GROUP D

The Great Firewall of China

Encoding and Decoding techniques used in Chinese Censorship

NATURAL LANGUAGE PROCCESSING

IE MBD
Spring 2020

Table of Contents

Introduction

Natural language processing techniques cover a wide variety of practical applications. NLP can be used for everything from judging sentiment to identifying spam, but one of the most dynamic applications is through encoding and decoding. While the idea of "codes" might bring to mind Indiana Jones movies, the modern use cases more often concern hiding content from government and private companies, and on the other hand, decoding said information. Tight censoring regulations in China have resulted in widespread use of modern language processing, with citizens developing an ever-evolving lexicon of codewords, while the government and social platforms try to keep up. This chase has led to a game of cat-and-mouse, that while deeply restrictive for the Chinese people, has also created practical applications for state-of-the-art natural language processing techniques.

China and Censorship

China is shaped by a unique cultural and historic background, which has greatly impacted interactions between citizens. The communist party in China utilizes censorship for several reasons: to eliminate illegal activity; to reduce undesired political discussions; to change the narrative about the cultural revolution; and ultimately, to maintain control over their population. China has received the worst possible ranking on the Freedom House index, and Freedom House has commented that " "state control over the news media in China is achieved through a complex combination of party monitoring of news content, legal restrictions on journalists, and financial incentives for self-censorship". The communist party has justified censorship by claiming it "protect[s] the country's culture"- under the guise of cultural protection, the party has censored content ranging from violent films to portrayals of LGBT characters in film and television [1].

Censorship in China comes in several flavors. The notorious Great Firewall of China is a catch-all type of censorship, where Chinese citizens are simply unable to access websites like Facebook. Soft censorship, on the other hand, uses NLP techniques to censor specific pieces of content [2]. This paper will be focusing on soft censorship. While sites like Facebook are blocked in China, comparing content on these sites to Chinese social media allows for insights to what content is most often censored.

In order for the Chinese people to communicate with each other and share their views on a variety of topics, they have turned to social media. Social media has developed into a sort of front-page news where Chinese citizens can be updated on topics that have been censored out of mainstream news [3]. Sina Weibo (the Chinese equivalent of Twitter) currently hosts half a billion users [1], and Renren (similar to Facebook) has another 257 million users [4].

While social media is censored, similar to news and television, this censorship often takes place after-the-fact, and at least temporarily, can serve as a more effective means of communicating citizens' feelings about 'sensitive' topics. Posts on popular social media sites can often be left up for days-to-weeks before being removed retroactively. This supports the claim that many social media sites are employing human censors, rather than relying exclusively on algorithms for detection. In fact, Seina Weibo, a microblogging social media platform, reportedly

employs up to 1,000 censors . In addition to this, China also utilizes Internet police (20,000 - 50,000 individuals) , and "50 cent party members" (volunteers paid by the government to manually check post content for censorship) which amounts to approximately 250,000-300,000 individuals [5]. The Chinese government relies upon domestic social media companies to patrol and censor their own websites under penalty of fines and shutdowns. In a sense, this creates a market for censorship, incentivizing social media platforms to perform their own illicit material detection. Generally, government- controlled web platforms have more rigorous censorship methods than their privately-owned counterparts. For instance, the same post may be blocked right away on one platform, but remain on the other. As in the graph below, different Chinese blogging websites censored anywhere from 1-60 of 108 posts made on their sites [7].

Different Chinese social platforms supressing the same censored posts.

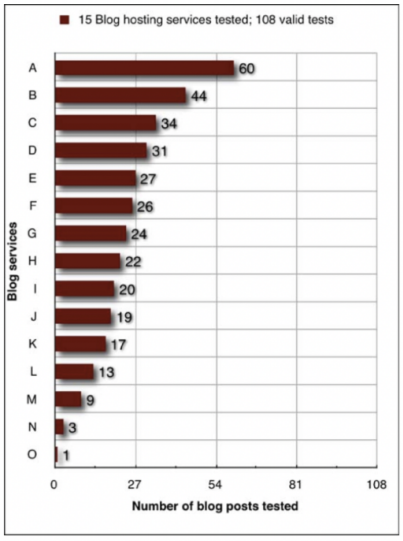


Figure 1

Regional Censorship

In an analysis of Sina Weibo, over 16% of messages are deleted. However, deletion of messages isn't consistent throughout China - outlying provinces have a much higher deletion rate (over 50%) than in areas like Beijing (around 13%) [2]. A higher deletion rate in outlying provinces could be due to greater censorship enforcements by the government, but could also be due to people in these regions self-censoring at a higher rate. As Chen et.al noted, up to "82% of tweets in some topics [are] censored [8]. However... censorship of a topic correlates with high user engagement, suggesting that censorship does not stifle discussion of sensitive topics." Additionally, as a topic becomes more censored, the community adapts by developing new language to discuss said topic- this can push down the censorship rate, but encoding also makes it more difficult to accurately gauge the censorship rate from the get-go. The overall rate of censorship on Chinese

Figure 2: Map of censorship

Posts originating from regions in conflict, such as Tibet and Qinghai,deleted at a higher rate than posts from other areas

