

Appendix to Catching Fire: National Humiliation,  
Emotional Contagion, and the Spread of Hostile Foreign  
Policy Preferences on Chinese Social Media

Michael Masterson

**Contents**

**1 Machine Learning Models** **2**

**2 Coding Dictionary** **3**

2.1 National Humiliation . . . . . 3

2.2 International Relations Outcomes . . . . . 4

2.3 Close Call . . . . . 7

**3 Keyword Lists** **7**

3.1 Keywords for National Humiliation . . . . . 7

3.2 Keywords for Policy Variables . . . . . 7

**4 References** **8**

# 1 Machine Learning Models

The content of the posts was preprocessed following Chang and Masterson (2020). The text was segmented using Jieba and punctuation and special characters were not removed (Sun 2015). Whether or not the content was vectorized by count or term frequency–inverse document frequency (TFIDF) was selected using cross-validation on the training set.

I follow Chang and Masterson (2020) on selecting models to classify texts with unbalanced categories. According to this paper, there are two few of the minority category posts in my training set to use Long Short-term Memory models. In these cases, they advise using either SVM models with weighted loss functions (so that misclassifications of the minority category are penalized more highly) or gradient boosting machines. I tried both and went with the SVM models with weighted loss functions because they performed better.

Models for each variable are trained and tuned independently. The table below shows the features that each model used in addition to the text content of the social media post as well as whether the text content was preprocessed by count or TFIDF vectorization. The features each model used were selected using cross-validation on the training set (with the exception of the model for national humiliation, which was only allowed to train on the text of the posts).

| Variable             | Additional Features Used   | Word Vectorization |
|----------------------|----------------------------|--------------------|
| National Humiliation | NA                         | Count              |
| Combined Force       | Device Type, and Post Year | TFIDF              |
| Territory            | NA                         | TFIDF              |
| Trade Barrier        | NA                         | TFIDF              |

Table 1 shows how each SVM model performed for each variable. To evaluate the models, I use a 0.8/0.2 train/test split. The models are highly accurate, each model classified at least 96% percent of the posts in the withheld test set correctly (approximating the 97% intercoder agreement among the human codings).<sup>1</sup> The models also perform well on the area under the receiver operating characteristic curve (AUC), which measures how well they are able to distinguish between classes. I also report precision and recall for the minority classes because this captures how often the models are correct when they label a post as being

---

<sup>1</sup>While the machine classifications are accurate, they are imperfect, so it is worth considering whether measurement error could bias my causal estimates. Researchers have sometimes argued that measurement error will, at worst, bias against their hypothesis by creating attenuation bias (Grumbach and Sahn 2020, 210). Unfortunately, this is only the case if the machine labels are uncorrelated with the residuals in the downstream regression and the classification errors are uncorrelated with observed covariates (Fong and Tyler, n.d.). While this bias cannot be estimated without knowing the true parameters, it is possible to reason about its severity. The first condition is isomorphic to the assumption made by any OLS regression that all independent variables are uncorrelated with the error term, so it is not uniquely of concern here. The second condition could be violated if, for example, the SVM that classified posts for military force was more or less likely to classify these posts accurately when a post contained a national humiliation narrative. While I cannot rule out this possibility, the facts that the training and evaluation of the models was conducted separately for national humiliation and the other variables means that the models classifying posts for the dependent variables were blind to whether the post also contained national humiliation. This makes it less likely that classification accuracy for the policy variables is correlated with the classifications of the national humiliation variable.

in the minority category and how many of the posts in the minority category are correctly labeled respectively. Labeling minority categories is a challenge for machine learning models, partly because the algorithm has fewer posts of this kind to learn from (Chang and Masterson 2020).

Table 1: Machine Learning Model Performance

| Variable             | Accuracy | AUC    | Minority Class Precision | Minority Class Recall |
|----------------------|----------|--------|--------------------------|-----------------------|
| National Humiliation | 0.965    | 0.8243 | 0.6792                   | 0.6667                |
| Force                | 0.986    | 0.6929 | 0.7                      | 0.3889                |
| Territory            | 0.964    | 0.8622 | 0.8                      | 0.7407                |
| Trade Barrier        | 0.996    | 0.7495 | 0.75                     | 0.5                   |

All results are for performance on the withheld test set using a 0.8/0.2 train/test split. The training set for national humiliation posts contains 4,000 posts, and the training set for each policy variable contains 3,999 posts.

