechanism to optimize stories: in (García-Ortega et al. 2015), we proposed the MADE framework: a parameterizable multi-agent system that allowed the generation of backstories in massive environments. A GA was used to optimize the parameters of the system, for instance the simulation time, the size of the world and the parameters of the behaviour of the agents, with respect to the appearance of different archetypes, such as the *hero* or the *villain*. These archetypes are defined by the possible actions that an agent may perform: for example, the archetype Villain appears when an agent fights against another for food. However, this form of evaluation was difficult to justify in order to measure the quality of the generated stories, since it was based on an objective decision: just the number of different character archetypes that emerge during the run of the world. Also, the list of possible actions for the agents was very limited. That is the reason that, later on, we proposed a more advanced model, with more complex agents and the possibility of extracting knowledge from a logical reasoner (García-Ortega et al. 2016). On this occasion we used as quality metric the appearance of the *hero's journey*, and the different archetypes that compose it. In order to do this, logical reasoning was applied based on the predicates produced by the different events that emerged in the system. However, as in previous work, the mere appearance of the *hero's journey* does not fully serve as a measure of the interest of the stories generated by the system.

That is why in this work we propose the combination of the previous ideas: We will make use of the tropes as a way to model stories, in our case, films, and we will optimize sets of tropes (our Film DNA) through evolutionary computation. We will use a Neural Network as surrogate model of a genetic algorithm as well, and we will train it with tropes from TV Tropes and features from IMDb, including the rating in order to predict the rating.

3 | METHODOLOGY

Our methodology is divided in four main steps, explained below and described in the Figure 1:

Step 1 Extract/scrape the tropes for every film and codify them as *Film DNAs*. As we will explain, our dataset will have limitations derived from the fact that is fed from the community, finding that popular films are broadly described in terms of tropes and unpopular films poorly described or directly missing. We will analyze this variability and how it could affect the performance of the prediction model.

Step 2 Extract ratings and genres from an external Film Database finding the unequivocal film names and cull the original TV Tropes dataset. This *extended dataset* will show limitations as well based on the original one and the automatic matching based on different heuristics. As we will see, a trope that is widely used does not need to be linked to good ratings, tropes that are present in bad films can become good in different combinations and vice-versa.

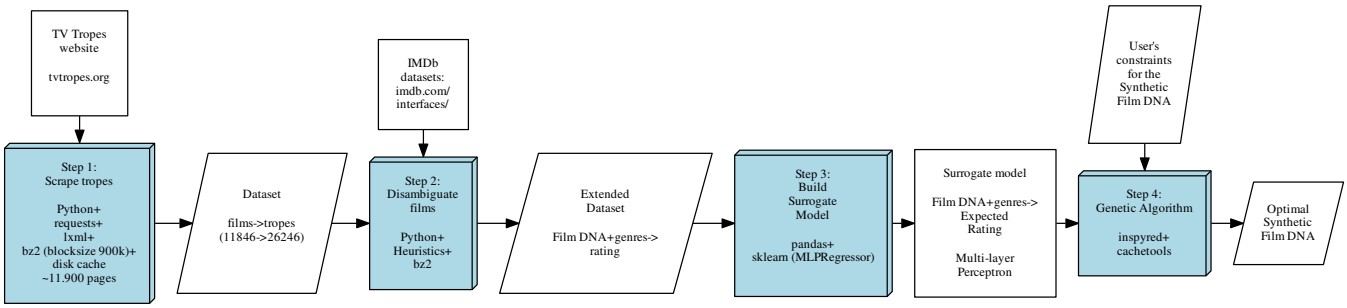


FIGURE 1 Methodology to generate constrained optimal Film DNAs using Genetic Algorithms with Neural Networks as surrogate models, fed from TV Tropes and IMDb.

Step 3 Build and train a surrogate model to predict the rating from a *Film DNA*. We will follow different *rules of thumb* to achieve a moderately good solution that serves our purposes. A multi-layer perceptron will handle the unknown relations between tropes and their combinations.

Step 4 Optimize the Film DNA with respect to ratings by building a Genetic Algorithm with specific operators that relies in the surrogate model previously built.

3.1 | Step 1: Extraction of tropes

We are going to use tropes as described in a live wiki called TV Tropes (2018), that is collecting thousand of descriptions and examples of tropes from 2014 until now. As the data is fed by a community of users, we could find the bias that popular films are better described and analysed in terms of the tropes than older or independent films, and that popular tropes are more recognised than very specific ones. Which means that, during the automatic generation of Film DNAs, tropes could be under/overrepresented, and that positive and negative estimation errors are possible. The semantic network of knowledge behind TV Tropes is huge and complex; it massively links hierarchies of tropes to their usage in creations for digital entertainment. The data, however, is only available through its web interface, which is why, in order to make it usable by the scientific community, Kiesel (2018) extracted all their data to a database so-called *DBTropes.org*. As the base of the research on automatic trope generation, we begun with a dataset based in the latest version of DBTropes, called PicTropes (García-Ortega et al. 2018) that included 5,925 films with 18,270 tropes. However, the last version of DBTropes is from 2016, and the community of users of TV Tropes has tripled the size of the database since then; in other words, we are not using it because it is outdated. If we work with the latest data from TV Tropes our machine learning algorithms would benefit from having much entries and hence, provide better results. That is why our first step is to extract the data directly from TV Tropes while making it available to the public and the researchers, in the context of the Open Science.

