CYBER SECURITY STRATEGIES ANALYTICS

Natural Language Processing + Machine Learning

Machine-based Text Analytics of National Cybersecurity Strategies

Yanlin Chen, Yunjian Wei, Yifan Yu, Wen Xue, Xianya Qin
https://github.com/Ychen463/Cyber

Introduction

1.1 *Objective*

We try to build an automatic sentence classification system that can automatically update training data when given new categories. And build a tool to analyze the text of Cybersecurity strategies of more than 75 countries to find their commonalities, differences, and key characteristics.

1.2 *Importance*

As time changes the needs of users also change. The topic and measures of cyber security change as society and information techniques develop. So, the training data for the model may be outdated. If the system could update the training data itself, it will save lots of time. But the hardest problem of building a model is to find the properly label the training data. That's why we come to search engine: use search engine to find category-related raw data and filter them to training data.

Methodology

2.1 Supervised Learning

2.1.1 Data Collection

Based on the suggested labels given by the headers in the cyberwallets profiles, we search category keywords in google and use Selenium to crawl the content of top 20 search results and use this data as training data. Our training goal is to optimize the accuracy of classifying a sentence. However, the performance of this method depends largely on the accuracy of the search engine and the result was not as good as expected. So, we reduce the crawling range to high-quality PDF documents and it turns out that PDF files have even more text than web pages. Then we download the most relevant 10 PDF files for each sub-category and use the content of these files as training data.

2.1.2 Data Processing

The policy documents were originally in PDF format. We use python with PDFMiner to transform into text file. In this way we can get the content of the file as well as its location (page number) in the original document file. The text was in bad format at the beginning, we have tried multiple ways to separate into sentences and remove unreadable gibberish and punctuations. From the documents of 63 countries, 22053 sentences were extracted at last, which are assigned with different document ID, Sentence ID (unique), Page No. and Sentence No. of the Document.

2.1.3 Text Mining

We use word2vec method to project each word into a 100 dimension and numeric vector so that similar words will be close to each other in the vector space. It makes the model robust to synonym. Then we use CNN to capture the context of a sentence, which means CNN can predict data on the sentence level other than word level. Combining the two methods together makes our model have a strong generalization ability.

2.2 Unsupervised Learning

Unsupervised learning is a kind of machine learning algorithm without labeled target compared with supervised learning. The aim of unsupervised learning is to learn the hidden patterns of input data. The common method is cluster analysis, which can help us find similarities between each input data. Compared with supervised learning, unsupervised learning cannot tell us what this cluster is. We can only manually label our result for future use.

2.2.1 Data Processing

First, we apply a common way to deal with the raw data. Tokenize whole text to words for future TF-IDF matrix calculation. Removing some meaningless but high frequent words is very important, in case these words would influence our results. Based on our exploration on the whole text, we add some words like “cyber”, “security” and” government”, because these words are exactly our topic keywords, and we only keep words longer than three characters because of English characteristic. Moreover, for the future words distance calculation, we remove country names. Meanwhile, words have different format in the text, so we break each word into their roots. In this step, we create a table to compare the original words and root in case we want to traceback the root word.

Then we calculate TF-IDF matrix for future dimension reduction. TF-IDF vectorizer means term frequency - inverse document frequency. It measures each word frequency in one document and how frequent this word appears in each document. If a word shows a high frequency in one document, and a lower percentage of document the very word in whole documents, it means the word is important and distinguished. To get a better result, we keep filtering some very high frequent words and some meaningless words like people’s name which show not frequent and carry no meaning. Meanwhile, not only create one-word tf-idf matrix, we still want to explore the combination of words. We creatively also look at bigrams and trigrams to see the possibility of each combination.

Therefore, we get a matrix which contains filtered words and it’s tf-idf weight with shape of 27297 and 5125.

2.2.2 PCA dimension reduction

Due to high dimension of our matrix, we decide to use principal component analysis to do dimension reduction. PCA can keep most of the characteristic of the data to present the whole data. In this matrix, we find 200 dimensions can explain about 70% of the data. Therefore, we only keep 200 columns to show our whole data.

2.2.3 Hierarchical clustering

Hierarchical clustering is a method of clustering analysis which seeks to build a hierarchical cluster between each point. We use hierarchical clustering to see clusters.