## 2 Coding Dictionary

Each section defines a set of variables to be coded. If a post contains content that fits the definition for a variable, it should be coded as 1 for that variable. Otherwise, it should be coded as 0 for that variable. There are subsections for each variable containing examples of posts that would be coded 1 for each variable. The variables in the national humiliation section and the international relations section are coded in separate coding sessions and should be coded independently of each other. This means that if a post is coded as 1 for a particular international relations variable, such as military force, that does not provide any information about whether the post should be coded 1 for national humiliation.

### 2.1 National Humiliation

The post represents the nation of China as being humiliated. For an event to be framed as humiliation, it must be claimed that someone or some group outside the Chinese nation did something morally wrong/unjust to China, which was humiliated. The post can refer to something happening now or to a historic humiliation. It does not necessarily need to contain the phrase “national humiliation” (国耻).

#### 2.1.1 National Humiliation Examples

Example H1: Forward Weibo @ 北其村 The Japanese Invasion army signed the Instrument of Surrender. Celebrate the commemoration of the 68th anniversary of Japan’s surrender! Never forget national humiliation and revitalize China!

转发微博 @ 北其村侵华日军投降书之六（完）庆祝、纪念日本投降 68 周年! 不忘国耻, 振兴中华!

Example H2:

The Chinese, united, produce the strong voice of the nation. From self-reliance comes strength. The Chinese dream shines in the East. @ 中国之声 [The 82nd Anniversary of the “Manchurian Incident” never forget national humiliation] Today is the anniversary of the Manchurian Incident. Today, sharp air defense alarms resounded through the land of China. In the past 82 years, China has undergone earth-shaking changes. The spirit of national self-improvement cultivated in the War of Resistance Against Japan has helped China to overcome difficulties and obstacles and become increasingly powerful. History has proven that China does not disgrace itself. Who can disgrace it? As a Chinese, we should always remember: never forget national humiliation and revitalize China.

中国人，团结起来，奏响民族的最强音。自立，至强。中国梦，在东方闪耀。@ 中国之声【“九一八”事变 82 周年勿忘国耻】又到“九一八”纪念日。今天，尖利的防空警报再次响彻华夏大地。82 年过去，中国发生了翻天覆地的变化，在抗日战争中培育起来的民族自强精神助推着中国走过艰难险阻，日益强大。历史证明，中华不自辱，谁人能辱之？作为中国人，我们时刻应该牢记：勿忘国耻，振兴中华

Example H3

It is thanks to Britain starting the Opium War that the Chinese people who have been humiliated for centuries will fight for the country and the nation on any occasion. Now itself in decline, Britain denounces the suffering it caused and feels a kind of sorrow. @ 太史公: The British media made a documentary praising their athletes as “the pride of England,” but criticizing Chinese athletes for fighting for “national honor.” Can this be the same country? @ 流回雪 # #Seeing the world in the blink of an eye The Olympics were originally pure, but some countries must put narrow political meanings on them all the time, they do not know how to respect people! They don’t know what the Olympic spirit is! From raising the national flag to deliberately picking on young athletes, everywhere they show their narrow-mindedness and ridiculous self-importance. Can this be effective? On the contrary! The dogs they raised will cooperate, but decline is not far.

正是拜英国发动的鸦片战争所赐，深受百年屈辱中国人才会在各种场合为国家而战，为民族而战。现在它自己没落了，却指责当年的受害者，感觉是一种哀鸣。@ 太史公说: 英国媒体为他们的运动员拍纪录片，歌颂他们是“英伦的骄傲”时，却指责中国的运动员为“国家荣誉”而战，难道，英国是国家，中国 @ 流回雪 # 明眼看天下 # 奥运会本来是纯净之地，但某些国家一定要把狭隘的政治含义无时无刻的镶嵌其中，他们根本不懂得如何尊重人！不懂得何谓奥运精神！从升国旗到刻意抹黑刁难小选手，处处显示了他们的小肚鸡肠和妄自尊大。这样普世能有效果吗？只能相反！他们养着的狗接了命令是会附和着叫的，但离落水不远了

Example H4

National humiliation, ethnic/racial humiliation, the Chinese people’s humiliation, only formidable talent can wipe humiliation away

国耻，民族耻，中国人的耻，唯有强大才能洗耻

## 2.2 International Relations Outcomes

The variables in this section are all about foreign policy responses that a post could advocate. To be coded as a 1, these cannot be posts that just describe events that took place in the

past (although these descriptions can be included as part of an argument about policy). The post has to take a position on the kind of policy that China should have.