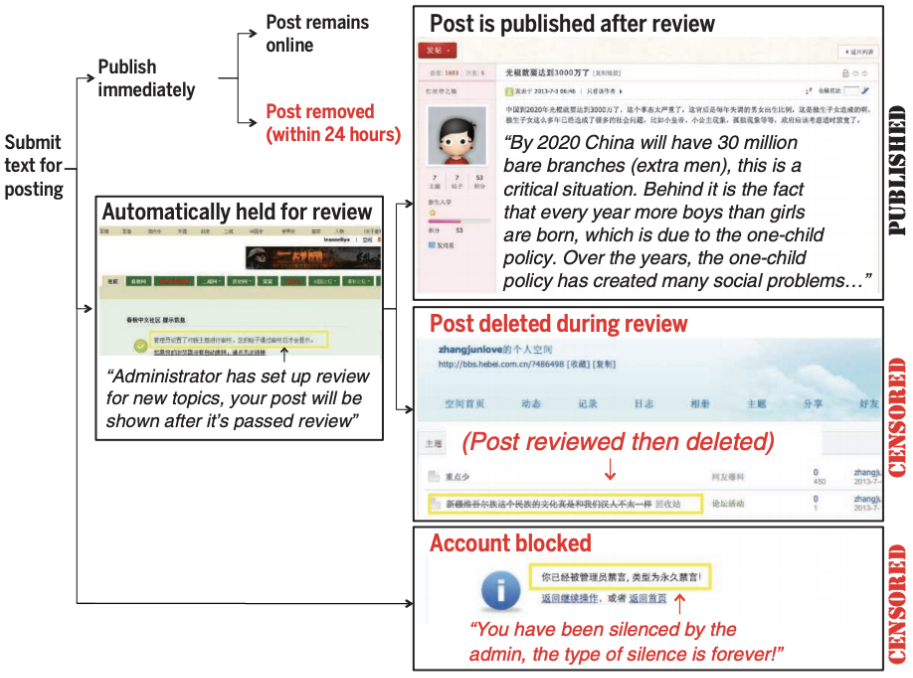
social media platforms varies greatly, with estimates ranging anywhere from 0.01% to 16% [8]. Chen also notes that censorship is applied unevenly across China, with "Users who discuss political issues and minority groups" bearing the brunt of censorship.

While individual posts are often censored as they are uploaded, search censorship is also widely used, where users are prohibited from searching specific terms [2]. Even if a user is trying to search positive information about a political leader, because "pro" and "con" might be the only indication that a post is for or against them, the user will simply not be able to search for that leader. For a more targeted filtering attempt, Chinese companies could consider using sentiment analysis to determine whether a post has a positive or negative view of a given public figure, allowing more targeted censorship. Commonly deleted posts often refer to controversial political figures, planning of protests, and posts near specific significant days (eg. Tiananmen square protest anniversary) [2].

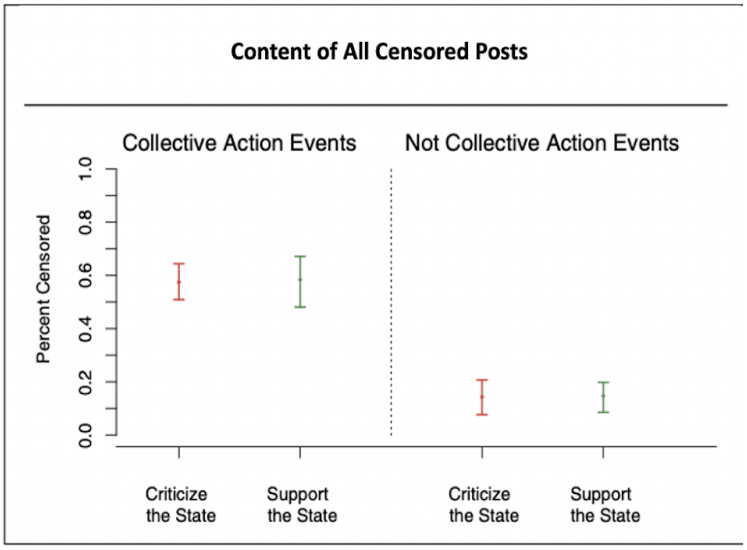
A major issue that arises when trying to detect this material is that sensitive material is often encoded to avoid censorship. This could take the form of anything from replacing the name of a controversial figure with a nickname, to developing a whole dictionary of codewords to refer to a topic. While encoded material can be easy to spot by the human eye, detection by algorithms offers several challenges- in addition to processing Chinese characters (which offers additional barriers in comparison to English), detecting encoding requires an understanding of the cultural and historic context of a tweet (censorship and deletion practices) [2].

Censorship Lifecycle

According to "*Reverse-engineering censorship in China: Randomized experimentation and participant observation*" by Harvard scholars Gary King, Jennifer Pan, Margaret E. Roberts, Chinese censorship of individual social media posts occurs in two phases. First, posts are manually censored by workers in the government and social media firms. Then, those posts are labeled and used to train algorithms in order to conduct automatic censoring. The figure below illustrates possible outcomes of a post on a Chinese social media platform [9]. Most posts are censored (out of experimental posts on 15 social platforms, 11 were deleted) [10].