First, we use three methods to connect each point which are “single”, “complete” and “ward”. And compare the performance of each method. We found “ward” calculation method can show a better performance because it achieves score with 0.73, which is better if it is closer to 1.

Then we draw dendrogram to show hierarchical clustering to check how each similar pattern combined and connected. The greater difference in height, the greater dissimilarities between each group.

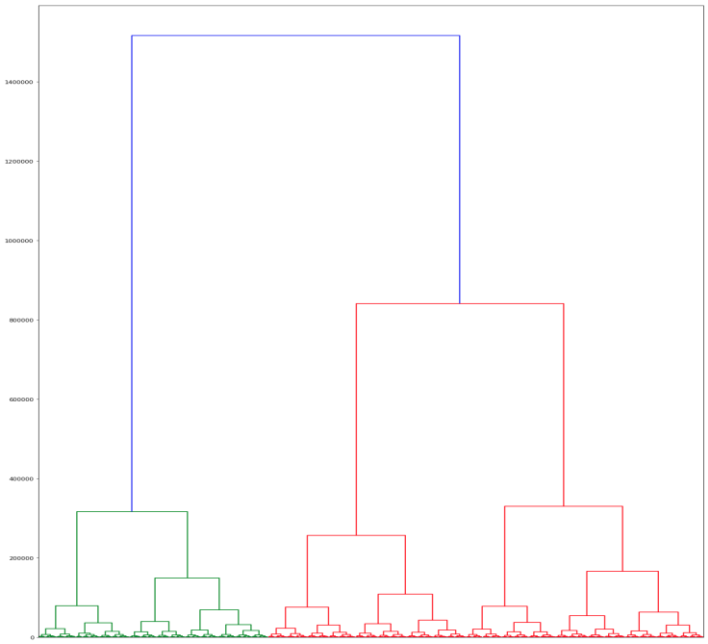


Figure 1. Hierarchical Clustering

Due to the specialty of text and the sparse of our matrix, we decide to use K-means to support our analysis. Meanwhile, results can be generated by different clustering methods to be more reasonable.

2.2.4 K-means

K-means is another method of clustering under unsupervised learning. It randomly chooses to point as centroid point, and calculate the distance to cluster each point. Then it would recalculate and correct the centroid point until it won’t change. And it is difficult to set the exact number of clusters.

Based on the hierarchical clustering, we decide to first to set the clustering number as six to see the overall category. And then compare the results under K-means and LDA to explore subcategories under each category.

2.2.5 LDA topic extraction

After six overall categories generated, we put these texts under different groups in LDA model we build to explore subcategories. We can extract important words under each group and find the similarities to form subthemes.

Result

3.1 Supervised Result

About 41% sentences are assigned ‘category unknown’ (see appendix 1), which is not surprised because we assume without context more than half of sentences are of no specific category.

The most popular topic is ‘organization measures’: about 17% sentences talk about it. And ‘Child Online Protection’ has the least sentences, which is 4.43%

3.2 Unsupervised Result

We decide to use six categories under unsupervised result. And use LDA model to generate different sub-themes.

