

The Physics of Culture

Predicting Success of Screenplays

ScriptBook

Epagagogix

Neural Networks

Artificial Intelligence

Alogorithms

Collection of Articles & Links

By

John Fraim

Desert Screenwriters Group

<http://desertscreenwritersgroup.com>

johnfraim@mac.com

ScriptBook

<http://scriptbook.io>

OUR MISSION

ScriptBook assists stakeholders in the film business with their greenlighting decisions by providing an algorithmic assessment of a script's box office potential prior to financing, producing & releasing a film.

Research shows that 87% of films lose money at the box office. At ScriptBook our aim is to help profitable films become the rule and no longer the exception, without giving up on creativity and compelling storytelling.

Our decision support system guides film studios, production companies, talent agencies, distributors & investors in selecting film projects with high commercial value by providing them with automated analysis & box office forecasts using nothing more than a script.

<http://www.engadget.co>

https://www.youtube.com/watch?v=C00QU-8S_yM

<http://www.dnaindia.com/entertainment/report-could-this-software-fine-tune-bollywood-s-scripts-top-screenwriters-are-sceptical-2154937>

<https://www.linkedin.com/in/azermainadira>

<https://www.linkedin.com/in/bartdm>

https://mobile.twitter.com/scriptbook_io

<http://www.teleread.com/writing/scriptbook-predicts-screenplay-success-novels-next/>

UPPING BOX OFFICE PROFIT AND BANKABILITY WITH HARD SCIENCE

Justine Harcourt de Tourville

Hardcourt de Tourville

29 JAN 2015

DEMAND FOR FILMS CONTINUES

Despite a weak economy in 2013, the global box-office reached a record high of nearly \$36 billion.

According to PriceWaterhouseCoopers, the global box-office will hit \$45.9 billion in 2018, in part due to expanding markets in China, and sustained growth in the US, UK and Japan.

In total, worldwide filmed entertainment revenue is projected to surpass \$100 billion in 2017.

MAKING FILMS REQUIRES SIGNIFICANT RISK

Deciding which script to greenlight carries tremendous financial responsibility—**average cost to produce, make and release a film is \$74.9 million** (MPAA 2011).

Box office disasters can be felt across the industry: movie studio stock prices fall, studio executives and literary agents are fired, actors & directors lose bankability.

Gigli, a 2003 film that cost \$75.6 million to produce, only earned \$7.2 million: "This is not just ordinary bad ... but a hypnotic, black hole of a movie that sucks reputations, careers and goodwill down its vortex." —Liam Lacey, *The Globe and Mail* of Toronto

The Lone Ranger, which came out ten years later in 2013, had to pare down the budget to be made for \$215 million. It failed miserably garnering scathing reviews (projected loss between \$160-190 million). The chief analyst for Boxoffice.com, Phil Contrino, called it "the kind of bomb that people discuss for years to come."

BLOCKBUSTERS SHOULDER THE BURDEN

Studies by DeVany & Walls have shown:

Overwhelming majority of films are bad bets: 87% lose money.

Few subsidize entire industry: 80% of Hollywood's total profit earned by a mere 6.3% of the films.

THE SCRIPT FACTOR

Many factors can contribute to films failure to earn a profit: Bad acting, bad directing; high production, distribution and/or marketing costs. But one of the better predictive factors to a film's success is the screenplay.

In his book, *Great Flicks: Scientific Studies of Cinematic Creativity and Aesthetics*, Dean Keith Simonton demonstrates statistically how an Oscar-winning screenplay has the highest correlated factor ($r=0.63$) to positive critical evaluations than any other element: directing, acting, picture, song, etc.

In other words, scripts have an inherent predictive value.

FINDING THE SCRIPT

The film industry relies on outdated and error-prone methodology when considering scripts.

In 2012 in the United States, a total of 677 movies were released in theaters (MPAA 2013). That figure compares to—using the best guess of industry insiders—the yearly “25,000 to 40,000 of scripts that filter through the studio acquisition system.” (Scott Meyers 2012)

The predominant method of acquiring a script involves literary agencies; they in turn employ story analysts to read scripts and make recommendations in what is

known as "coverage." Scripts with positive coverage are then forwarded to the film studio or production company for consideration.

Whether a script should be pursued starts with a subjective decision based on the story-analyst's personal preferences. Some high-potential scripts never make the cut, while others (which will eventually bomb at the box office) are recommended.

*“The most famous dictum about Hollywood belongs to the screenwriter William Goldman. ‘Nobody knows anything,’ Goldman wrote in Adventures in the Screen Trade ... ‘Not one person in the entire motion picture field knows for a certainty what’s going to work. Every time out it’s a guess.’ One of the highest-grossing movies in history, Raiders of the Lost Ark, was offered to every studio in Hollywood,...and every one of them turned it down except Paramount: ‘Why did Paramount say yes? Because nobody knows anything. And why did all the other studios say no? Because nobody knows anything. And why did Universal, the mightiest studio of all, pass on Star Wars? ... Because nobody, nobody—not now, not ever—knows the least goddamn thing about what is or isn’t going to work at the box office.’” — from Malcolm Gladwell's article "The Formula" in *The New Yorker* 16 October 2006*

HOLLYWOOD’S PATCH

In 2005, an executive with Leonard diCaprio’s production company asked development executives about the best script (not in production) they read in the past year. The results were tallied; “The Black List” was launched, creating an internal industry ranking system among those with the expertise—and capacity—to greenlight a script and put it into production.