Many Posts, Few Resources



Since analyzing exobytes of text generated each minute requires both massive computational and human labour, it is imperative to manage the distribution of resources in a form that helps to detect and eliminate censored material within the shortest period possible. There are two main factors that contribute to the difficulty of censoring: high social involvement (number of active internet users per capita is one of the largest in the world) and population of 2 billion people. The sheer quantity of social engagement happening in China each day surpasses the bureaucratic capacity of the state. The combination of these two factors, coupled with high population density, possess a real danger to the government of expansion past the digital realm.

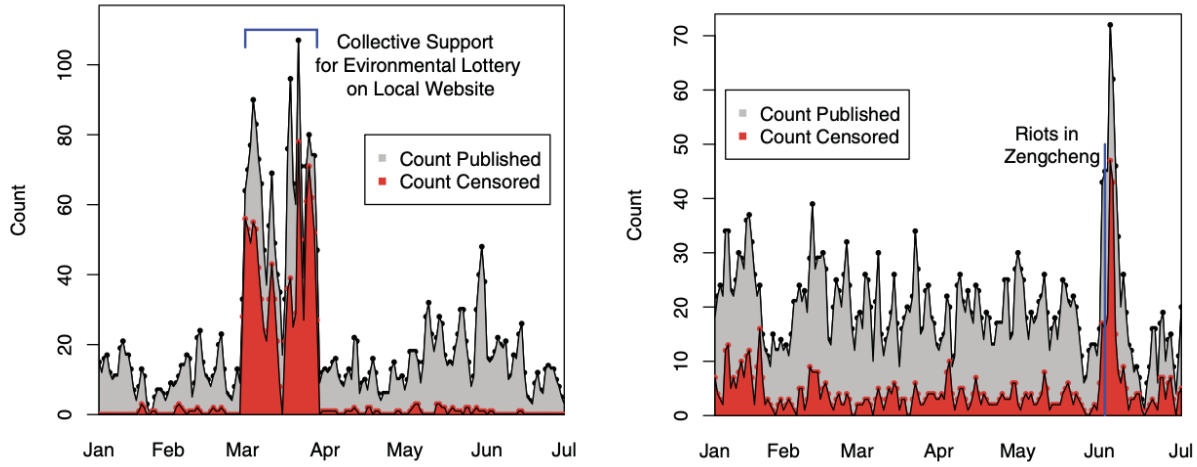
In order to efficiently distribute resources, web platforms may use preliminary statistical techniques to partition vast amounts of data into topic-based chunks [5]. This sub-categorization simplifies the search for both human censors and machine algorithms. Assigning each post to a predefined topic may also allow the usage of more specified techniques based on the domain features.

The methods and algorithms used by both the Chinese government and various online platforms are rigorously protected from external access. Given this, the best way to learn about these censorship techniques is by hypothesizing based on current NLP research and validation based on experiments that imitate the web-posting activity of Chinese citizens.

King et.al. outline that increases in posts (called “volume bursts”) about a given topic generally occur after an event relating to this topic has taken place (ie. elections, political protests, etc.). These volume bursts can serve as identification for the government to focus resources into that specific area and increase the rate of censorship. As shown in the table below these volume bursts also tend to lead to an overall higher censorship rate about said topic. It’s important to note here that the purpose of censorship is not to suppress all criticism towards the Chinese government, but rather to ensure state stability and discourage disorder by suppressing information spread that might provoke collective actions. Volume burst combined with a call-to-action are highly censored, while simply having a volume burst doesn’t necessarily lead to censorship [9].

Below we can observe that indeed the censorship rate is higher during the occurrence of related events, which confirms the hypothesis that volume bursts lead to an increase of censorship rates on platforms.

Figure 5. High Censorship During Collective Action Events (in 2011)



(a) Chen Fei's Environmental Lottery

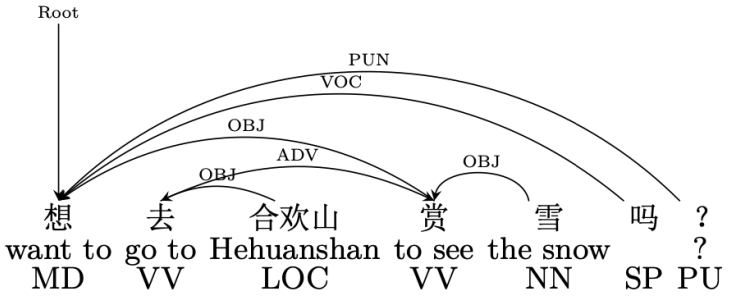
(b) Riots in Zengcheng

Challenges with NLP and Characters

Translating

To find censored terms, words need to be identified from Chinese characters- this is called Chinese Word Segmentation (CWS). A word can be made of up to 5 characters, and spaces aren't used between words, making them more difficult to identify. The meaning of a specific character isn't complete, and relies upon the other characters in the word to shape the context. For example, the Chinese word for cucumber is 黄瓜, with the first character meaning "yellow" and the second meaning "mellon". For English language speakers, Chinese must also be converted into English. Converting Chinese to English also offers the benefit of access to english language NLP libraries.