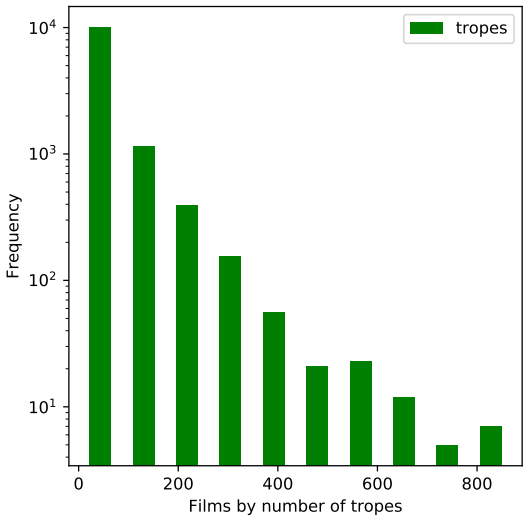
Our scraper, which is also released as free software in the Python ecosystem under the `tropescraper` name, and is also available from GitHub (`https://github.com/raiben/tropescraper`), extracts all the categories from the main categories page and, for every one of them, it extracts all the film identifiers assigned to it. Finally, for every film page, it extracts all the trope identifiers, building a dictionary of films and tropes. Trope identifiers are written in *CamelCase* format and may include the year to avoid ambiguity. Some technical details are listed in Figure 1.

The resulting dataset includes 11846 *Film DNAs* and 26246 tropes. In both cases, the number of tropes by film and film by tropes follow long tail distributions, where a large number of occurrences are far from the "head" or central part of the distribution, as shown in Figure 2. 60% of the films have 40 or less tropes but there are films with more than 800 tropes. On the other hand, most tropes appear in 6 films, but there are tropes with more than 3000 occurrences in films. These figures will have to be taken into account when we analyse the expected quality of the evaluator and the distribution of evaluation errors, and during the experimental setup, in order to make decisions according to the observed bias.

It is part of the current research to analyze the expected effect of this distribution in the results of applying our methodology. The first conclusion is that we have many more samples with a small number of tropes than with many; however, at this step we do not have enough information to elucidate if this situation is explained by the fact that it is user-generated data and the popularity defines how well described are the films in terms of tropes, but we can assume that, in general, that is the case. Furthermore, we cannot make out yet a relationship between the number of tropes in a film and its rating, but according to the Figure 2c, the films with the highest number of tropes are mostly last-generation superhero movies are popular and broadly acclaimed by the critic, and that suggests a positive correlation between rating and number of tropes. However, as the next section complements the tropes with additional information, such as the rating, the genres or the number of votes, we will be in a position where we can find correlations that help us explain the possible results of the experiments in a better way.

	tropes
count	11846
mean	53.3779
std	72.5272
min	0
10%	8
20%	14
30%	19
40%	24
50%	30
60%	40
70%	52
80%	72
90%	118
max	871

(a)



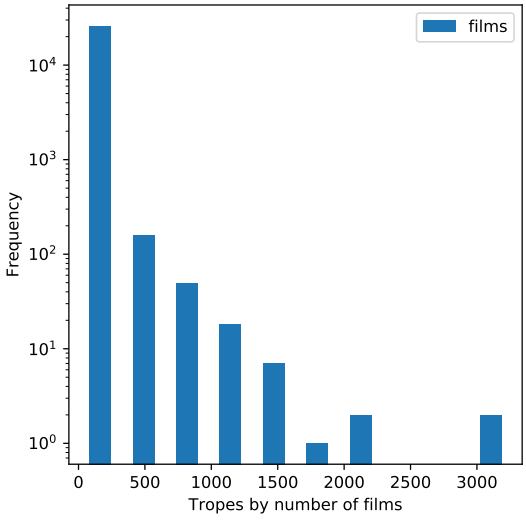
(b)

	tropes
XMenDaysOfFuturePast	871
XMenApocalypse	867
ThorRagnarok	860
Thor	843
TheDarkKnightRises	814
XMenFirstClass	813
CaptainAmericaTheFirstAvenger	792
AvengersAgeOfUltron	779
Deadpool2016	779
TheLastJedi	729
CaptainAmericaTheWinterSoldier	727
WhoFramedRogerRabbit	719
TheDarkKnight	681
Deadpool2	670

(c)

	films
count	26246
mean	24.0919
std	77.9833
min	1
10%	1
20%	2
30%	3
40%	4
50%	6
60%	9
70%	14
80%	24
90%	50
max	3267

(d)



(e)

	films
BigBad	3267
ShoutOut	3066
OhCrap	2108
ChekhovsGun	2011
Foreshadowing	1842
BittersweetEnding	1607
Jerkass	1586
DeadpanSnarker	1572
TitleDrop	1530
RunningGag	1391
DownerEnding	1391
LargeHam	1313
HeroicSacrifice	1184
LampshadeHanging	1176

(f)

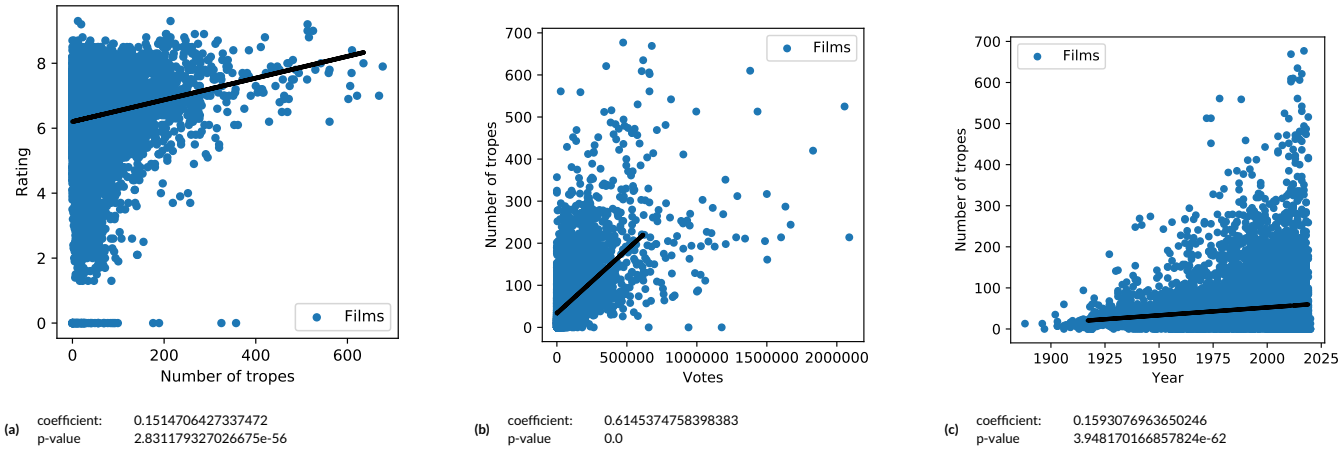
FIGURE 2 (a) Descriptive analysis of the Tropes by appearance in films. **(b)** Histogram of number of tropes by film (with logarithmic y axis). **(c)** Top films by number of tropes. **(d)** Descriptive analysis of the tropes by number of films in which they appear. **(e)** Histogram of number of films by tropes (with logarithmic y axis). **(f)** Top tropes by number of films.