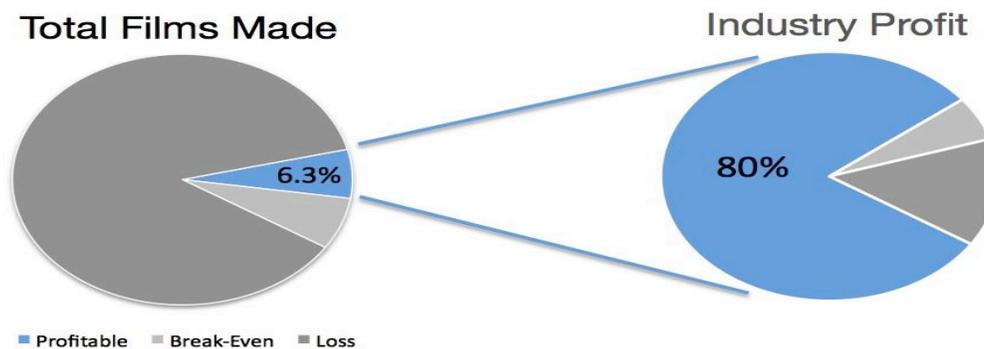
The Black List has grown in size and reputation (up to 500 executives participate in the tally). It has its own official blog and script-reading service. But as Indiewire.com reports, The Black List’s track record is the same as chance: “scripts near the top have been made into terrible movies, and those in the lower reaches have proven to be among the best.”

A TECHNOLOGY SOLUTION

Movie star Will Smith is a believer in pattern. According to a MIT paper on

prediction and customer wants, Smith methodically reviews the box office top performers to look for patterns and identified trends that helped him choose movies. His strategy paid off. His 2007 film *Hancock*, suffered terrible reviews but still earned \$625 million worldwide. Smith attributed his selecting films according to trends in high box office performance: special effects, special effects with creatures, and special effects with creatures and a love story—and chose films with these elements with grand success. (Davenport and Harris MIT Sloan Review Winter 2009).

Similarly, film production companies & studios need a decision support tool for scripts in the greenlighting stage. In order to capture more of the commercial potential between what is made and what is successful, increasing the odds of an investment return begins with better scripts.



Most academic research focuses on forecasting box-office performance *after* a film has been produced and sunk costs have already been incurred. The industry needs an objective tool to assess a script's commercial potential *prior* to production, investment and sunk costs.

PREDICTIVE SUCCESS

Other creative industries have already looked to predictive intelligence to deal with subjectivity in finding commercially-viable work with success. With revenue diminishing by 50% over the past 15 years, the music industry had to explore other channels to find hit music. MusicXray created a web application that performs multivariate analysis on a song based on elements like beat, melody, harmony, etc.

MusicXray can predict hits with 80% accuracy. For the music industry, minimizing the subjective guessing of A&R executives in charge of locating talent, as well as the risk of signing unknown artists, was an economic advantage in lean times.

FIRST ATTEMPTS IN FILM

Epagogix a U.K. company led by Nick Meaney relies on neural networks to make predictive analyses about films that should or should not go into production by looking for script elements that correspond with either success or failure at the box office. They try to assess risk, for example, by looking at factors such as cast and location. **Epagogix has found that A-list actors and directors are for the most part irrelevant to a film's bottom line.**

The highly successful *The Pirates of the Caribbean* and the ill-fated *Lone Ranger* films illustrate this finding. Both films were produced by the same studio (Disney), same director (Gore Verbinski) and boasted the same star (Johnny Depp). While the same writing team from *Pirates* worked on the *Ranger* film in the development at the mid-point, the *Lone Ranger* was subjected to numerous and substantial rewrites. The script was considered significantly weaker.

One problem with neural network analysis, however, is the immense need for input. Outcomes are rooted in data generated from screenplays past—problematic for cultural evolution. Data from neural networks, for example, would predict James Bond needs to be a white male to perform successfully at the box office; yet, the black actor Indris Elba is currently the subject of a public petition promoting his selection as the new 007.

LOWERING RISK FOR HIGHER REWARDS

Concentrating exclusively on storyline, with highly accurate predictive success, Nadira Azermai and Bart De Maesschalck founded ScriptBook in Belgium. Spe

At its core, ScriptBook's predictive system relies on textual information, but integrates screenwriting expertise so it is able to distinguish characteristics between genres. A comedy will not have the same considerations as a drama, nor an indie-cult film the same as an action thriller; creative features are incorporated into the evaluative process. ScriptBook is a product designed to be a tool for the

numbers-centric studio heads, in addition to filmmakers.

ANALYSIS

Just as every sector has been challenged by technology, creative industries are not immune to the shifts and change instigated by the digital world.

With funds tightening, studios either cut back or hedge bets with franchises. Either way, investors are increasingly careful about the size and frequency of their investments. Even though formulaic films are viewed as "safe" (the story has already been proven), film franchises steadily decrease their box office return over time (witness: *Indiana Jones*, *Star Trek*, *The Hangover*, and *The Pirates of the Caribbean*)—ultimately leading to "franchise fatigue," which *The Hollywood Reporter* attributes as the source of 2014's poor year in film:

*"The only thing the theater business needs is more movies; big four-quadrant movies [meaning appeals to all demographics, ed.], quiet dramas, niche comedies — you name it," says Patrick Corcoran, V.P. of the National Association of Theatre Owners. "There were 45 percent fewer \$100 million- budgeted movies in the summer, and box office was off 15 percent. That's the entire difference between 2013 and 2014." —from Pamela McClintock's article "What's Behind 2014's Box-Office Slide: Franchise Fatigue, Fewer Big Movies" in *The Hollywood Reporter* January 7, 2015*

Demand is there, but originality and breadth are missing.