Chen et.al. suggest translating by creating a Chinese-English dictionary using Chinese wikipedia entries that have an English equivalent. Once a dictionary is created, an n-gram technique is used to check all character combinations up to 5 characters in a lexicon [2]. Another approach to Chinese word identification proposed by Gao et.al. is to identify five types of words using "(1) word breaking, (2) morphological analysis, (3) factoid detection, (4) named entity recognition (NER), and (5) new word identification (NWI)". This technique first predicts the word types by considering the likelihood of each word belonging to a word type. Then, it estimates "the likelihood of a character string, given the word type". Because words in Chinese are composed of characters with different meanings, the context with which characters are used can impact the meaning of the word. Xiong et.al describe that "for example, "自觉" is a word that usually means



"conscientiously" in the newswire domain, but are two words "自" (self) and "觉" (feel) that mean "feels ... by himself/herself" in the clinical domain" [11].

Processing characters also adds difficulty to POS tagging, given that there are no morphological changes between the different forms of a word. For example, "there are different morphologies in English for the word "毁灭 (destroy)", such as "destroyed", "destroying" and "destruction". But in Chinese, there is only one form". This difficulty has resulted in poorer performance in Chinese POS taggers, with state-of-the-art taggers performing around 93% (compared to 97% for English) [11]. The below image illustrates how a POS tagger can be used for Chinese characters.

Once parts of speech are identified, Name Entity Recognition (NER) identifies names of people, locations, and organizations. While identifying these in English simply requires spotting a capital letter (ie. "Madrid", "Jose", "Instituto de Empresa"), Chinese characters can't be capitalized, adding difficulty to making these identifications. Without a capital letter, NER relies highly on context. As Wu et.al. commented, training a Chinese NER can be time-consuming to annotate as well as computationally expensive. In their paper, they discuss using a CNN technique to jointly train a model to identify NER and word segmentation.

One final struggle faced in tagging Chinese is identifying temporal phrases (when something happened, how long it occurred, dates, times, etc). China uses both the Gregorian and lunar calendar, with the Lunar calendar being used primarily to mark important holidays. The use of both calendars means that Lunar calendar dates need to be converted to Gregorian calendars to be processed. Additionally, both Arabic (1, 2, 3, … 9) and Chinese (一, 二, 三, … 九) numerals are used for counting, as well as the English "AM" and "PM" [11].

Methods of Encoding

Chinese citizens utilize a variety of techniques to bypass censorship and communicate through social media without having to encrypt. This challenge of encoding and decoding is known as the "prisoners' problem", where two prisoners want to exchange secret information but must pass notes through a warden, so have to hide information in plain sight, without agreeing on specific codes [12].

Most morphs are encoded based on semantic meaning and background knowledge instead of lexical changes, so they are closer to jargon. A good code is something that is easy to remember

and understand, and can be easily spread through a social network, without being flagged by censors. While most codes are generated organically by users, Zhang et.al. have proposed several methodologies to generate codes [13]. Similar descriptions of techniques can be found in a variety of NLP research, but these techniques must be modified when being used with characters, rather than spelled words.

Phonetic substitution: Similar to English pronunciation, Chinese pronunciation has several sounds that are similarly pronounced (g, k), (z, c), etc. By substituting in characters that are phonetically similar (in that they sound the same but may be written differently), codes can easily be created. An example suggested by Zhang is to "substitute the characters of "比尔 盖茨 (Bill Gates) [Bi Er Gai Ci]" with "鼻耳 (Nose and ear) [Bi Er]" and "盖子 (Lid) [Gai Zi]" to form new morph "鼻耳 盖子 (Nose and ear Lid) [Bi Er Gai Zi]". Zhang notes that those codes with a negative connotation are perceived as funnier, and therefore more easily remembered. [13]

Spelling decomposition: by breaking down Chinese characters into their building blocks (radicals) and then combining these radicals in new ways, a huge range of possible codes can be created. For example, the radical "艹" could be converted into "草" (grass). [13]

Nickname generator: by taking the last character of a name and repeating it, a nickname can be easily generated. For example,,"_"幂幂 (Mimi)" to refer to "杨 幂 (Yang Mi)"". This technique also frequently works in English, by taking the first syllable of a name and repeating it (eg. JoJo for Jonas). [13]

Translation and transliteration: this technique works by translating Chinese names into English and then identifying if any components of the translated word are common English words. Those common words are then translated back as a new word, rather than the name. For example, the nickname ""拉里 鸟儿 (Larry bird)" for "拉里 伯德 (Larry Bird)" could be created by replacing the last name "伯德 (Bird)" with its Chinese translation "鸟儿 (bird)"" [13]

Semantic interpretation: character dictionaries offer definitions for specific Chinese characters. This technique works by searching the characters of an entity in a character dictionary, and checking to see if that character is used in the definition as well. If it is, that word using the character can be replaced in the original entity. Zhang gives the example of ""薄 胡来 (Bo Mess)" for "薄熙来 (Bo Xi Lai)" because the semantic interpretation sentence for "来 (Lai)" includes a negative word "胡来 (Mess)"". The English equivalent of this would be replacing a last name with a similarly- spelled name. [13]

Historical figure mapping: this technique maps famous historical figures, and ranks them based on their similarity to a given entity. Morphs are then generated based on the most similarly-mapped historical figure. For example, ""太祖 (the First Emperor)" is generated for "毛泽东 (Mao Zedong)" who is the first chairman of P. R. China and "高祖 (the Second Emperor)" for "邓小平 (Deng Xiaoping)" who succeeded Mao". [13]