3.2 | Step 2: Disambiguation of films to get the rating

TV Tropes is a huge yet very specific database of tropes but it does not include a rating or links to an external database that we could use as a rating source; on the other hand, IMDb offers their database for non-commercial use and they provide datasets with lots of interesting features, including the rating and the number of votes. Our research just needs a way identify a movie in TV Tropes with another in IMDb.

IMDb Datasets are a compendium of information that IMDb offers for personal and non-commercial use (IMDb Datasets 2019). Our current research will make use of these datasets to extend the film information from TV Tropes, in particular, *titles*, which contains metadata from the films such as the title, the year, the genres and the duration, and *ratings*, which contains the rating and the number of votes.

Items in IMDb that don't relate to films are excluded (tvEpisode, tvSeries, tvSpecial, tvShort, videoGame, tvMiniSeries, titleType) because they are not in our TV Tropes scraped dataset and they would only increase ambiguity as more films might match the same name. In order to be able to map the film names, films names are normalized in both cases, TV Tropes and IMDb, converting *CamelCase* format to *Title case*, removing non-alphanumerical values and extra blanks, splitting name and year when required, and converting to lowercase, considering the original title and the English title. Normalized names in TV Tropes and IMDb are matched, ideally {1->1}, but in practice, especially when the year is not declared, we tend to find a big list of candidates for every single film in TV Tropes. In order to reduce ambiguity, if the year is present in TV Tropes's identifier, we



(a) coefficient: 0.1514706427337472
p-value 2.831179327026675e-56

(b) coefficient: 0.6145374758398383
p-value 0.0

(c) coefficient: 0.1593076963650246
p-value 3.948170166857824e-62

FIGURE 3 Scatter plot (with fitting) to show the relation between: **(a)** Rating vs. # of tropes **(b)** Votes vs. # of tropes **(c)** Year vs. # of tropes

reduce the search to the specific year in IMDb, and, in any case, we select the candidate with the highest number of votes. This heuristic relies in the fact that both data sources (tropes and votes) are generated by different communities of users, enthusiasts in both cases, so if there is a film in TV Tropes and there are many films with the same name in IMDb, it will probably be the one with highest popularity, that is reflected in the number of votes.

First, we need to explore if there are significant biases in the original set, since it's generated by the user, and we have included features such as the rating, the number of votes and the year. For this reason, the Figure 3 shows the scatter plots of the relations between the number of tropes and the rating, the rating and the number of tropes, and the number of votes and the year in the new *extended dataset*, consisting in 10766 films linked to 26246 tropes. Figure 3b shows that there is a significant positive correlation between the number of tropes and the rating; in other words, a higher number of tropes is related to higher rating, and vice-versa. It is important to remark that the films up to $\sim$ 53 tropes (the average) or less, show the widest range, $\{\sim 1.3 - \sim 9.5\}$, for the rating, whereas films with more than 600 tropes have a rating in the range $\{6 - \sim 9.5\}$. If we want to test the methodology, choosing a fixed number of tropes lower or equal to 53, wuch as the median the median (30), which what we are doing, would allow us to find an optimal solution in a wider range while keeping the search space small. At the same time, a number of tropes in this range has an average rating in the range $\{6 - 6.5\}$, according to the regression function, which is good enough. No search algorithm will guarantee that a synthetic set of tropes will reach that average, but at least we know that we will be able to generate movies with good or excellent rating if we work with that specific number of tropes.

There seems to be also a significant positive correlation between the number of tropes and the votes (popularity) as well, which might explain the long tail distributions in the previous step, Figure 2; in other words, the more popular the film is, the better it is described by the community and the more tropes are found. This outcome implies that, if the experimental setup fixes the number of tropes for the synthetic Film DNA, it does not necessary imply fixing the complexity but fixing the detail, that probably ends up determining the evaluation error. Finally, there is a significant positive correlation between the number of tropes and the year, according to the scatter plot, the year of the film limitates the maximum number of tropes that describe the films in our dataset.

These three findings will help us unravel the limitations of a surrogate model fed from the *extended dataset* and used to predict the rating from a set of tropes. The analysis points out that choosing a small Film DNA will lead to films not very well described, with potential to have a rating in a wider range of values, but also a bigger chance to be evaluated with errors than films with more tropes.

As we stated before, the occurrence matrix of tropes in films is very sparse because most of the films in TV Tropes are described with just a few tropes whereas the minority have a huge number, and the reason is that it is user-generated data. However, we confirm that the films have a mean of 2.415 genres, so we consider that the estimation will benefit from considering the genres as features, especially when the films are poorly described. That is the reason why, although genres and tropes are different concepts, we will consider both in our *extended dataset* during the next steps, as they both serve to describe the story at different layers.

3.3 | Step 3: The surrogate model

We will use the *extended dataset*, to build a trope-to-rating approximator that can then be used as a surrogate model by the GA.