Jeffrey Katzenberg remarked "the movie marketplace is very different today than it was three or four years ago," as he stated in *Variety* magazine, 'It's much more competitive,' in terms of playability, marketability and the availability of release dates." (Graser *Variety* 22 January 2015)

But with 87% of movies failing at the box office, it's easy to understand the reluctance to change— no industry wants to lose *more*.

How can hard science help?

Already in use within the creative-based industries in the form of artificial intelligence and behavior prediction, content-filtering is used by Amazon and

Netflix to suggest products, books, and movies based on customers' previous purchases. The music industry has adapted to using predictive intelligence by parsing song components in order to scout artists. Both Will Smith and some major film studios already look at data-driven analyses to forecast box office performance. Data science in film would be neither novel nor unwarranted in this financial climate.

Implementation remains tricky, however. Capitalizing on his background in actuarial science, Nick Meaney has positioned Epagogix as a process that offers film investment risk mitigation with predictions extracted from massive data inputs that can save studios money. Epagogix still has two challenges:

- 1) The inability to forecast future shifts in cultural mores, i.e. the outdated tendency to cast villains who are dark-haired or foreign, for example, will automatically suggest that future films continue to typecast villains along these lines.

- 2) Technological tools feel threatening to Hollywood insiders because it renders hard-earned human expertise and insight obsolete. Bankability, box office points, and perks (like invites to the important parties) stem from having power and knowledge. A tool perceived to eliminate creative—*human*—input, such as pointing out that movies, where the hero-wearing-a-blue shirt has greater box office potential, is going to ruffle creatives, who envision the hero in a red shirt.

Still, studio heads are increasingly engaging Epagogix to find ways to save—often discovering that they will make *more* money *without* a certain mega-star attached, which dispels conventional wisdom that stars are box office draws and essential to movie success.

More friendly to stars and creatives, their teams, and ultimately the studios, ScriptBook focuses on the storytelling metric. Stories—the backbone of a script—are time-tested; the three-act structure is found in the writings of Ancient Greeks and Shakespeare.

Not only are agents given a reprieve from reading scripts late into the night because there is a subjective tool to separate the good from the bad, ScriptBook has the ability to position itself as a support tool to creatives. Since outcomes are

derived from creative elements like story structure, language, and genre, ScriptBook generates data unaffected by social mores—and well-before the investment and greenlighting phase, too. Using story alone, ScriptBook can predict that a Hitchcock thriller ranks high, *Gigli* and *The Lone Ranger* do not; an analysis that could have saved hundreds of millions of dollars—and Jennifer Lopez and Johnny Depp some credibility.

A story-based predictive system benefits actors, directors and agents, in addition to film studios, because not only investment loss is averted, but stars' bankability remains intact. ScriptBook steers users away from a bad script allowing an actor or director (agents and production companies) to avoid being attached to box office bombs, or worse seen as overpaid or lackluster. Investors win because bad stories, the best predictor of success, go unmade.

Like Epagogix, ScriptBook's predictive prowess improves with more data, though ScriptBook only requires a simple input—screenplays. ScriptBook's biggest challenge will be developing product awareness and connecting with Hollywood players from their tiny offices in Belgium. Epagogix has done fine from London, proving that there is both a demand and need for big data in the film industry, irrespective of headquarters. Whether creatives embrace either hard-science product, however, remains to be seen.

BIBLIOGRAPHY

As Studios Cut Back, Investors See Opening

By Michael Cieply, *The New York Times* November 14, 2010 http://www.nytimes.com/2010/11/15/business/media/15cash.html?_r=0

The Cruel Math Behind Why Streaming Will Never Save The Music Industry By Mark Rogowsky *Forbes.com* 3/20/2014

The Formula - What if you built a machine to predict hit movies? By Malcolm Gladwell *The New Yorker* October 16, 2006

20 Best & Worst Films Made From Black List Scripts

By The Playlist Staff, *indiewire.com* April 10, 2014 <http://blogs.indiewire.com/theplaylist/the-10-best-10-worst-movies-made-from-black-list-scripts-2014041>

The Definitive Spec Script Sales List (1991-2012): 2012

by Scott Myers *Gointothestory.blcklst.com* - Official Screenwriting Blog of The Black List July 4, 2012

Odds of Selling a Screenplay

by Unk *screenwriterunknown.com* July 14, 2011 <http://www.screenwriterunknown.com/observations/odds-of-selling-a-spec-screenplay>

Theatrical Market Statistics 2013. Motion Picture Association of America March 2014 http://www.mpa.org/wp-content/uploads/2014/03/MPAA-Theatrical-Market-Statistics-2013_032514-v2.pdf

Global and entertainment media outlook 2014 – 2018. *Filmed Entertainment*. Paul Orchard & Alayna Francis, PriceWaterhouseCoopers <http://www.pwc.com/gx/en/global-entertainment-media-outlook/segment-insights/filmed-entertainment.jhtml>

When Number Crunching Met the Creative Industry - (video)

Nick Meaney, Keynote Speaker at Casualty Actuarial Society Seminar, March 13, 2013 <http://www.epagogix.com/media.html>

Great Flicks: Scientific Studies of Cinematic Creativity and Aesthetics. By Dean Keith Simonton, Oxford University Press, January 26, 2011 p.182/240 pp.