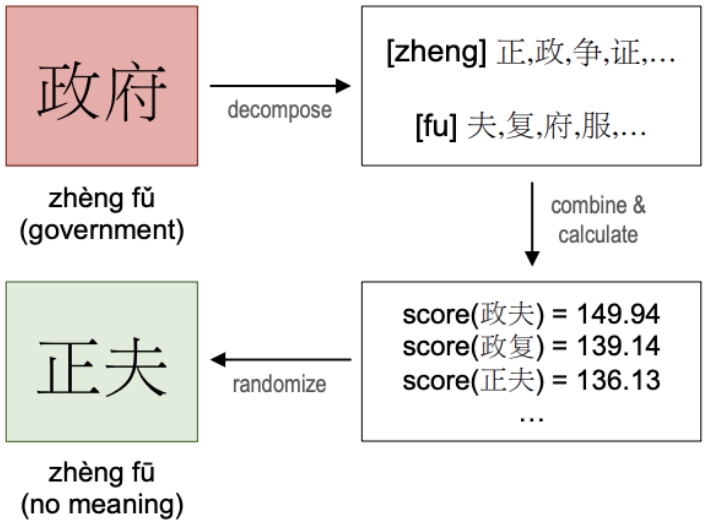
Characteristic Modeling: this technique uses several large corpora as well as news and web documents, and maps the similarity of an entity to the words in each corpora. Words are then ranked on cosine similarity, and are appended to the entity's last name. Zhang describes that this technique "generate many vivid morphs such as "**姚 奇才** (Yao Wizard)" for "**姚 明** (Yao Ming)"" [13]

Linguistic steganography: a method not noted by Zhang, linguistic steganography involves using word ordering techniques alongside existing embedding algorithms to hide information in plain sight. [13]

Homophones: One of the most common methods currently used for avoiding censorship in China is replacing censored keywords with homophones. Given that mandarin has many characters that refer to the same sound, replacement with homophones is fitting for the language. As Li et.al noted, "80% of the monosyllable sounds are ambiguous, with half of them having five or more corresponding characters"

Due to the specifics of mandarin, two words with distinct meaning and spelling can have the same or very similar pronunciation. For example, "when a river crab meme spread across Sina Weibo, it did not really refer to river crabs. Rather, it stood for a protest against Internet censorship, as the word for harmonize (**和谐**, pronounced he xi ´ e´,), slang for censorship, is a homophone of the word for river crab (**河蟹**, pronounced he´ xie`) (Zuckerman 2008)". While hard to identify by machines, homophones are highly interpretable by users. The main issue is that if the censoring system is using n-grams based word frequency models to detect oddity of sentences, homophones are still quickly detained by automated censorship systems.

One of the most successful and promising algorithms is based on a non-deterministic approach, where after transforming the censored keyword into its homophonic substitution, it finds the ones with the highest frequency score. The logic behind a high frequency score is that since those characters also appear frequently in all regular posts, they are a less unusual combination within the text and also more easily interpretable by users. In fact, many Chinese people might not even know the pronunciation of rare characters. Closer to the algorithm itself. It was created by scholars from Georgia Institute of Technology and tested on the platform similar to Twitter called Sina Weibo.



As seen from the figure above, the word Government is composed of 2 characters that are pronounced "zheng" and "fu" respectively. There are many other Chinese characters with the same pronunciation, creating many possible

homophones, ignoring the dictionary (if such word exists or not), then calculate the frequency score of each created homophone. In order to avoid rare occurrences that increase the detectability, only the 20 most frequent homophones are selected. Then to avoid the usage of only one homophone substitution for all posts, different homophones out of the top 20 are randomly selected. This makes detection more difficult for censorship systems to add to a list of censored keywords. Also "adversaries cannot simply add all homophones of censored keywords to a blocked keyword list because it would mistakenly censor a large portion of Sina Weibo's daily messages (one estimate in this paper suggests a figure of 20M posts per day, or 20% of daily messages). Rather, it seems likely that Sina Weibo would have to turn to human labor to defeat it."

This results in posts using homophones staying on social sites up to 3 times longer, and leading to higher interpretability by mandarin speakers. In a test, 99% of homophone-transformed words were understood by native speakers. This additionally results in more costly censorship for social media platforms- "coping with homophone transformations is likely to cost the Sina Weibo censorship apparatus an additional 15 hours of human labor per day, per censored keyword."

Algorithm 1: Homophone generation

GetTopHphone

Input: W: Word for which to generate homophone
Output: $\widetilde{W}$: A homophone of W with frequency score in the top k

$\widetilde{W}_h \leftarrow GenHphone(W)[rand(1,k)]$
$n \leftarrow len(W)$
if $n < 4$ **then**
  $\widetilde{W} \leftarrow \widetilde{W}_h$
else if $n = 4$ **then**
  $\widetilde{W} \leftarrow rand(\{\tilde{w}_h^1\tilde{w}_h^2, \tilde{w}_h^3\tilde{w}_h^4\})$
else if $n = 5$ **then**
  $\widetilde{W} \leftarrow rand(\{\tilde{w}_h^1\tilde{w}_h^2, \tilde{w}_h^3\tilde{w}_h^4\tilde{w}_h^5, \tilde{w}_h^1\tilde{w}_h^2\tilde{w}_h^3, \tilde{w}_h^4\tilde{w}_h^5\})$
return $\widetilde{W}$

