

GAOSU

Machine Insights on Chinese Film Audiences

Yves Bergquist

JANUARY 2019

Contact:
Yves Bergquist
Founder & CEO
yves@corto.ai

BACKGROUND

- **China is the world's second-largest theatrical movie market (after the US), and it's become the fastest growing (9% growth in 2018).** Box office receipts in China have reportedly grown from \$8.6 billion in 2017 to \$8.9 billion in 2018, according to Xinhua News Agency. Chinese companies are reported to have added more than 9,000 news screens throughout China in 2018 (bringing the total to more than 60,000 nationwide).
- **Despite this growth, it's been reported that the market share for US studio films has been down sharply in China (24%, according to one estimate, from \$2.17 billion in 2017 to \$1.64 billion in 2018).** As theatrical distribution pushes through more rural areas, domestic production seems to be at a substantial advantage.
- **This said, the success of such films as "Avengers: Infinity Wars" and "Aquaman" in the Chinese theatrical window is impressive and promising, and it indicates that there is a path to success for US films.** Even Chinese audiences' enthusiasm for the "Mamma Mia!" Sequel seems to indicate that there is a potentially strong market for American musicals there.
- **But with tight government oversight, and a short marketing window, theatrical distribution of US films in China remains a substantial challenge.**
- **It's been reported that word of mouth is playing an increasing role in driving theatrical success in China.** The surprise performance of such small budget Chinese films as "Dying to Survive", "Project Gutenberg" and "A Cool Fish" is a sign that audience-driven buzz is playing an increasing role in driving theatrical box office there.
- **But Chinese audiences are elusive. Relevant conversations happen across a wide number of sources, and their velocity is high. Different channels seem to be sensitive to different attributes of films (Douban users tend to pay more attention to the plot of films, for example). This is why building the capacity to comprehensively monitor and analyze Chinese audience conversations in real-time is important.**

PROJECT GAOSU

2

- **ETC has been building the technical capacity** for our members to access audience conversations about entertainment in foreign territories where financial stakes are high, but machine monitoring and analysis of audience conversations has been limited.
- **For the past 18 months, ETC's "AI & Neuroscience in Media" Project has been collecting hundreds of millions of conversations about film in China, SE Asia** (Indonesia, Philippines, Malaysia, Singapore, Thailand) and Latin America (every country except Guatemala, Nicaragua, El Salvador, Costa Rica, Panama and Belize).
- **Our coverage has been particularly extensive in China**, where a team of 4 PhD-level Mandarin computational linguists has been working for 12 months to build an application capable of collecting and analyzing all conversations about film in China, in real time.
- **This project has been called "Gaosu"** ("tell me" in Mandarin), and will be launched as part of the Corto suite of features in Q2 2019. Gausu will provide real time monitoring and analysis of Chinese film audience conversations across dozens of sources.
- **This document will provide ETC members with an overview of the output of Gausu** for 4 films which have already been distributed in China (Aquaman, Skyscraper, Jurassic World, and the Mamma Mia Sequel), two which trailers have recently come out (Pikachu and Godzilla), and one which has just started production (Dune).

Sources: Tianya, AcFunTv, Youku, Bili Bili, Sina Weibo, Douban, Maoyan, Tudou, Mtime, Gewara, Toutiao, ZhiHu, qq, IQiYi, Baidu, Nuomi, Baidu Tieba, Tencent Weibo.