Masked Lawmen Stumbles at the Gate

by Brooks Barnes, *The New York Times*, July 7, 2013. <http://www.nytimes.com/2013/07/08/movies/masked-lawman-stumbles-at-the-gate.html>

What's Behind Hollywood's Big Slide: Franchise Fatigue, Fewer Big Movies

by Pamela McClintock, *The Hollywood Reporter*, January 7, 2015 <http://www.hollywoodreporter.com/news/whats-behind-2014s-box-office-761337>

Contact: justine@hadto.be

Epagogix

<http://www.epagogix.com>

info@epagogix.com

Studios, TV Industry & Investors

Epagogix helps studios to identify, enhance and deliver on-screen success, and guides investors in the creation of winning film and television-related investments. Epagogix works confidentially with the senior management of major film studios, large independents and other media companies, assisting with the selection and development of scripts by identifying likely successes and probable 'Turkeys'; helping to quantify a script/project's commercial success; and advising on enhancements to the Box office/audience share potential.

Epagogix's approach helps management of this most critical financial risk by delivering accurate predictive analysis of the Box Office value of individual film scripts, and by identifying and quantifying how and where to improve their commercial value. If requested, Epagogix sensitively bridges the gap between the financial and creative aspects of film production by providing quantified insights and advice to those responsible for script development.

[Hedge Fund and Investment Managers](#) work with Epagogix, determining highly targeted, non market-correlated risk/reward strategies for asset allocation, and developing fund-raising vehicles.

Links On Epogix

<http://www.epagogix.com/media.html>

<http://www.newyorker.com/magazine/2006/10/16/the-formula>

<http://www.marketplace.org/2013/07/19/business/whats-behind-future-hit-movies-algorithm>

<http://www.telegraph.co.uk/culture/film/film-news/10547268/The-geeks-who-are->

[directing-Hollywood.html](#)

<https://en.m.wikipedia.org/wiki/Epagogix>

<http://sushi-vegas.net/supercrunchers/>

<https://www.youtube.com/watch?v=fHc6QtdZ5FQ>

<http://www.intelligentlifemagazine.com/content/features/anonymous/slaves-algorithm>

<http://soreeyes.org/archive/2006/10/20/calling-bullshit-on-epagogix/>

<https://www.quora.com/How-might-the-technology-used-by-Epagogix-affect-the-movie-business-and-should-filmmakers-and-screenwriters-be-worried>

<http://www.irishtimes.com/business/technology/hollywood-s-best-friend-at-the-box-office-just-might-be-an-algorithm-1.1648288>

General Notes

The Physics of Culture

Epagogix (8 min. 33 seconds into presentation)

Uploaded on Jul 21, 2011

<http://www.ted.com> Kevin Slavin argues that we're living in a world designed for -- and increasingly controlled by -- algorithms. In this riveting talk from TEDGlobal, he shows how these complex computer programs determine: espionage tactics, stock prices, movie scripts, and architecture. And he warns that we are writing code we can't understand, with implications we can't control.

“Prismatic chaos, like all Netflix algorithms, determines in the end, 60% of what movies end up being rented. So one piece of code with one idea about you is responsible for 60% of these movies. But what if you could rate those movies before they get made?” Kevin Slaven, TED Talks, 2011

<https://www.youtube.com/watch?v=TDaFwnOiKVE>

* * *

Opinion: Are computers creative thinkers?

Waterloo Region Record

By Luke Dormehl

Dec 07, 2015

<http://www.therecord.com/opinion-story/6161718-opinion-are-computers-creative-thinkers-/>

“In south London, a company called Epagogix is fine-tuning a neural network — a vast artificial brain — that can help Hollywood predict hit movies by analyzing scripts using thousands of metrics. It can also advise how to tweak a script to make

the film more profitable, much as a creative producer would do. (Epagogix won't divulge client details — suggesting that, although Hollywood producers have long since embraced computer-generated effects in movies, they're not yet ready to admit that robotic attention can benefit screenwriting too.)

This year, Google employed the same 'deep learning' technology that powers Epagogix's neural network to build its Deep Dream project, which uses image recognition software to generate entirely new pictures. (The computer scans millions of images of, say, a chair, and then tries to draw one from scratch.) Other tech researchers are similarly using algorithms to write novels and generate music.”

* * *

Inceptionism: Going Deeper into Neural Networks

Wednesday, June 17, 2015

Posted by Alexander Mordvintsev, Software Engineer, Christopher Olah, Software Engineering Intern and Mike Tyka, Software Engineer
Google Research

Update - 13/07/2015

Images in this blog post are licensed by Google Inc. under a Creative Commons Attribution 4.0 International License. However, images based on places by MIT Computer Science and AI Laboratory require additional permissions from MIT for use.

Artificial Neural Networks have spurred remarkable recent progress in image classification and speech recognition. But even though these are very useful tools based on well-known mathematical methods, we actually understand surprisingly little of why certain models work and others don't. So let's take a look at some simple techniques for peeking inside these networks.