Methods of Decoding

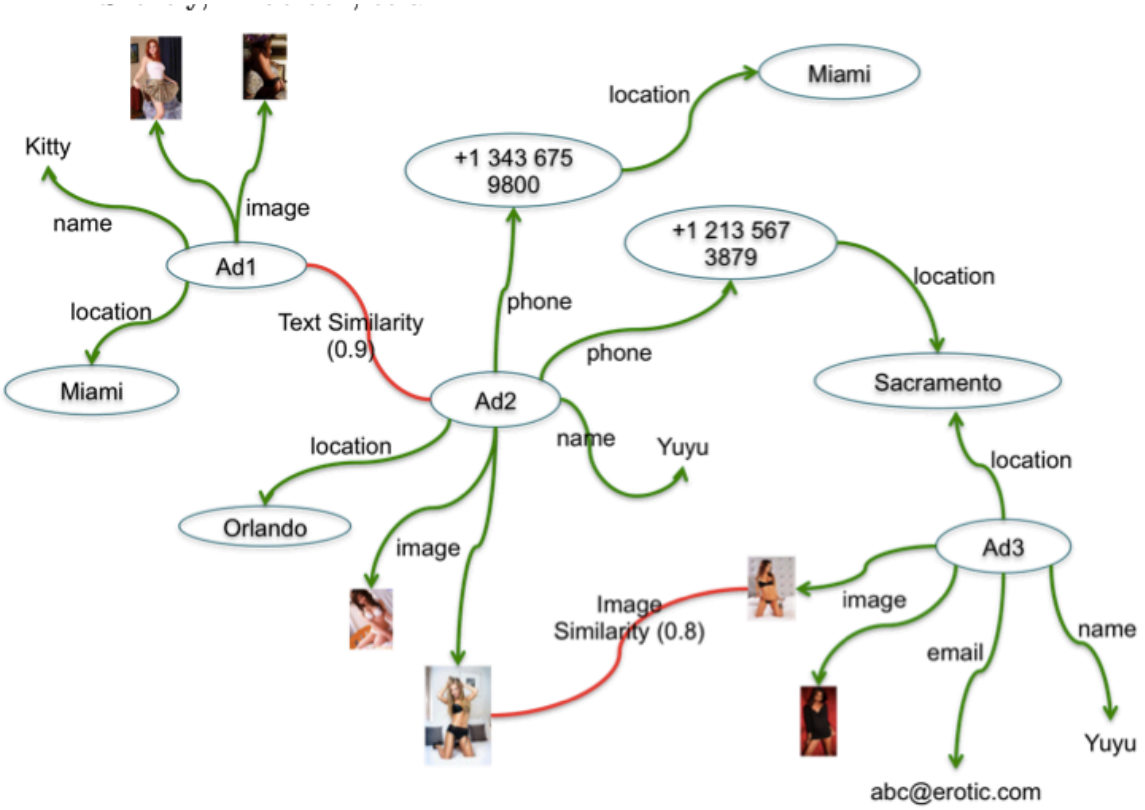
In order for both the Chinese government and private companies to flag sensitive material, several methods of decoding are used. These methods range from those that are meant to decode illicit information through the use of knowledge graphs in order to visualize and identify online ads that could potentially be related to illegal activities (DIG system). To other techniques that aim at decrypting the specific meaning of the words and the overall context of messages, that are often manipulated by the user in order to hide the real information (Common Semantic Features, Poly Substitution, Maximum Entropy, and Context-Aware methods).

While specific material on how China applies these methods are scarce, by understanding these decoding methods, hypotheses can be made on specific application. These different methods that other organizations and governments are implementing, are summarized in this section and they can be potentially implemented along with other measures by the Chinese government as decoding methods for censoring sensitive information:

Domain-Insight-Graph System (DIG system)

Domain-Insight-Graph System (DIG System) is a useful tool in identifying illegal activity by connecting networks of suspicious material and building knowledge graphs. While the tool is used primarily for identifying human trafficking networks, there are also applications for other illegal networks [14]. DIG systems refers to a broad technology that permits fast construction of information graphs (edges and nodes) for specific content along with query, visualization tools and analysis capabilities that enable end-users (in this case institutions like governments or NGOs), to resolve complex problems such as illicit weapons, counterfeit electronics, identifying patent trolls, and identifying research trends in material science and autonomous systems [14]. Although DIG is not primarily a natural language processing tool, it incorporates NLP in identifying relationships.

The figure below illustrates an example of a knowledge graph. In this case, information on human trafficking is being graphed, with nodes representing ads and data obtained from these ads: images, phone numbers, etc. Other attributes such as title and text of ads and physical attributes (ethnicity, eye color, hair type, etc.), are also considered in the process. The graph incorporates edges which represent the output of analytic processes such as Jaccard similarity between the transcript of ads and resemblance of the images.



A key aspect in the DIG system during the data processing phase is the identification of similarities in the data extracted. DIG offers capabilities to subtract similarity for pictures and for text data by using DeepSentiBank, a deep convolutional neural networks approach. The method extracts around 2,000 features from each image and computes solid hash codes that are used to

retrieve similar images [14]. The most significant advantage of these hash codes is that the similarity algorithms do not have to be trained with other images in the field of interest.

The DIG system comes along with several challenges and limitations:

- No agreement between organizations upon APIs or schemas, since they tend to focus on different problems and encode their extractions differently. Some of them do it in relational databases, while others produce JSON objects,etc [14].