### 2.2.1 Military Force

The post advocates using or threatening to use China's military force against another country. There are two separate variables for military force. `mil_force_imp` is coded 1 if the post's advocacy of the use or threat of military force is only implicit and 0 otherwise. Some posts in this category might say that talking will not solve this dispute, or criticize the government for *merely* issuing a statement condemning another country's incursion into Chinese territory. Others might refer to past conflicts and a current international dispute and say that China cannot tolerate or must put a stop to what is going on in the current dispute. In this case, the earlier reference to force implies that force will be used or threatened as a means to end the current dispute. The second category is `mil_force_exp`, which is coded 1 if the post explicitly advocates the use or threat of military force and 0 otherwise.

### 2.2.2 Implicit Military Force Examples

#### Example Imp1

Kim Jong-un's current behavior is more arrogant than Kim Jong Il's. He is totally ignorant of China. North Korea has always done things in its own way. Even while resisting US aid, Kim Il Sung planned to drive the Chinese army out of North Korea. North Korea consumes the security benefits China provides but heedlessly takes actions that harm China's interests. China will not tolerate this. Much less will China tolerate North Korean criticism of China. Kim must understand that he can't humiliate China!

金正恩当前的行为比金正日更加放肆，对中国根本无视。朝鲜历来是我行我素，就连当初抗美援朝后，也是金日成一手策划把中国军队赶出朝鲜的。一味着靠吃中国给予其提供的安全保障而吃红利，却全然不顾损害中国利益的行为，中国不会包容，更不会容忍朝鲜评中国。金必须明白他不可以屈辱中国！

### 2.2.3 Explicit Military Force Examples

Example M1: Facing Japan's provocation and infringement, China should take the initiative and attack. [This] would make Japan know that China protects its sovereignty and territory with resolute determination. China cannot be passive. If [we] give them an inch on the Diaoyu Island outrages, they will take a mile. [China] cannot hesitate to take firm action to make Japan pay a painful price for its mistakes.

面对日本的挑衅和侵犯，中国应该应该主动出击，让日本真正认识到中国捍卫领土和主权的坚定决心，不能处处被动，任其在钓鱼岛上胡作非为得寸进尺；应该果断行动，让日本为其错误的行径付出惨痛的代价。

#### Example M2:

China will not give territory in exchange for peace. Do not hesitate to go to war to defend [our] interests <http://t.cn/S5R0R2>

罗援：中国不会以土地换和平捍卫利益不惜一战 <http://t.cn/S5R0R2>

#### Example M3:

The power of our national unity is diplomatic and combat power. Yes, *The Art of War* says the best policy is to avoid war, but we must fight without hesitation! @ 工者 [seeking the patriotic signatures: asking the People’s Liberation Army to fight the Philippines]: 1 The Philippines is gambling the Chinese people lack the moral backbone, betting our strategic decision making is weak. 2 We lost Outer Mongolia. Will we also lose the South China Sea? Will we also lose our navigation rights? 3 The sea area near Huangyan Island contains the oil and natural gas needed for future development. Now, if [we] do not fight, [we] will not completely destroy the Aquino administration’s hopes for the United States. China will completely lose its resources. 5 The people need an unyielding voice! !

国人的团结力就是外交战斗力，也是孙子兵法不战而屈人之兵的上策，但我们要不惜一战! @ 践行者周宇【求爱国连署：请求解放军对菲律宾开战】理由如下：1 菲律宾赌中国人没骨气，赌决策软弱；2 我们失去外蒙，还能失去南海吗？还要失去本属于我们的航行自由？3 黄岩岛附近海域蕴藏着未来发展必需的石油和天然气，现在不打，不彻底摧毁阿基诺政府对美国的奢望，中国将彻底失去资源；5 民间需要强硬的声音!!

#### 2.2.4 Territorial Dispute

The post refers to one or more of China’s territorial disputes. Examples of territorial disputes include but are not limited to Taiwan, the Diaoyu/East China Sea dispute, the South China Sea (including the Paracel Islands and the Spratly Islands), China’s territorial disputes over its border with India, and China’s territorial disputes over its border with Russia. The post has to refer to the dispute as a dispute or claim that the disputed territory is part of China. For example, posts about the weather in the South China Sea or that mention that the media in Taiwan are reporting something should not be coded as 1.

#### 2.2.5 Territorial Dispute Examples

Example T1: “Taiwan is China’s sacred territory” 台湾是中国神圣的领土

M1, M2, M3, and TB1 are also examples of posts that mention China’s territorial disputes, but not all posts advocating military force or military spending and resources will necessarily mention territorial disputes.

#### 2.2.6 Trade Barrier

The post advocates boycotting or raising trade barriers to goods from one or more other countries. Trade barriers include both tariffs and non-tariff barriers such as increased inspections. The post does not need to give a reason for why Chinese consumers should boycott this country’s goods or why the Chinese government should raise trade barriers.