<http://googleresearch.blogspot.ch/2015/06/inceptionism-going-deeper-into-neural.html>

* * *

FILMONOMICS: Predictive Analytics vs Humans

By Ted Hope

<http://issuesandactions.hopeforfilm.com/2013/08/filmonomics-predictive-analytics-vs-humans.html>

Those that think *Twilight*, *The Green Hornet*, *The Beach*, *The Abyss* and *Boston Shuffler* are a bunch of movie titles are only partially correct. They are all nicknames for just a few of the algorithms used on Wall Street to give firms those precious milliseconds of trading advantage. Some two thousand physicists and mathematicians work in the financial sector cooking up these computational black boxes – and a handful are now applying such predictive modeling and risk evaluation skills to the film industry in order to determine why some movies click and many others don't. With worldwide spending on filmed entertainment climbing towards **an annual \$100 billion** and beyond, the pay-off is self-evident.

Cinema historians will point out that movies have always been subject to generic modification in order to maximize their appeal across multiple markets and cultures. Studio distribution executives will tell you that older-female-targeted films tend to “over-perform” in Germany and Australia, while Latin America has an outsized soft spot for family fare. And Hollywood tailors its films accordingly. But such factory-belt fine-tuning is nothing compared to the scientific wizardry being applied today. Consider the following:

- *Netflix*, awarded \$1 million to the team that created the collaborative filtering algorithm known as *Pragmatic Chaos* that predicts user ratings for films based on previous ratings. According to TED speaker Kevin Slavin, in his **memorable talk on how algorithms are shaping our world**, this one piece of code accounts for 60% of the movies that end up getting rented.
- Studio marketers – **and some trailblazing producers** – are turning to neuroscience to do what test screenings and focus groups have been unable to do very reliably: pinpoint the levels of audience

engagement. Using brain scans and other biometric data, firms such as *MindSign*, *NeuroFocus* and *EmSense* **are not only evaluating trailers, but also individual scenes, characters and effects.**”

- As mentioned in a previous **Filmonomics blog**, *Epagogix* uses artificial neural networks to analyze screenplays with a view to determining their probability of commercial success. Meanwhile, professors at NYU Stern and Pennsylvania’s Wharton School claimed to have devised
- **a reliable screening method for choosing movie scripts** based on “textual information” – including the use of specific words.

While different in their methodologies, these algorithmic approaches share a fundamental trust in data analysis as the best filtering system. For those who think that the creative vetting process is best left to humans, there is a contrasting set of computational tools that might be broadly termed social recommendation engines. These rely on the curatorial instincts of one’s peer-groups. These influential tribes could be your friends on Facebook banging the drum for the film they just saw at the multiplex or film festival. Or they could be the membership network of 300 movie professionals whose votes determine those as-yet-unproduced screenplays that deserve to be included on **The Black List**.

Conceived by **Franklin Leonard**, VP of Creative Affairs at Will Smith’s *Overbrook Entertainment*, The Black List is perhaps the nearest we get to an instant snapshot of Hollywood’s collective taste buds. And it’s persuasive too: more than 125 past Black List scripts have ended up getting produced and being released theatrically, generating \$11 billion between them. Successes include four of the last eight Oscar-winning screenplays.

Building on its hit-making potential, The Black List announced last year the launch of an online members’ community that will make algorithmic screenplay recommendations based on individual tastes. Users can now explore real-time updated lists of Hollywood’s most liked scripts. By offering a blend of human insight and artificial intelligence, The Black List is affirming the need for both sets of tools in order to decode the DNA of filmmaking success. As we are learning, each inspires the other. It’s a no-brainer.

“We all fear the ‘future script’ spit out by a robot. But we can take comfort in the fact that the human brain is its own vast data storage program. A serious artist does his or her own rigorous study of successful works to lodge the common elements and patterns into their unconscious and inform their artistic output,” notes **Jennine Lanouette**, a story consultant who lectures at both Lucasfilm and Pixar (and will soon be contributing periodic blog posts on screenwriting analysis for Slated). “As I see it, there is room for both – the organic artistic process and the set of objective measurements. The best application I can imagine of these computer prediction methods is in determining the size of the budget for a given project, and discovering niche market or sales territory potentials. The danger, in our winner-take-all economy, is everyone wanting to compete for the same gold ring – the billion dollar box office take. This is when art loses out. A computer might be able to maximize return on investment, but it will never create art. You need a human soul to do that.”

Colin Brown

* * *

CASE FILE

Who are we?

Dirk de Wit, Student number: **0783878**

Larn van de Vries, Student number: **0791869**

Summary of the case

Eindhoven University of Technology

5612 AZ Eindhoven, Netherlands

<http://sushi-vegas.net/supercrunchers/#whoarewe>

Dick Copaken started a company named Epagogix (from the Aristotelian idea of inductive learning) that uses a trained neural network to crunch numbers and predict how much a movie will make before it has even been recorded. The prediction is based only on characteristics of a movie script, even before the stars and director of the movie are casted.

Out of nine films the neural equation was able to accurately predict the profit of six films based on characteristics of the movie scripts only, while experts were able to predict the profits of a movie only three out of nine times. A large studio could net around a billion dollars profit a year by using the neural equation.

Epagogix works directly with movie writers to help them developing a script to optimize the box office. Epagogix thus prevents bad movies from ever coming into existence. If a script has flaws Epagogix can calculate the expected revenues for possible changes.