- Legal requirements (provenance): organizations need to trace back from the knowledge graph to the original documents as they must cite the raw documents from the web site providers [14].

- A general lack of knowledge by most organizations on the use of Semantic Web technologies [14].

Common Semantic Features(CSF).

Metaphor detection issues have gained much attention in natural language processing research recently. The decoding and interpretation of metaphors is crucial for many practical language processing tasks such as information extraction, summarization, opinion mining, and translation [15]. As using metaphors are a common technique to hide information in plain sight, being able to decode these on a large scale offers a huge advantage to both the Chinese government and private businesses.

The Common Semantic Features (CSF) method is cross-lingual (meaning it works for multiple languages). Therefore it could potentially be used to decode messages in Chinese, even though the model is originally trained in English [15]. Its main objective is identifying the presence of metaphors in a given piece of text. This technique works by establishing the basic connotation of a lexical unit and testing if this interpretation applies to the current context. CSF requires a dependency parser and a target English dictionary. The approach to detect metaphors is based on semantic and not lexical features [15].

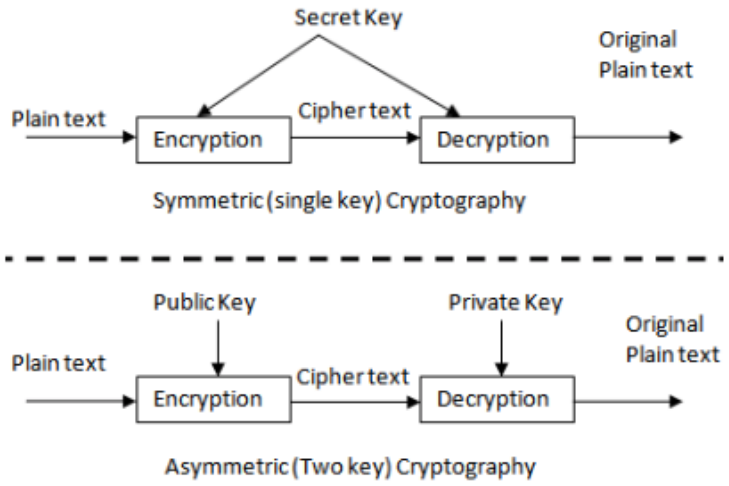
CSF works by firstly training a model using content from the “Wall Street Journal, WordNet (an English lexical database containing synonyms), the MRC dictionary (containing linguistic attributes),and finally, Word Representations via Global text, which is a collection of 100,000 words along with their vector representation” [15]. CSF considers the metaphor detection problem as a task of binary classification of sentences (literal sense or, otherwise, metaphoric) [15]. “Sentences are broken down into Subject-Verb-Object triples. These triplets are no more than a combination of features where the ‘S’ and ‘O’ parts contain information such as the semantic categories of a word, the degree of abstractness of the word and the types of named entities”. Then, a SVO relation is classified as literal or metaphorical using a Logistic Regression Classifier [15]. If all relations representing a sentence are classified literal by the CSF model, then the whole sentence is tagged literal. Otherwise, the sentence is tagged metaphoric [15].

Although the CSF model is considered to be cross-lingual, some languages such as Farsi have a bigger proportion of metaphors based on figurative use of adjectives and nouns [15]. This is a limitation that the model presents, and it would be useful to expand the set of semantic features in order to make it more precise in these types of languages.

Poly Substitution Method

This method could be used for both, encoding and decoding of any given text. The logic behind this technique is simple: if for example, two Chinese citizens want to communicate about a sensitive topic (say, a political leader), the first will encrypt their text, and will develop a cipher. Because the second citizen knows the encryption method, they will be able to use this (along with a secret key) to decipher the text.

To generalize, this method combines the features of Genetic Algorithms (inspired by darwin and genetic sequencing) and poly-substitution (where letters or characters are replaced) to generate ASCII values [16]. "ASCII, or American Standard Code for Information Interchange, replaces 128 English characters as numbers, with each letter assigned a number from 0 to 127" [18]. For instance, the ASCII code for uppercase M is 77 [16]. Overall, this method could be symmetric or asymmetric depending on the number of keys as we can see below. The encoding and decoding process could be summarize in the following steps:



Maximum Entropy (ME)-based model.

The Maximum-Entropy method (ME), is a metaphor detector (similar to the previously mentioned CSF method) that uses a logistic classification problem to identify metaphors. ME works as a multi-class problem, which allows more room for classification of different levels of metaphorical language [21]. The advantage of the ME is that it was trained using the Chinese language, specializing in Chinese nominal metaphor recognition and it is proven to give better accuracy and F1 scores when identifying Chinese metaphors [17].

Maximum Entropy works by looking at vocabulary and location of words. While ME functions similarly to a logistic regression, it tends to perform better for metaphor detection. For improved detection, semantic information could also be included. One of its biggest limitation

is its lack of identifying "noun+"的(of)"+noun" metaphorical patterns. It only recognizes metaphors in target single words [17].