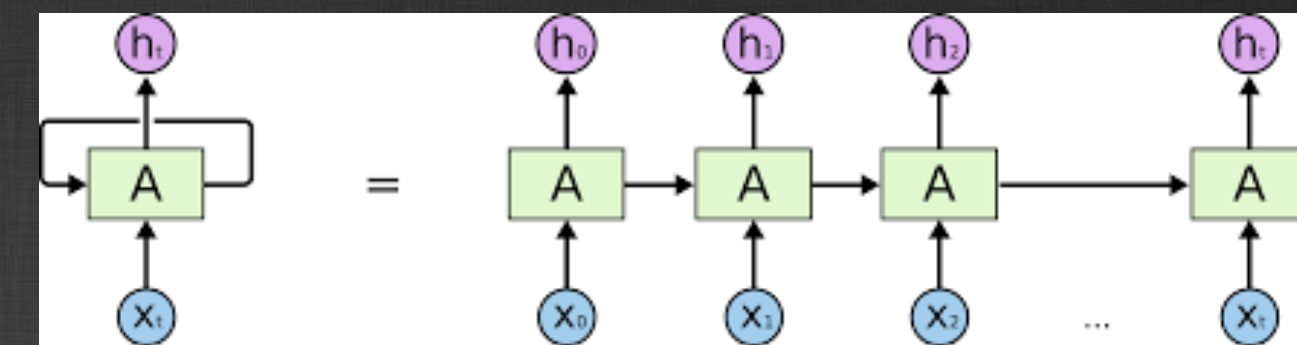


METHODOLOGY

3

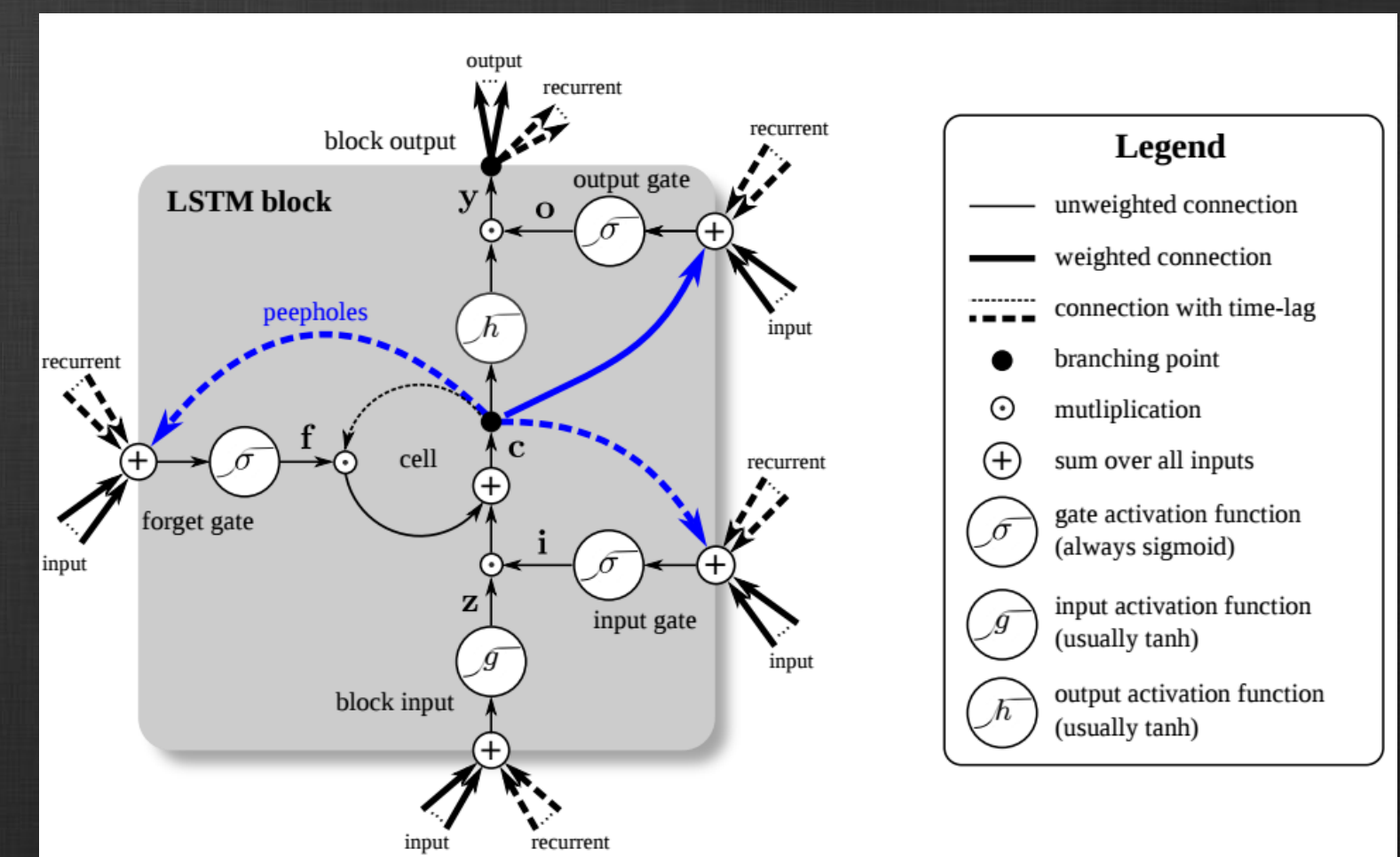
- Data collection was automated through scraping pipelines. We can access any and all web content in China in near-real time (30 minute lag).
- For Natural Language Processing, the Corto-ETC team used its own NLP tools for Mandarin.
- Our biggest challenge was to classify sentiment for each entity (cast, plot, director, VFX, etc.). At this stage of NLP research, it's easy to score the polarity of sentiment for a corpus of text. But parsing that text to classify sentiment polarity about specific attributes/symbols/objects/entities in the text is still a challenge, as we have to create a machine representation of grammatical dependency.
- To create a more semantic view of the text itself, we represented all collected conversations in a directed hypergraph structure. This data structure allowed us to unpack the temporal and semantic structure for each sentence.
- We then used an LSTM neural network to perform "Aspect-Level Classification", or classification of sentiment for each object in the sentence.
- "Long Short Term Memory" Networks (LSTM) is a class of neural network (in the family of Recurrent Neural Networks, or RNNs) that organize each individual element (they're called "blocks" in a "loop", which allows information to persist not only within that block, but at each level's layer. RNNs perform feature representation on sequential data with high temporal dependencies, like video, audio, and text. LSTMs are a class of RNN that can better handle long-term dependencies than traditional RNNs, hence are particularly useful in classifying natural language across a large corpus of text.