As explained in the Section 3.2, the tropes and genres have a relationship of belonging with the film, and hence, the proposed way to represent the input of the evaluator is a list of boolean whose indexes map to the list of possible tropes.

	activation	alpha	hidden layer sizes	learning rate	max iter	solver	mean	std
0	relu	0.0001	(162,)	constant	100	sgd	0.250449	0.00775062
1	relu	0.0001	(162,)	adaptive	100	sgd	0.247197	0.00662991
2	tanh	0.0001	(162,)	constant	100	sgd	0.242164	0.0088337
3	tanh	0.0001	(162,)	adaptive	100	sgd	0.240557	0.00857194
4	relu	0.0001	(883, 29)	constant	100	sgd	0.224481	0.00950363
5	relu	0.0001	(883, 29)	adaptive	100	sgd	0.216897	0.0208249
6	tanh	0.0001	(883, 29)	constant	100	adam	0.173729	0.04267
7	tanh	0.0001	(883, 29)	adaptive	100	adam	0.166007	0.0396141
8	tanh	0.0001	(883, 29)	constant	100	sgd	0.164613	0.0280966
9	tanh	0.0001	(883, 29)	adaptive	100	sgd	0.16356	0.0320443
10	tanh	0.0001	(162,)	adaptive	100	adam	-0.132811	0.0604104
11	tanh	0.0001	(162,)	constant	100	adam	-0.136999	0.0436053
12	relu	0.0001	(883, 29)	constant	100	adam	-0.444568	0.100033
13	relu	0.0001	(883, 29)	adaptive	100	adam	-0.450668	0.203108
14	relu	0.0001	(162,)	constant	100	adam	-0.708584	0.0805029
15	relu	0.0001	(162,)	adaptive	100	adam	-0.723516	0.0966577

TABLE 1 Hyper-parameters evaluation using 3-fold cross validation, sorted by validation score

The output is a continuous numeric value that represents the rating of the film, theoretically from 0 to 10. As the average number of tropes by film is 53, and the possible number of tropes is 26246, 99% of the cells in the matrix will have a value of 0, in other words, it will be very sparse, with a small representation of the tropes in the catalogued films. The deep learning technique most suitable in this context needs to expose feature-extraction capabilities in order to deal with unknown and unbalanced relations between tropes to achieve a specific rating. Although there are different candidates that could perform properly under these circumstances, in our current research we choose a neural network (Schmidhuber 2015).

The goal of our research in this initial stage is to evaluate a methodology, so tuning the surrogate model is carried out as far as it suits the needs in terms of quality of the estimations, with reasonable performance and a low error rate. There are many decisions that can define the quality of the model; some of them will be made based on the state of the art and, for others, we will have to make hyper-parameters search. In general, although there are many rules of thumb to build acceptable neural networks, results may differ drastically depending on the nature of the problem and it is recommended to do a hyper-parameters evaluation. We selected the multi-layer perceptron (MLP), the most widespread neural network architecture, because it has been proven to be able to approximate any function that we require, the so called *Universal Approximation Theorem* Hornik, Stinchcombe, and White (1989). In order to choose hyper-parameters that get along with the nature of our problem we did a preliminary search with all the combinations in a domain of possible values for the activation (ReLu or tanh), the number of hidden layers (1 or 2), the number of neurons in each layer (162 or 883/29) according to the geometric pyramid rule proposed by Masters (1993), the learning rate (constant or adaptive) and the solver (Adam or SGD). We applied 3-fold cross validation and obtained the average and the standard deviation.

The results in Table 1 show that a MLP with the structure '[26273/162/1]', using ReLu activation, constant learning rate and SGD solver provides the best validation score. After training the MLP using the selected hyper-parameters, the *extended dataset* as input and the rating as output, until it does not improve more than the tolerance for 10 consecutive runs, the evaluation converged to a training mean squared error (MSE) of 0.410 and the a validation MSE of 0.357, that implies that the model is a good predictor as both values are quite similar, and that it is not overfitting. The root mean squared error (RMSE), that is a metric in rating units, has the value of 0.597 for new predictions, that implies that in some cases the evaluations might be above 10.

The multi-layer perceptron will be used as a surrogate model in a specific experiment in further sections, so, in order to anticipate the results, we need to analyze how well it predicts the ratings grouped by the number of tropes, since we will have to limit them during the optimization process. We calculate the error for every sample of our *extended dataset*; Figure 4(a) presents it as a hexagonal binning plot. This plot shows that most of the cases are positioned close to an error of value 0 (high density). There are under and over-estimations, especially in the range of tropes with more occurrences {0-100}, but the it's more likely that rating is under-estimated; in general, the surrogate model will tend to under-estimate for a small number of tropes, and estimation will be increasingly accurate when the number of tropes is increased. This is consistent with the fact that movies with more tropes are more popular and they are probably better described.

Figure 4(b) and (c) shows the average absolute error and absolute standard deviation by the number of tropes. Again, we can observe that a small number of tropes (less than 200) implies higher errors and higher deviations than a bigger number of tropes. That is explained because we