Although Epagogix focuses mainly on characteristics of the script they do take actors and directors into account. Epagogix found that actors and directors form a very large factor in predicting box office results. However, the location of the set has a much larger effect. Epagogix does not want to hurt stars. It is in fact helpful for actors to decide whether they are interested in playing in a movie or not. Epagogix can even help actors decide if they would like their money up front or if they would like a share in the profit.

Unfortunately some studios are entirely closed-minded to a neural equation that could decide whether a project should be green lighted or not. Even though fund

managers are willing to invest millions of dollars into movies that are optimized for box office and even if the neural equation would be right a hundred percent of the time, some studios will still not change the way they green-light movie scripts.

Main Question

As stated above, Epagogix is a company that tries to predict box office gross gain, by analyzing movie scripts and estimating the returns. But essentially or necessarily not for the money itself. Copaken and his co-workers are trying to optimize their neural network system so that ‘bad movies’ won’t even get the green light for development after testing the movie-script with the neural network. Instead of, using the old term, studios ‘shooting turkey’s’. Epigagox is making sure these movies won’t even happen.

The second part being discussed in the case is the acceptance of movie studios and the society, which we will come back to later in this document in the ‘societal impact’ section.

Why

After being graduated on Harvard Dick Copaken started to work as a lawyer at the Washington Office of Covington & Burling. Being a lawyer he crunched numbers for his legal clients. Dick Copaken was a man who has seen two to three movies a week for 50 years. He filed hundreds of plots, characters, and scenes in his mind. Dick Copaken met Nick Meaney on a case years ago. Copaken and Meaney shared a fascination of a computerized learning system called intelligent neural networks. Meaney has a background in risk management and wanted to bring that principle to the world of movie business. Meaney told Copaken about a friend of him who together with a business partner developed a system for evaluating the commercial potential of stories. The two men were able to define a script into different categories so that they could assign scores to each category.

Together with the two men, and another former colleague of Meaney named Sean Verity, Meaney and Copaken started a company named Epagogix.

New results

For any producer that has to make movies continuously and any distributor that has to select promising new movies, reduction in the uncertainty of the market is of great importance (Lee & Chang, 2009). Everywhere around the globe movies are being produced. This is thus also the case for Korea, where two researchers named Lee and Chang did research in 2009 on box-office performance of both imported movies and Korean movies.

Conducted research on box-office performance found that box-office performance could be divided into psychological or behavioral and economical perspectives. The psychological perspective focuses on individual decisions to visit a movie such as needs, attitudes, opinions and personality traits. The economical approach aims to finding the attributes that are related to financial performance of a movie (Sharda & Delen, 2006). In previous research the attributes affecting box-office performance were studied without taking into account the interaction effect between the attributes. Lee and Chang (2009) replicated a study in which they focus on the causal relationship among all attributes related to movies (Lee & Chang, 2009). For this study Lee and Chang (2009) collected information about directors, actors, genre, ratings, producers, distributors, budget, related articles and critics of moviegoers and analyzed 100 movies, which were released between January 2005 and October 2006. The total number of moviegoers per movie was defined as the dependent variable (Lee & Chang, 2009). All the attributes selected in the research by Lee and Chang (2009), were based on earlier research by Elberse and Eliashberg in 2003 and 2006. The prediction was done with a Bayesian belief network (BBN). All attributes used could be subdivided into four categories (i.e. Box-Office Performance, Production, Distribution and Exhibition). Lee and Chang (2009) found that a large screen number, peak season release, good word-of-mouth and a moderate budget were the most important factors for a non-blockbuster movie to be successful. While for a blockbuster movie factors like a large screen number, star cast, good word-of-mouth and a low level of competition were most important. By taking the interdependency between the movie attributes into account the predicted box-office performance results are more accurate than results through any other method available. The predictive power of the BBN however could be weakened by the absence of any data on marketing expenditure (Lee and Chang, 2009).

In China, Zang, Luo and Yang (2009) have conducted comparable research. In their study they compared three different neural networks, a Back-Propagation (BP) neural network, a Multi Layer BP (MLBP) neural network, and a Multi-Layer Perceptron (MLP) neural network. BP neural networks are the most widely used neural networks and are considered as the workhorse of Artificial Neural Networks (ANNs) (Zang, Luo & Yang, 2009). In this study the researchers have used attributes as propaganda, content category, showing time, competition, and cinema information. But also actors, director, nation, and screen number. Zang et al. (2009) compared the accuracy of BP, MLBP and MLP by comparing the pinpoint accuracy 'Bingo' and the within one category '1-Away' percentages obtained by each ANN for the years 2005 and 2006. The BP neural network has a pinpoint accuracy of 68.1% and a within one category of 97.1%, therefore the prediction accuracies are significantly better than the prediction accuracies of the MLBP and MLP neural networks (Zang, Luo & Yang, 2009). However, in this research Zang et al. (2009) concluded that a valid theory is required to reduce the number of predictors used. Much additional work needs to be done according the modeling extensions, experimentation for testing the performance and applications to other media (Zang, Luo & Yang, 2009).

In the paper by Gazly, Clark and Sinha (2011) state that also awards, budget and symbolism (a story translated from another medium like a book) have a direct correlation on short-term box-office success, while advertisement, critical reviews and stardom have a direct correlation with perceived quality of a movie by consumers (Gazly, Clark & Sinha, 2011).