One block in a RNN



Source: <http://colah.github.io>

One block in a LSTM



Source: NVIDIA

TEAM:



YVES BERGQUIST

Director, AI & Neuroscience in Media Project, ETC. Founder/CEO, Corto.

- Data scientist.
- Director of the “AI & Neuroscience in Media” Project at the University of Southern California’s Entertainment Technology Center. Funded by (and reports to) all 6 CTOs of Hollywood studios.
- 20 years of consulting and modeling experience.
- Lead architect of “Contender”, the world’s first AI-driven content marketing intelligence engine, which mines social media conversations for deep semantic insights on how specific narrative structures impact ad performance for brands.



JASON KRAFT

CTO, Corto.

- Former Co-founder & CTO, MagicBus (Y Combinator Alum).
- 4x startup veteran.
- Full-stack developer and data scientist.



MACIEJ DRWAL, PhD

Chief Data Scientist, Corto.

- Associate Professor of Computer Science, University of Toronto (Geoff Hinton’s team).
- ex-Google.
- PhD, Wroclaw University of Science and Technology (Poland).



CHANG WU, PhD

Chief Computational Linguist, Corto.

- Computational Linguistics Professor, Xiamen University.
- National Distinguished Scientist Award Recipient, 2017.
- One of China’s foremost NLP academic, with 100s of publications.



RUITING LANG, PhD

Computational Linguist, Corto.

- Former Head of Computational Linguistics, Hanson Robotics.
- PhD in Computational Linguistics, Hong Kong PolyU.



XIANGYU WANG

Data Scientist, Corto.

- PhD Candidate, USC Viterbi School of Engineering.
- Master’s Degree in Computer Science, USC Viterbi School of Engineering.
- Master of Science in Engineering, Shanghai Jiao Tong University.
- Computer Vision and NLP expert.