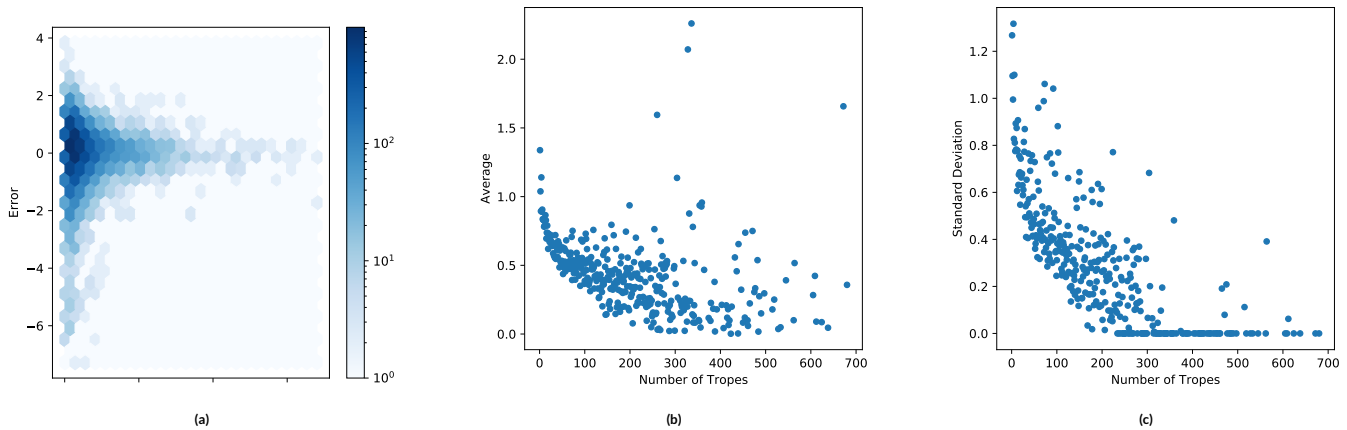


FIGURE 4 (a) Hexagonal binning plot of the evaluation errors for all the films (color in logarithmic scale). **(b)** Average absolute value of the error by number of tropes. **(c)** Standard deviation of the absolute error value by number of tropes

are working with user-generated content and less popular films are poorly described in terms of number of tropes, hence their rating is harder to predict. At the same time, it points out that the number of tropes is a good predictor, and, if the films were described with an equal level of detail, it would imply more tropes and less errors.

Analysing the weight of the genre on the rating is also a good outcome in order to know if there are rating biases regarding the genre. We did the evaluation of all the Film DNAs that contain only one genre out of the 27, getting an average value of 5.511(+-0.010), so, in principle, choosing a single genre or another does not seem to limit the rating. Intuitively, that tells us that all the genres have good and bad films in the same proportion and our synthetic Film DNA will not suffer from strong biases due the genre. However, it is the combination of genres what might boost the rating, and this is precisely where the optimization process comes into play.

3.4 | The Genetic Algorithm

At this step we already have a set of candidate tropes for the Film DNA and the surrogate model trained and evaluated in the previous step, so our goal is to use a system that allows us to generate a Film DNA that optimizes the rating. We determine to use a GA because it is a mechanism that allows us to explore the domain of Film DNAs, in other words, of combinations of tropes, getting high-quality solutions and dealing with global optimization problems.

Our chromosome will be the Film DNA, that is, a set of tropes and genres. In practice it will be encoded as an array of different indices on a dictionary of the total set of tropes available without value repetitions, given that, *a priori*, the order of the tropes is irrelevant (according to their nature, they may refer to specific moments, but also to narrative structures and general settings) and also, our rating evaluator does not consider weights or multiple occurrences of the trope in the movie (both, training data and model do not consider multiple occurrences of a trope in a single Film, in other words, the trope appears or does not appear). In practice, it will be a list of 30 tropes and genres with no repetitions.

The mutation operator changes a trope of the Film DNA by another that is randomly chosen from the set of tropes not included of the Film DNA, which avoids repetition of tropes/genres, and allows an exploration of new tropes. The crossover operator will make a superset of tropes from the parents' Film DNAs and randomly selects two subsets for the offspring. This way, the offspring will have Film DNAs whose tropes are exclusively from their parents, allowing the exploitation of the data.

As we explained above, we have to rely on a surrogate model for the evaluation of Film DNAs, and our approach uses a neural network trained with existing movies, the one that has been presented in the previous section. The fitness of our GA will be the result of evaluating the set of tropes and genres the neural network, that is, their predicted rating.

According to our tests, the GA has proven to converge towards optimal solutions efficiently, given the simplicity of the operators based on set algebra and fitness calculation. It is important to remark that the GA performs an optimization step that is essential for the research as far as it leads us to our goal, that is to prove that the methodology works, and hence, an exploratory analysis of the parameters of the GA is reasonable in this context, as we will see in the next section, 4.

4 | EXPERIMENTAL SETUP

The evaluation of the methodology is performed unequivocally through its testing and a further analysis of the results. In this research, we want to know if the methodology can be used to synthesise an optimal Trope DNA for a standard movie in terms of potential rating.

Our experiment will use a fixed size Film DNA of 30 tropes, which is the median of tropes per film in the original dataset and therefore, a candidate for a 'standard' film. As benefits, a Film DNA of size 30 is easier to interpret in natural language than one of size 200. Furthermore, as we saw in the Section 3.2, this number of tropes is a good candidate because it can lead to a wide range of ratings, good and bad, and the GA will have a better chance to evolve solution from bad cases; however, as we saw in the Section 3.3, a Film of size 30 implies more chances to under or over-evaluate, and the rating will be less accurate. We also use a fixed length chromosome because at this stage we were only interested in a proof of concept and, otherwise, the results could be harder to explain and interpret.

The most suitable parameters to solve a specific problem require an extensive calibration, which is outside of the scope of this paper. However, to choose the set of parameters that we will use in all the experiments in this paper, we have made a preliminary selection executing 30 times every possible combination of Population (P) = {50, 100, 200}, Mutation probability (Mp) = $\{2 \div \text{size}_{dna}, 1 \div \text{size}_{dna}, 0.5 \div \text{size}_{dna}\}$ and Crossover Probability (Cp) = {0.25, 0.5, 0.75} and we have chosen the combination P = 200, Mp = $1 \div \text{size}_{dna}$, Cp = 0.5, because it obtains better results on average. In the experiments we have set the stop criterion to a minimum of 3000 evaluations and until the best generation does not improve during 10000 evaluations. We will run the GA 30 times with different random seeds and we will select the run that gives the highest fitness in order to analyse it.

As previously stated, a set of tropes is not a full-fledged film, so we need to have an idea of what a movie with that set of tropes would look like. This is why, as tools to interpret the results, we will use metrics of similarity between finite sample sets. The first one is a metric called *Jaccard coefficient* and is defined as the cardinality of the intersection of Film DNAs divided by the cardinality of the union of the Film DNAs. The Jaccard coefficient is interesting to measure not only what two sets have in common, but also, how different their sizes are, penalizing big differences in length of the sets. However, as we will be comparing our synthesised Film DNA of a fixed size of 30, popular films with hundred of tropes will be penalized just because they are too broadly described in comparison; In order to address this problematic, we will also use a coefficient based only in the common elements and is defined as the intersection between the sets divided by the length of our synthetic Film DNA.