The research described above concerns a different aspect of predicting box-office performance than the approach of Epagogix. However, it shows that box-office performance is globally concerned and that there are more ways to accurately predict box-office performance. While the approaches described above were both based on box-office predictions done just before a movie hits the theatre and are only useful for the first two weeks a movie is aired (Sharda & Delen, 2002; Gazly, Clark & Sinha, 2011), the Epagogix prediction model is based on nothing more than the movie script. As Epagogix is keeping its formula secret we don't know much about the construction of their algorithm.

However there is some research done about a similar approach. Eliashberg, Hui, and Zhang have written a paper in 2007 about green lighting movie scripts. Their approach is based on the domain knowledge of screenwriting, natural language processing, and statistical learning methods. Their goal is to predict movie returns on investment (ROI) entirely based on textual information available in movie scripts (Eliashberg, et al. 2007). Every year 15,000 screenplays are registered at the WGA, while just 700 movies are made each year. Thus, a reliable approach is needed to accurately green light good movies. In the approach of Eliashberg et al. (2007), compared to the researches mentioned above, factors as actors, director, producer, number of screens, and competitions are not included. Their approach is based on the storyline only (Eliashberg, et al. 2007). However, movie scripts are not readily available on the Internet, which is why Eliashberg and his colleagues used film spoilers for their model. A spoiler is an extended summary of a movie storyline, written by viewers after watching the movie. By calculating the correlation between a movie script (when it is available) and a movie spoiler, the researchers have found that there is a positive and significant correlation between the two. A set of 281 movie spoilers is used for their approach, 200 movies were used for training the model and 81 for testing it (Eliashber, Hui & Zhang, 2007). As the first step towards a model, to represent the data and decrease the number of dimensions, the bag-of-words model can be used. The bag-of-words model is developed in natural language processing. The model tries to represent a document by the words that it contains and the number of times a word appears in the document without taking into account the word order. The model is able to pick up the themes, scenes, and emotions of a script. Since the model does not take into account the word order, it cannot be used to capture the story line of a movie. Two sentences written with the exact same words can have a totally different meaning when the word order is different (Eliasher, Hui & Zhang, 2007).

That is why Eliashberg and his colleagues used the knowledge of screenwriting experts to evaluate scripts. The experts defined some criteria a script should possess for a higher box-office success. Because a computer cannot understand a storyline a questionnaire is defined consisting of 22 questions. Human judges are hired to read the movie spoilers and answer the questions. Moreover, the components of a script can have (as also described in the previous examples) very complex relations with each other. That is why the analytical tool Eliashberg and his colleagues

propose must be flexible so that it can accommodate all possible interaction among different components in a script (Eliashberg, Hui & Zhang, 2007). Eliashberg, Hui and Zhang (2007) choose to use Bag-CART an extension to CART (Classification and Regression Tree), a technique that is able to uncover complex interactions between predictors that otherwise might be impossible to uncover, with this technique the results of the decision aid will be easy to interpret for movie producers. CART is a technique that is used in many different disciplines as a prediction method. Unfortunately the CART technique suffers of an estimation instability problem, a slight change in the data can result in a whole different structure of the resulting regression tree. This problem is serious when it is used with a small sample size. To solve this problem a bootstrap method can be used called 'bagging' (Eliashberg, Hui & Zhang, 2007). More information for the prediction comes from the bag-of-words model to identify the frequencies of important words in the spoiler. With a stemmer the same words of a different form are grouped together. The 100 words that occur most frequently are determined as most important and are used for the analysis. The CART tree can interpret the frequency of each predictor so that the most relevant predictors can be identified. Eliashberg and his colleagues found that passive sentences, violent sentences, and the number of sentences required to describe the story line are the most relevant predictors. The movies can be classified into two groups; above median and below median, the median ROI is 27.2%. The correct classification rate using their prediction model is 61.7%. The approach is able to significantly improve the studio's profitability, relying only on story line information derived from movie scripts (Eliashberg, Hui & Zhang, 2007).

Scientific follow-up

The scientific follow up is actually kind of shallow so far. The main article written by Malcolm Gladwell isn't even a paper really. Though Gladwell did write a nice story about the history of the idea behind Epagogix and Dick Copaken, much remains behind closed doors. As stated in the SuperCrunchers book, it's a secretive company. This is also the case when searching for additional literature. Most of the literature on using neural networks for predicting box office gross doesn't even mention Epagogix. Though this might be because these papers were published earlier than the Epagogix story came out. But obviously this shouldn't apply for the papers that were 'recently' published at least 2007 and onwards.

More and more research is being done on the use of neural networks on box office gross. There already used to be a lot of research with standard statistical models on box office gross. And this research is taken as a base for the use of neural networks in later researches such as Sharda, R., & Delen, D. (2006) and Zhang, L., Luo, J., & Yang, S. The standard statistical approaches are getting outperformed by the neural networks at almost every turn now simply because of the fact that neural networks can handle more complex models than the standard models researchers were using. One downside of this is that a neural network is more of a 'black box' than the usual regressions or factor analyses.

Though you cannot say it is 'common practice' to research box office gross with neural networks. These box office gross problems are so complex that the use of a neural network (to get any decent result) is almost mandatory. So they are quickly becoming the norm, so to speak, to use as a method for these problems.

Sadly as stated before there has not been done much research on these cases yet. Most researchers find these problems too hard or too much work to really spend any time on it. This is a reason why most of the research in this area comes from the same people, e.g.: Sharda and Delen and Eliashberg etc.