GAOSU OUTPUT (1)



AQUAMAN

	Positive	Negative
General	0.812199037	0.187800963
Cast	0.917525773	0.082474227
Director	0.932075472	0.067924528
Plot	0.269972452	0.730027548
Trailer	0.87826087	0.12173913
VFX	0.872289157	0.127710843

Some common words under **positive reviews**:

画面, 特效, 温子仁, 媚拉, 紧凑, 马王
(visual, vfx, James Wan, Mera, fast-paced, Khal Drogo)

Some common words under **negative reviews**:

剧情, 老套, 爆米花电影, 音效
(story, cliched, popcorn movie, sound effect)

- Chinese audiences were overwhelmingly positive about Aquaman. They loved the visual effects and the fast pace of the film.
- The visual effects elicited the most positive comments, with users comparing the quality to that of "Avatar" or "Lord of the Rings". All the Atlantis scenes were listed as looking really great.
- Large volume of positive conversation focused on the cast and its physical attractiveness. The film has gotten Jason Momoa big star power in China (people mentioned he could be the next Dwayne Johnson).
- Complaints and negative comments were around the story itself and the music/sound effects. Chinese audiences liked the fast pace of the film, and did not complain much about logical errors in the plot itself.
- The trailer seems to have been extremely effective in driving audiences to see the film: audiences didn't expect the trailer to be so good after "Justice League" disappointed.
- The director got particularly good reviews, and the fact that he is Chinese was a central feature of affinity for him and his work.

GAOSU OUTPUT (2)



	Positive	Negative
General	0.698717949	0.301282051
Cast	0.893617021	0.106382979
Director	1	0
Plot	0.260273973	0.739726027
Trailer	0	0
VFX	0.814814815	0.185185185

Some common words under **positive reviews**:

惊险, 刺激, 震撼, 视觉冲击, 强森, 家庭, 亲情

(exciting, shocking, visual shock, Dwayne Douglas Johnson, family, family love)

Some common words under **negative reviews**:

套路, 没有新意, 烂, 弱智, 无聊, 开挂

(cliche, boring story, stupid, illogical)

- In general, a majority of people thought it was a good commercial film, with great entertaining value and good visuals. VFX and fight scenes were a big highlight in the positive comments, as well as Dwayne Johnson of course (he is very popular in China).
- Douban audiences seem to be more sensitive to the quality of the plot than audiences on Weibo, for example, which are more sensitive to the visuals.
- However there was large volume of conversations complaining about the "simplicity" and "unrealistic nature" of the plot, and not just on Douban. Lots of people complained that it seemed like a carbon copy of "Die Hard" and "Mission:Impossible".
- We found fairly large volume of conversation expressing the desire that Dwayne Johnson would start in "more sophisticated" films, which we found interesting, as it seems a recurring theme in conversations about him.
- There was also a fair amount of conversation about some CGI visuals in some scenes not meeting the standards of quality.

GAOSU OUTPUT (3)



JURASSIC WORLD: FALLEN KINGDOM

	Positive	Negative
General	0.655279503	0.344720497
Cast	0.611111111	0.388888889
Director	0.857142857	0.142857143
Plot	0.2	0.8
Trailer	0	1
VFX	0.936507937	0.063492063

Some common words under **positive** reviews:

特效, 技术, 恐龙, 刺激, 娱乐

(vfx, technique, dinosaur, exciting, entertaining)

Some common words under **negative** reviews:

逻辑, 三观, 套路, 无聊

(illogical, wrong values, cliched, boring)

- Overall, the latest film in the Jurassic Park franchise has gotten positive reviews in China. The visual effects are the dominant attribute driving positive audience comments.
- There's also a certain volume of negative comments, centered on the plot and the "values", which a lot of comments refer to.
- There are many positive reviews of the cast, with a majority centered on Bryce Dallas Howard's "beauty". But there are also a certain number of people complaining about actors' performances.
- Interesting to note that a large volume of comments on Baidu seem to be from bots (large volume of the same comment). We did not find substantial bot traffic with other films we looked at.

GAOSU OUTPUT (4)



**Mamma Mia! Here
We Go Again**

	Positive	Negative
General	0.75	0.25
Cast	0.890909091	0.109090909
Director	0	1
Plot	0.222222222	0.777777778
Trailer	0	0
VFX	0	1

Some common words under **positive reviews**:

音乐, 舞蹈, 梅姨, 柯林, 双线, 风景

(Music, Dance, Meryl Streep, Colin Firth, Double storyline, Beautiful scenes)

Some common words under **negative reviews**:

剧情, 特效, 年轻演员

(Story, vfx, young actors/actress)

- Those who commented on, and reviewed this film, felt generally positive about it, hence the 75% positive volume of review overall. We found a lot of people stating that this was of their favorite musicals, and it seems that the genre as a whole has a lot of promise in China.
- We found overwhelming enthusiasm for Meryl Streep and Colin Firth, who both seem to have a very passionate following in China. However, the younger cast did not elicit the same positive reviews.
- We found a lot of negative volume for the plot. While it seems that audiences liked the parallel plot structure, they were disappointed that it did not integrate enough with the first film.
- We also found a lot of negative reviews with regards to the VFX, especially the fact that often actors seem to be cut from the chroma key in some scenes.