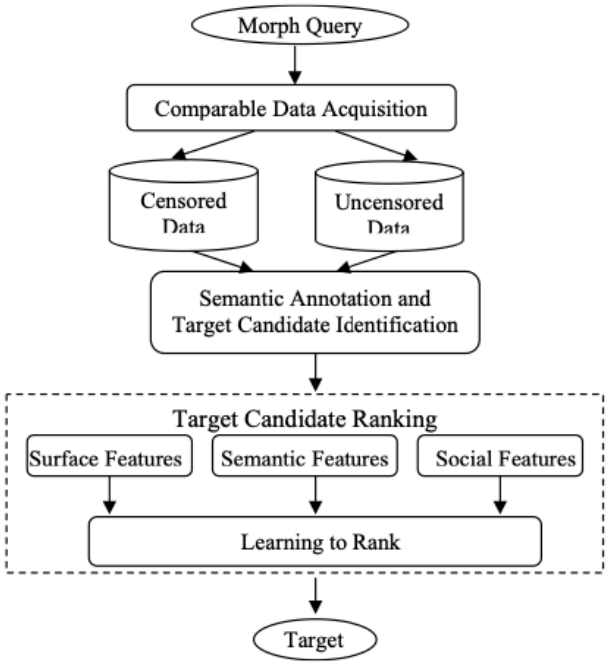
Context-Aware Entity Morph Decoding

Morphs refer to a special type of word that people create as an alternative name, to achieve a certain communication goal. "The Context-Aware system can mechanically identify, disambiguate, verify morph references based on specific contexts, and resolve them to target entities" [20].

Morph	Target	Motivation
Peace West King	Bo Xilai	Sensitive
Blind Man	Chen Guangcheng	Sensitive
Miracle Brother	Wang Yongping	Irony
Kim Fat	Kim Joing-il	Negative
Kimchi Country	South Korea	Vivid

What really characterizes morphs is that they tend to be informal, and usually evolve and have new meanings over time. For instance, "nearly all 平西王 (Conquer West King) appearances in Wikipedia refer to the ancient king instead of the modern politician 薄熙来 (Bo Xilai)" [20]. In the table below, you can see an example of different morphs, their target meaning and the motivation behind the morph:

This system first defines all the possible target meanings that a morph can have and ranks them. Finally, based on the ranking, it defines the final target node. This process is shown below [20]: The Context-Aware methodology also makes use of meta-path-based and semantic similarity measurements in order to decode morphs [20]. The model aims at measuring the similarity between two nodes over heterogeneous networks, by differentiating neighbors in three different types according to the network schema. Afterwards, meta-path-based similarity measures are put in place. These measures are defined over heterogeneous networks to extract semantic features. Moreover, the model also takes advantage of cosine-similarity normalization method, in order to make sure that each morph node is connected to the correct target node [20].



As a final thought, we can say that one of the biggest limitations that this model has is its low capacity to identify possible morphs from scratch, and

also discovering morphs for a specific target, based on anomaly analysis and textual coherence modeling [20].

Conclusion

As with many Natural Language processing techniques, one of the largest hurdles faced by encoding and decoding algorithms is the understanding of non-direct and metaphorical language. Decoding messages on Chinese social media compounds this issue by providing an ever-evolving set of metaphors and codes that are often difficult to spot even with the human eye. While there are a wide range of options to encode any language, character-based languages like mandarin offer a wider range of options. In addition to playing with homophones and metaphors, the similarity of many characters creates the opportunity for many new code words that are not possible with alphabet-based languages like English. Decoding on the side of the Chinese government and private businesses is a game of cat-and-mouse where the target continually changes. Difficulties in training models to identify coded language include the challenge of identifying words and entities in mandarin and not having access to an updated set of training data.

Sources

1. https://en.wikipedia.org/wiki/Censorship_in_China
2. https://firstmonday.org/ojs/index.php/fm/article/view/3943
3. http://www.ccs.neu.edu/home/leonchen/papers/weibo-cosn13.pdf
4. https://www.statista.com/statistics/227059/number-of-renren-com-users-in-china/
5. https://gking.harvard.edu/files/gking/files/censored.pdf
6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2104894
7. https://firstmonday.org/article/view/2378/2089
8. http://www.ccs.neu.edu/home/leonchen/papers/weibo-cosn13.pdf
9. https://gking.harvard.edu/publications/randomized-experimental-study-censorship-china
10. https://firstmonday.org/ojs/index.php/fm/article/view/2378
11. https://www.aclweb.org/anthology/J05-4005.pdf
12. https://www.cl.cam.ac.uk/~rja14/Papers/jsac98-limsteg.pdf
13. https://www.aclweb.org/anthology/P14-2115.pdf
14. https://usc-isi-i2.github.io/papers/szekely15-iswc.pdf
15. https://www.aclweb.org/anthology/W13-0906.pdf
16. https://www.researchgate.net/publication/44250729_Information_Security_Text_Encryption_and_Decryption_with_poly_substitution_method_and_combining_the_features_of_Cryptography
17. https://nadesnotes.wordpress.com/2016/09/05/natural-language-processing-nlp-fundamentals-maximum-entropy-maxent/
18. https://www.researchgate.net/publication/221628828_Chinese_Noun_Phrase_Metaphor_Recognition_with_Maximum_Entropy_Approach
19. http://comp.social.gatech.edu/papers/icwsm15.algorithmically.hiruncharoenvate.pdf
20. https://www.aclweb.org/anthology/P15-1057/
21. http://www.cs.cornell.edu/courses/cs5740/2016sp/resources/maxent.pdf