5 | RESULTS

The goal of this specific experiment is to check if the surrogate model can be used successfully for optimizing a Film DNA by its potential rating. According to the experimental setup in Section 4, the selected length of the Film DNA is 30, as it has been analysed to be a good candidate. The Film DNA will include tropes and genres, as discussed previously in Section 3.2, because the estimation will benefit from considering the genres as features, specially when the number of tropes is small.

The repetition of the GA 30 times has resulted in an average fitness of 9.783 (+-0.431). The best solution across all runs has a rating of 10.313 and a Film DNA with the following tropes.

$$\begin{aligned}\text{DNA}_{\text{Film}} = \{&\text{ActionHeroBabysitter}, \text{DeathByFlashback}, \text{DisneyVillainDeath}, \text{DuelToTheDeath},\\ &\text{EarlyBirdCameo}, \text{FightingFromTheInside}, \text{HandsOffParenting}, \text{Homage},\\ &\text{ImNotAfraidOfYou}, \text{JumpCut}, \text{MouthingTheProfanity}, \text{NoSympathy}, \text{OminousFog},\\ &\text{OneHeadTaller}, \text{PoorMansSubstitute}, \text{PragmaticAdaptation},\\ &\text{RichIdiotWithNoDayJob}, \text{SomeoneToRememberHimBy}, \text{SpitefulSpit}, \text{TalkingHeads},\\ &\text{TitledAfterTheSong}, \text{WeaponOfXSlaying},\\ &[\text{GENRE}]\text{Animation}, [\text{GENRE}]\text{Documentary}, [\text{GENRE}]\text{Drama}, [\text{GENRE}]\text{History},\\ &[\text{GENRE}]\text{Mystery}, [\text{GENRE}]\text{Romance}, [\text{GENRE}]\text{War}, [\text{GENRE}]\text{Western}\}\end{aligned}$$

The synthetic Film DNA belongs to a multi-genre film (historical documentary romantic animation drama, set during a war, that includes western and mystery settings, not comedy). However, according to one of the tropes, even if it is historical/documentary, it is completely an adaptation from the author with clear differences with the overall known story for pragmatic reasons (*PragmaticAdaptation*). The following explanations are derived from the current definitions in TV Tropes and, in any case, they should be interpreted as just one example out the vast number of possible implementations, specially considering that tropes and genres evolve through time and are continually discovered, adapting to the new films and cultural trends.

In the synthetic Film DNA, there are tropes that define the main characters: one of them is a rude action character that, in this case, is compelled to have a position of responsibility for children (*ActionHeroBabysitter*), that could be related to the existence of irresponsible parents (*HandsOffParenting*). There is a female character, as required by the trope *SomeoneToRememberHimBy*, and there is a couple, according to *OneHeadTaller*, where one character is clearly taller or shorter than the other. One of the characters is rich and has a lot of free time (*RichIdiotWithNoDayJob*), and, according to the trope *DisneyVillainDeath* there is a villain as well. There are tropes that define the conflicts and how they will develop: characters resisting an external influence acting on them (*FightingFromTheInside*), a villain is finally fought by a protagonist when this character is *not afraid* anymore (*ImNotAfraidOfYou*), there is a duel (*DuelToTheDeath*), a special weapon is used (*WeaponOfXSlaying*), the villain falls off (*DisneyVillainDeath*), one of the protagonists heroically dies as well, and in the end it is discovered that the female protagonist is or was pregnant, as required by the trope *SomeoneToRememberHimBy*. There are also tropes that define the setting: there is fog (*OminousFog*) as part of the mystery genre and a clear *'Homage'* to a classic or well known artwork in the same genres is present. According to the narrative perspective, the story includes a flashback that points to the death of the main character/s (*DeathByFlashback*), uses 'Jump Cuts' as an editing technique (*JumpCut*) and make a character appear before his/her introduction (*EarlyBirdCameo*). Some tropes also define very specific sequences: in some cases the film includes spits (*SpitefulSpit*) and the characters swear, although it cannot be heard by the audience (*MouthingTheProfanity*). There is also a scene where terrible things have happened to the main characters but no-one acts as it is really important (*NoSympathy*), and there are scenes with no action, just long conversations where people do not move from their place (*TalkingHeads*). According to the meta-tropes present, the action character use to be an *action hero* in previous movies, and, in contrast, in the current one the character deals with unfamiliar problems and domestic situations. Also, one of the actors is relatively unknown but looks alike another well known one (*PoorMansSubstitute*) and the film's name is a reference to an existing song (*TitledAfterTheSong*). As we mentioned above, the film DNA has 8 genres (out of 27), that can be seen as a high number for a film, but there are examples of TV series with a high number of genres as well, like Dr. Who or Stranger Things, with 5 genres each.

As part of the analysis, we need to confirm the originality of the Film DNA, so we perform a similarity analysis against the whole *extended dataset* and we find that the coefficients are small: According to the Jaccard metric, the most similar films have a value of 0.1, in other words, they are 10.000% similar at maximum. According to the Common Elements metric, the most similar films have a value of 0.167 that is higher than the Jaccard coefficient because it is not penalizing the difference of length of the Film DNAs, in other words, the maximum number of common tropes/-genres between the whole *extended dataset* and the synthesised Film DNA is 5. There are only 6 films in the whole dataset with that similarity ('Arrival', 'Senso', 'Richard III', 'Jane Eyre', 'Lantana' and 'Bridget Jones: The Edge of Reason'), presenting all of them the genre Drama and the tropes PragmaticAdaptation, most of them the genre Romance, and to a lesser extent the genre Mystery and the trope Homage.