The latest papers in the area of using neural networks are now focusing on getting more accurate results while using less predictors. This means easier (and faster) models. While these models are getting less complex it's also possible to start combining and averaging different kinds of Neural Networks together for a more optimal result. These Neural Networks can be tested by splitting the data. One of these halves will be used to 'learn' the Neural Network a better fitting model. The other half will be used to test the newly learned model on real data, to see if it is truly a better fit to the data.

Also, so far the current research has been very specific, such as only movies from certain regions like: New Zealand, or Korea. Current research is trying to expand and globalize the models so that they are usable for other countries as well. Lastly, something that has been discussed in the paper by Delen, D., Sharda, R., & Kumar, P. (2005). Research is currently also focusing on getting the theory to work

in practical systems that can be used by certain target groups. These target groups contain people from various fields such as directors, movie studios, actors etc. Right now the most progress is made on web-based systems. Though no doubt there are private systems that are not publicly available in papers or the like.

Critique

Our main point of critique is basically the amount of research that has been done on this topic. So far most researchers have been taking the 'easy route' on this subject, or haven't even started researching this subject at all. This is mainly because the subject is rather difficult to research and also very time consuming. We would rather see people embracing this subject as its both very interesting and could also be a giant money saver for multiple parties in the industry.

Another point of critique we have is that the research so far has been focusing on the first two weeks of the release of the movie. As stated before, this is because it's difficult to estimate what the box office gross is going to be after these two weeks. But we would've liked to see research done on the subject of improving this problem instead of just focusing on the 'easier' part. Neural Networks might have the power to see this subject over a longer terms and they may just 'crack the code' in that regard.

Maybe not necessarily a critique point but something we would like to address. There are multiple kinds of neural networks, and one of the most used ones is the CART network. This network is used a lot because of its power as a neural network but it has the downside of being very instable. Maybe it's a good thing to start researching more about this instability and try to improve this issue.

Another fact is that it's still necessary to have humans in the process of neural network data entry. These neural networks are not fully autonomous yet as they require humans to read the script (or spoilers in this case) to enter the textual data in the system. As explained in the paper of Zhang, L., Luo, J., & Yang, S. (2009) there's also a problem with the order of the words. Sentences with the exact same words but in a different word order, or sentence structure, cannot be separated from each other by the neural network and thus still require human input. If there would be more research on the topic there might be some progress in this area, which

could help neural networks to gain accuracy and obviously a lot of speed in the decomposition of the scripts.

Societal impact

Most of the societal impact has already been discussed above in various chapters of this document but here is a quick recap.

So far the use of neural networks to number crunch movie data has not been common practice yet but is sprouting from the ‘standard’ statistical methods used to predict box office gross. There has been done various kinds of research on the effect of certain predictors such as actors, rating, ticket price etc. to estimate the amount of ‘return of investment’. But all this research focuses on the time period after a movie has already been released. There hasn’t been that much research yet about actual ‘green-lighting’ of scripts for example. And especially not when combined with neural networks. Though it is becoming more common practice to research this.

Something that was also to be read in the SuperCrunchers book, and is in fact a problem with almost all number crunching examples, is the fact that society is not really taking advantage of it yet. There are already systems that have a very decent hit rate on the detection of a scripts worth, but studio’s, directors and actors are not yet willing to invest or even use the systems. They still just don’t believe in the feasibility or maybe they still fear for their jobs. Hopefully with the addition of for example: web based systems, this ice can be broken and the acceptance of these systems could start to rise.

The companies that do see the value in these systems however are jumping on the opportunities. Companies like Epagogix themselves are investing a lot of time and money in getting their neural network trained and are getting more accurate with their predictions after every adjustment to the model. These companies sadly work in high secrecy as to make optimal profit, which of course would not help for a ‘consumer’ centered version of the neural network.

So to conclude, thus far there has not been that much research on the specific use of neural networks on green lighting movies. Or at least not publicly, but there are

certainly already powerful networks that can already be used in the business environment. Sadly many studios are still often rejecting the idea and the ones that do embrace the idea are often not publishing their findings because of financial reasons.

References

Gazley, A., Clark, G., & Sinha, A. (2011) 'Understanding preferences for motion pictures'. *Journal of Business Research*, Vol. 64, 854-861.

Delen, D., Sharda, R., & Kumar, P. (2005) 'Movie forecast guru: a web-based DSS for Hollywood managers'. *Decision Support Systems*, Vol. 43, 1151-1170

Eliashberg, J., Hui, S., & Zhang, J. (2007) 'From story line to box office: A new approach for green-lighting movie scripts'. *Management Science*, Vol. 53, No. 6, 881-893.

Lee, K., & Chang, W. (2009) 'Bayesian belief network for box-office performance A case study on Korean movies'. *Expert Systems with Applications*, Vol. 36, 280-291.

Zhang, L., Luo, J., & Yang, S. (2009) 'Forecasting box office revenue of movies with BP neural network'. *Expert Systems with Applications*, Vol. 36, 6580-6587.

Sharda, R., & Delen, D. (2002) 'Forecasting Box-Office receipts of motion pictures using neural networks'. *Department of Management Science and Information Systems, Oklahoma State University, Stillwater*.

Sharda, R., & Delen, D. (2006) 'Predicting box-office success of motion pictures with neural networks'. *Expert Systems with Applications*, Vol. 30, 243-254