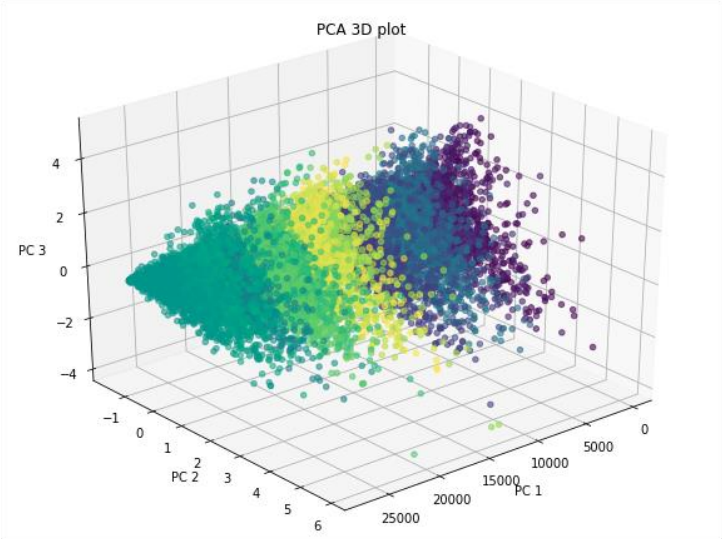


Figure 2. PCA 3D plot

The first cluster is capacity building from presented words. There are also four sub-categories under this cluster.

```
[(0,
  '0.013*"development" + 0.010*"objectives" + 0.009*"project" + 0.008*"in
creased" + 0.008*"initiatives" + 0.008*"research" + 0.007*"processes" + 0
.006*"focus" + 0.006*"service" + 0.006*"element"'),
 (1,
  '0.013*"environment" + 0.011*"development" + 0.010*"system" + 0.010*"pl
an" + 0.010*"infrastructure" + 0.009*"secure" + 0.008*"framework" + 0.006
*"project" + 0.006*"technology" + 0.005*"kingdom"'),
 (2,
  '0.011*"management" + 0.009*"industry" + 0.008*"threats" + 0.007*"aware
ness" + 0.006*"infrastructure" + 0.006*"response" + 0.006*"system" + 0.00
6*"cybercrime" + 0.005*"capabilities" + 0.005*"private"'),
 (3,
  '0.013*"international" + 0.010*"risk" + 0.009*"state" + 0.009*"ensure"
+ 0.008*"implementation" + 0.007*"management" + 0.007*"activities" + 0.00
7*"project" + 0.007*"standards" + 0.007*"protection"')]
```

Figure 3. Result of the 1st Clustering

The second group is about cooperation. At first, we remove words cooperation and international to achieve better performance. Words like “organization”, “agencies”, “support” tell us that this topic is around international cooperation.

```
[(0,
  '0.013*"access" + 0.012*"measures" + 0.012*"risk" + 0.012*"management"
+ 0.009*"risks" + 0.009*"business" + 0.008*"control" + 0.007*"organisatio
ns" + 0.007*"secure" + 0.007*"responsibility"'),
 (1,
  '0.012*"private" + 0.010*"development" + 0.009*"agencies" + 0.009*"work
" + 0.007*"develop" + 0.007*"defense" + 0.007*"response" + 0.006*"infrast
ructure" + 0.006*"department" + 0.006*"coordination"'),
 (2,
  '0.012*"ensure" + 0.009*"achieve" + 0.009*"organisations" + 0.007*"poli
cies" + 0.007*"standards" + 0.006*"outcomes" + 0.006*"help" + 0.006*"mand
ated" + 0.005*"secure" + 0.005*"uganda"'),
 (3,
  '0.020*"infrastructure" + 0.014*"communications" + 0.014*"critical" + 0
.010*"attacks" + 0.009*"networks" + 0.008*"public" + 0.006*"capabilities"
+ 0.006*"support" + 0.005*"vulnerabilities" + 0.005*"secure"'),
 (4,
  '0.009*"privacy" + 0.009*"technology" + 0.007*"march" + 0.007*"computer
" + 0.006*"industry" + 0.006*"education" + 0.006*"civil" + 0.005*"private
" + 0.005*"system" + 0.005*"science"')]
```

Figure 4. Result of the 2nd Clustering

Technical measures is the third cluster. Words like “development” and “technologies” help us learn more about this topic.

```
[(0,
  '0.018*"development" + 0.016*"research" + 0.014*"innovation" + 0.009*"a
ctivities" + 0.008*"economic" + 0.008*"major" + 0.007*"system" + 0.006*"s
cience" + 0.006*"support" + 0.006*"scientific"'),
 (1,
  '0.030*"technologies" + 0.011*"developing" + 0.009*"energy" + 0.009*"de
velop" + 0.009*"priorities" + 0.008*"ministry" + 0.008*"equipment" + 0.00
8*"system" + 0.007*"development" + 0.007*"resources"'),
 (2,
  '0.012*"private" + 0.011*"critical" + 0.010*"threats" + 0.009*"public"
+ 0.009*"infrastructure" + 0.006*"agencies" + 0.006*"computer" + 0.005*"r
esponse" + 0.005*"secure" + 0.005*"risks"')]
```