#### 2.2.7 Trade Barrier Example

Example TB1: “Resisting Japanese goods and not trading with Japanese people is the humiliated Chinese nation’s most fundamental principle. However, pouring out soup, throwing water bottles, pulling down the national flag, and breaking cars are acts of bandits and villains. @ 冷雷冷雨: @ 金毛鼠 007: Resolutely boycott Japanese goods until China fully

controls the Diaoyu Islands ! @ 海浩瀚: Resist the Japanese goods, it should be normal behavior, even the totally subservient must persist. @ 网络新趋势 The Chinese must defend the Diaoyu islands from the Japanese dwarf [derogatory] pirates. After the national humiliation of The Treaty of Shimonoseki, the butcher knife lifted at Nanjing. China's big market towers over Japan's small sprouts. Japanese goods are suddenly on fire; gentlemen do not forgive easily. Reject Japanese goods for 1 month and 1000 Japanese companies will teeter. Reject Japanese goods for half a year and half of Japanese enterprise will flee. Reject Japanese goods for a year, and patriotic blood will burn. A Chinese gentleman will forward this post one hundred times. Everyone forward this letter. The devil is coming. -Forward!" \ 抵制日货，不与任何日本人往来，这是饱受屈辱中国人两条最基本的原则。但是，泼面汤、扔水瓶、拔国旗、砸车、拧车螺丝都是暴徒与小人所为 @ 冷雷冷雨: @ 金毛鼠 007 : 坚决抵制日货，直到中国全面控制钓鱼岛! @ 大海浩瀚: 抵制日货，应该作为常态，即使他俯首帖耳，也要坚持不懈 @ 网络新趋势 华人必保钓，倭寇莫叫嚣。马关丧国耻，南京举屠刀。中国大市场，扶桑小命苗。日货猛于火，诸君莫轻饶。一月拒日货，日企千家翘。半载拒日货，日企半数逃。一年拒日货，日本经济倒。年年拒日货，爱国血在烧。君是中国人，转发一百条。人人转此信，鬼子末路到。-转!

## 2.3 Close Call

For every variable that is coded as a 0 or a 1 there is a close variable for that category that is coded as a 1 if the coder is uncertain about how to code that variable based on the text of the post.

# 3 Keyword Lists

## 3.1 Keywords for National Humiliation

Posts selected as possibly containing national humiliation had to contain at least one of the following keywords:

国耻, 耻辱, 耻笑, 耻于, 辱, 屈辱, 侮, 靖国神社, 九一八, 南京大屠杀, 百年国, 冤枉, 贬低, 欺负, 鄙视, 卢沟桥, 八国联军, 侵华, 甲午战争.

## 3.2 Keywords for Policy Variables

The posts selected as possibly containing one or more of the policy variables could not contain the keyword "advertisement" (广告).

These posts also had to contain at least one of the following key words:

贸易, 大国, 钓鱼岛, 东海, 舰队, 领土, 主权, 中美, 南海, 访华, 访美, 外交, 军事, 安理会, 和平, 核心利益, 中方, 世贸, 安理会, WTO, wto, 对华, 关税, 战争, 冲突, 爱国, 攻击, 核武器, 导弹, 冷战, 霸权, 多边主义, 祖国, 民族荣耀, 民族利益, 航空母舰, 战斗机, 潜艇, 士兵, 将军, 海军, 空军, 人民解放军, 北约, 自卫队, 军国主义, 法西斯主义, 民族主义, 敌军, 边防, 战备, 老将, 战区, 陆军, 特务, 情报, 炸弹, 杀手锏, 富国强军, 不惜一战, 西沙, 王毅, 华春莹, 南海, 非政府组织, 钓鱼岛.

## 4 References

- Chang, Charles, and Michael Masterson. 2020. “Using Word Order in Political Text Classification with Long Short-Term Memory Models.” *Political Analysis* 28 (3): 395–411. <https://doi.org/10.1017/pan.2019.46>.
- Fong, Christian J., and Matthew Tyler. n.d. “Machine Learning Predictions as Regression Covariates.” [https://www.dropbox.com/s/tgptv3hbcqemga9/Predicted\\_Covariates.pdf?dl=0](https://www.dropbox.com/s/tgptv3hbcqemga9/Predicted_Covariates.pdf?dl=0).
- Grumbach, Jacob M., and Alexander Sahn. 2020. “Race and Representation in Campaign Finance.” *American Political Science Review* 114 (1): 206–21. <https://doi.org/10.1017/S0003055419000637>.
- Sun, Junyi. 2015. “Jieba.” Baidu. <https://github.com/fxsjy/jieba>.