Our synthesised Film DNA not only has a very high potential rating but also is very different to the rest of the films in our *extended dataset*, that is positive because it proves that the GA is exploring combinations that are far from those already realized in real life

6 | CONCLUSIONS

Tropes in films and their relations can be seen as an adaptive, un-predictable and dynamic complex system, with an emerging behavior that would be the overall effect provoked on the public, that can eventually be measured by the rating given to them. Optimizing such a complex system can be approached from a number of different ways, which is why the goal of this research is to evaluate a methodology that automatically synthesises sets of tropes in a way that they maximize the potential rating of a film that uses them. In order to do so, we have first extracted all the tropes from the TV Tropes sites and the genres and ratings from IMDb, merging them in a single *extended dataset* and dealing with technical challenges as the devise of heuristics to disambiguate film names and the consequent analysis of distribution of the data, that affects the further steps. The scraped we have used has been released in the Python ecosystem as TropeScraper `https://pypi.org/project/tropescraper/`. This will allow easy updating of the tropes database for new experiments. Our research group also supports open science, so the results of all experiments and the data used in them is published in `https://github.com/raiben/made_recommender`.

We used the dataset to train a MLP that predicts the rating from a Film DNA from its set of tropes and genres, and used it as the surrogate model in an evolutionary algorithm that tries to find a set of tropes and genres that maximizes rating; this evolutionary algorithm uses a fixed-length set of tropes and is able to find a set with a rating that is, in fact, higher than the maximum rating found in the training set.

This first work on the subject establishes the methodology through a proof of concept, and will also be used to establish a series of baseline measurements that can eventually be used to compare with further developments of this methodology. This proof of concept already proves that the GA that uses a MLP as surrogate model to evaluate fitness of sets of tropes is able to find genuine Film DNAs with a very high potential rating, and is able to deal with the intrinsic difficulties of working with this *complex system* of films, tropes, genres and ratings.

The best result, which includes several genres among the synthetic film DNA, shows that, even if the individual contribution of every genre to rating is not too important, combining several genres, even more so in an environment such as this proof of concept where the number of tropes/-genres in the chromosome is limited, makes sense and boosts the rating, on average (9.7) and also in the best case analyzed, which includes 8 genres. The effect of using a different number of tropes, even unlimited, or limiting the genres or not including them at all, is unknown and interesting as these changes could affect the expected rating and the applicability of the tropes. This is left, however, as future work. On the other hand, building a Film DNA might be the most direct application of the methodology, and a good fit for a proof of concept but there are other applications that could have explained the boundaries of the proposal and the limitations of the surrogate model, such as improving an existing film or mixing films.

We also conclude that the tropes and the genres are good estimators of the quality and that the neural network architecture chose, a MLP, works well as rating estimator, with a low error rate that decreases as the number of tropes grow. However, a MLP is a black box, hard to optimize and to explain, so we wonder how using other techniques could affect the rating and the possible applications; this exploration is also left as future work.

Moreover, during this proof of concept we discovered limitations of the dataset that could compromise the quality of the evaluator: first of all, due the very nature of the user-generated data, the most popular and new films get a more detailed description in terms of tropes, hence the films with less tropes are more prone to be evaluated with errors. We decided to add the genres to mitigate this effect on unpopular films, but other techniques to reduce the dimensionality could have help us, making the dataset more dense and reducing the time to train the surrogate model. However, we decided not to use them because they would bring more complexity to the proof of concept.

A synthetic Film DNAs directly applicable to help on the preparation and design of a film as we shown in the results section, because every trope or genre has a different and specific translation according to TV Tropes: in some cases, it affects to the structure, narration, characters, elements, actions, moments and/or places in a general or specific way. in other cases, it affects to elements out of the story, like the casting, direction guidelines or production. So essentially, some tropes and genres will be achievable and other will not, and this brings two new questions: what to do if a trope needs to be avoided in a solution and what to do if a trope is required to complement the story. Our assumption is that both questions can be tackled through the modification of the GA operators but that would also benefit from having explicit dependencies between the tropes and the genres. At the same time, we want to remark that the final application of a Film DNA in the creation process, working with the coherence among tropes and genres, filling the gaps with elements in the story and narrating them in a compelling way is still required, not part of the scope of the current research and, as far as we know, mostly a human responsibility that the final rating depends on.

All previous considerations may open a promising and encouraging research line, hence we propose that further research should be undertaken in different areas. First, the GA could be improved by implementing constraints such as limiting the genres, dealing with tropes dependencies, enforcing the appearance of tropes/genres, penalizing tropes/genres or promoting certain Film DNA similarity. This way, further experiments could help us define the problems that the methodology can address, for example, allowing the generation without trope number restrictions, improving an existing Film DNA in a way that some tropes are replaced but a minimum similarity is achieved, mixing the tropes of two or more films in a way that the resulting film DNA keeps the essence of them while maximizing the rating, or even massively generating film DNAs that share common tropes, leading to backstories that cross in time, and ultimately, can help us populate the virtual world of a videogame.

The surrogate model could be improved as well in any number of different ways; in order to simplify the arithmetic of tropes/genres in the set and explain hidden relations between them, further researches could implement a technique similar to word2vec, assigning a vector in a multi-dimensional space to every trope/genre, in a way that the ones that share common contexts in the corpus are located close in the space. Also, the quality of the dataset could be improved by applying techniques to reduce the dimensionality, such as *principal components analysis* and *feature selection*, improving the performance of the evaluator.

Finally, any further research focused in automating the translation of tropes to coherent actions in a virtual world would be a decisive contribution to make the results suitable for a content generator of stories.

ACKNOWLEDGEMENTS

This work has been partially funded by projects DeepBio (TIN2017-85727-C4-2-P) and TEC2015-68752 and “Ayuda del Programa de Fomento e Impulso de la actividad Investigadora de la Universidad de Cádiz”.

7 | BIBLIOGRAPHY

References

Baldick, C. (2015). *The Oxford dictionary of literary terms.* OUP Oxford.