GAOSU OUTPUT (5)

Sources: Tianya, AcFunTv, Youku, Bili Bili, Sina Weibo, Douban, Maoyan, Tudou, Mtime, Toutiao, ZhiHu, qq, IQiYi, Baidu, Baidu Tieba, Tencent Weibo.



PIKACHU
TRAILER

	Positive	Negative
General	0.6647727273	0.3352272727
Cast	0.2553191489	0.7446808511
Director	0	2
Visual	0.8133333333	0.1866666667

Some common words under **positive** reviews:

萌, 可爱, 皮卡丘
(cute, pikachu)

Some common words under **negative** reviews:

长毛, 配音, 违和
(fur, voice actor, strange/unnatural)

- Chinese audiences have very high expectations for the film (Pokemon has a very large fan base in China).
- The Pikachu trailer was generally a success ("very cute", seemed to be recurrent in conversations).
- However, Ryan Reynolds as Pikachu seems very divisive: Chinese audiences either love it or hate it (some don't like that he's speaking in English and would like a kid's voice for him).
- Another big point of contention is the fur. Some fans like it, others hate it.
- There was virtually no mention of the director (but only negative mentions).

GAOSU OUTPUT (6)



	Positive	Negative
General	0.883870968	0.116129032
Cast	0.97826087	0.02173913
Director	0	1(6/6)
Plot	0.166666667	0.833333333
VFX	0.710526316	0.289473684

Common words under **positive reviews**:

燃, 特效, 卡司, 小11, 章子怡, 预告

(Marvelous, VFX, Cast, Millie Bobby Brown, Ziyi Zhang, Trailer)

Common words under **negative reviews**:

剧情, 黑, 看不清

(Story, Dark)

- The Godzilla trailer has been a big success in China. It has seen an overwhelming number of positive conversations, with even a lot of commentators saying it's "the best trailer of 2018".
- Many comments mention the trailer's music (Debussy) and how perfectly it fits into the trailer.
- Audiences feel very positive about the cast. Millie Bobby Brown has a growing fan base in China.
- The film's director, on the other hand, is receiving a lot of negative coverage. Chinese audiences generally do not like his previous work, and his involvement is mentioned as "something that could drag the quality of the film down".
- Regarding the plot, Chinese audiences overwhelmingly express the desire that the plot be simpler and less of a "soap drama" (terms mentioned frequently) than the last Godzilla film.
- We have seen a substantial number of commentators worrying that the trailer was going to be much better than the film.
- Lots of comments also expressed the wish that the new Godzilla be visually brighter than the last one, which a lot of people complained was often too dark for the audience to see.

GAOSU OUTPUT (7)



	Positive	Negative
General	0.964285714	0.035714286
Cast	1(19/19)	0
Director	0.866666667	0.133333333

Common **positive terms**:

维伦纽瓦, 甜茶, 维神

(Villeneuve, Timothee Chalamet, 'God' Villeneuve)

- Almost all the conversations found about the upcoming Dune was around how excited audiences were that Denis Villeneuve is directing and the film is starring Timothee Chalamet, who is popular in China (but they have very high expectations for his acting performance).
- Denis Villeneuve fans in China were ecstatic about the new "Blade Runner", which they considered a "visual masterpiece".
- Most people talking about the film in the digital sphere are big fans of the original novel. They acknowledge how difficult it will be to adapt the novel, but they trust Denis Villeneuve to do a good job.
- We found a big thread in the conversation about Alejandro Jodorowsky's original concept for the film, anyhow great it would have been to see that film.

yves@etcusc.org