Bitsilli, P. M. (1983). *Chekhov's art, a stylistic analysis.* Ardis.

Bui, V., Abbbass, H., & Bender, A. (2010). Evolving stories: Grammar evolution for automatic plot generation. In *Ieee congress on evolutionary computation* (pp. 1–8).

Campbell, J. (2008). *The hero with a thousand faces* (Vol. 17). New World Library.

Delmas, G., Champagnat, R., & Augeraud, M. (2007). Bringing interactivity into Campbell's hero's journey. In *International conference on virtual storytelling* (pp. 187–195).

Field, S. (1982). *Screenplay*. Delacorte New York.

García-Ortega, R. H., García-Sánchez, P., Merelo Guervós, J. J., Arenas, M. I. G., Valdivieso, P. Á. C., & Mora, A. M. (2015). How the world was MADE: parametrization of evolved agent-based models for backstory generation. In A. M. Mora & G. Squillero (Eds.), *Applications of evolutionary computation - 18th european conference, evoapplications 2015, copenhagen, denmark, april 8-10, 2015, proceedings* (Vol. 9028, pp. 443–454). Springer.

García-Ortega, R. H., García-Sánchez, P., Merelo Guervós, J. J., Ginés, A. S., & Cabezas, Á. F. (2016). The story of their lives: Massive procedural generation of heroes' journeys using evolved agent-based models and logical reasoning. In G. Squillero & P. Burelli (Eds.), *Applications of evolutionary computation - 19th european conference, evoapplications 2016, porto, portugal, march 30 - april 1, 2016, proceedings, part I* (Vol. 9597, pp. 604–619). Springer.

García-Ortega, R. H., Merelo-Guervós, J. J., Sánchez, P. G., & Pitaru, G. (2018). Overview of PicTropes, a film trope dataset. *arXiv preprint arXiv:1809.10959*.

Gervás, P. (2013). Propp's Morphology of the Folk Tale as a Grammar for Generation. In M. A. Finlayson, B. Fisseni, B. Löwe, & J. C. Meister (Eds.), *2013 workshop on computational models of narrative* (Vol. 32, pp. 106–122). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. Retrieved from http://drops.dagstuhl.de/opus/volltexte/2013/4156 doi: 10.4230/OASIcs.CMN.2013.106

Guarneri, A., Ripamonti, L. A., Tissoni, F., Trubian, M., Maggiorini, D., & Gadia, D. (2017). GHOST: a GHOst STory-writer. In *Proceedings of the 12th biannual conference on italian sigchi chapter* (p. 24).

Guzdial, M., Harrison, B., Li, B., & Riedl, M. (2015). Crowdsourcing open interactive narrative. In J. P. Zagal, E. MacCallum-Stewart, & J. Togelius (Eds.), *Proceedings of the 10th international conference on the foundations of digital games, FDG 2015, pacific grove, ca, usa, june 22-25, 2015*. Society for the Advancement of the Science of Digital Games.

Hennig-Thurau, T., & Houston, M. B. (2019). Entertainment product decisions, episode 1: The quality of the entertainment experience. In *Entertainment science: Data analytics and practical theory for movies, games, books, and music* (pp. 295–312). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-89292-4_7

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Hsu, P.-Y., Shen, Y.-H., & Xie, X.-A. (2014). Predicting movies user ratings with IMDb attributes. In *International conference on rough sets and knowledge technology* (pp. 444–453).

Imabuchi, S., Akimoto, T., Ono, J., & Ogata, T. (2012, November). KOSERUBE: An application system with a Propp-based story grammar and other narrative generation techniques. In *The 6th international conference on soft computing and intelligent systems, and the 13th international symposium on advanced intelligence systems* (pp. 248–253). doi: 10.1109/SCIS-ISIS.2012.6505320

IMDb Datasets. (2019). https://www.imdb.com/interfaces/. [Online; accessed on 11-July-2018].

Juarrero, A. (2000). Dynamics in action: Intentional behavior as a complex system. *Emergence*, *2*(2), 24–57.

Kiesel, M. (2018). *DBTropes – skipforward*. http://skipforward.opendfki.de/wiki/DBTropes. [Online; accessed on 11-July-2018].

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann.

Mellina, C., & Svetlichnaya, S. (2011). *Trope propagation in the cultural space*. Stanford CS224W: Social and Information Network Analysis, Autumn. Retrieved from http://snap.stanford.edu/class/cs224w-2011/proj/staceys_Finalwriteup_v1.pdf

ois Truffaut, F. c., Hitchcock, A., & Scott, H. G. (1985). *Hitchcock*. Simon and Schuster.

Pang, B., & Lee, L. (2007). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135. Retrieved from https://doi.org/10.1561/1500000011 doi: 10.1561/1500000011

Propp, V. (2010). *Morphology of the folktale* (Vol. 9). University of Texas Press.

Sack, G. A. (2014). Character networks for narrative generation: Structural balance theory and the emergence of proto-narratives. In *Complexity and the human experience. modeling complexity in the humanities and social sciences* (pp. 81–104). Pan Stanford Publishing Pte. Ltd.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.

Thompson, M., Padget, J., & Battle, S. (2018). Governing narrative events with tropes as institutional norms. In P. R. Lewis, C. J. Headleand, S. Battle, & P. D. Ritsos (Eds.), *Artificial life and intelligent agents* (pp. 133–137). Springer International Publishing.

TV tropes. (2018). https://tvtropes.org/. [Online; accessed on 11-July-2018].

Vogler, C. (2007). *The writer's journey*. Michael Wiese Productions Studio City, CA.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, *4*(2), 65–85.

Yannakakis, G. N., & Togelius, J. (2018). Generating content. In *Artificial intelligence and games* (pp. 151–202). Springer International Publishing.

How to cite this article: R.H. García-Ortega, P. García-Sánchez, J.J. Merelo (2019), StarTroper, a film trope rating optimizer using Deep Learning and Evolutionary Algorithms, , .