Figure 5. Result of the 3rd Clustering

These words are around the topic measurement of organization. Words show below indicate that these sentences are around different measurement to organize online policies.

```
[(0,
  '0.010*"awareness" + 0.008*"protection" + 0.007*"necessary" + 0.006*"co
mmunication" + 0.006*"risk" + 0.006*"level" + 0.006*"measures" + 0.006*"m
anagement" + 0.006*"relevant" + 0.006*"access"'),
 (1,
  '0.013*"cooperation" + 0.012*"development" + 0.011*"measures" + 0.009*"
response" + 0.009*"implementation" + 0.008*"necessary" + 0.007*"computer"
+ 0.007*"attacks" + 0.006*"incidents" + 0.006*"well"'),
 (2,
  '0.012*"development" + 0.009*"develop" + 0.008*"human" + 0.007*"telecom
munications" + 0.007*"access" + 0.006*"provide" + 0.006*"objective" + 0.0
05*"resources" + 0.005*"promote" + 0.005*"standards"'),
 (3,
  '0.013*"critical" + 0.008*"economic" + 0.008*"threats" + 0.007*"infrast
ructures" + 0.007*"citizens" + 0.006*"electronic" + 0.006*"secure" + 0.00
6*"private" + 0.006*"society" + 0.006*"development"')]
```

Figure 6. Result of the 4th Clustering

Child online protection is the fifth topic. Themes of education and protection are shown in the chart below.

```
[(0,
  '0.009*"society" + 0.007*"research" + 0.007*"centre" + 0.007*"situation
" + 0.007*"authorities" + 0.006*"framework" + 0.006*"education" + 0.006*"
state" + 0.005*"actors" + 0.005*"different"'),
 (1,
  '0.016*"businesses" + 0.012*"authorities" + 0.010*"society" + 0.009*"ci
tizens" + 0.009*"business" + 0.009*"solutions" + 0.006*"opportunities" +
0.005*"defence" + 0.005*"number" + 0.005*"functions"'),
 (2,
  '0.012*"access" + 0.010*"networks" + 0.009*"protection" + 0.007*"critic
al" + 0.007*"economic" + 0.006*"measures" + 0.006*"social" + 0.006*"secur
e" + 0.005*"ensure" + 0.005*"citizens"'),
 (3,
  '0.014*"content" + 0.010*"electronic" + 0.008*"implementation" + 0.008*
"management" + 0.007*"plan" + 0.007*"action" + 0.006*"well" + 0.006*"thre
ats" + 0.006*"educational" + 0.006*"process"')]
```

Figure 7. Result of the 5th Clustering

The last topic is the measurement of legal laws. Apparently, the second group is about criminal legislation and implementation.

```
[(0,
  '0.008*"critical" + 0.008*"threats" + 0.006*"ensure" + 0.006*"private"
+ 0.006*"measures" + 0.005*"capabilities" + 0.005*"defence" + 0.005*"prot
ection" + 0.005*"networks" + 0.005*"response"'),
 (1,
  '0.006*"crime" + 0.006*"implementation" + 0.005*"businesses" + 0.005*"r
isk" + 0.005*"industry" + 0.005*"research" + 0.004*"education" + 0.004*"p
rovide" + 0.004*"attacks" + 0.004*"access"')]
```

Figure 8. Result of the 6th Clustering

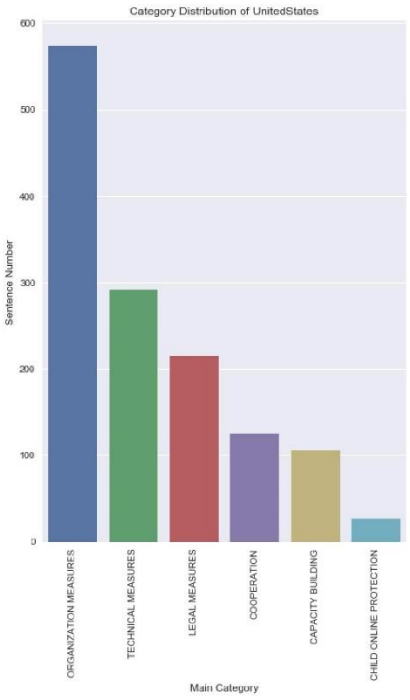
Demo of sentence search engine

This tool enables user to interact with the data and the classification result we got. Once the user chooses a country, the information of categorized sentences will show up, which can give user an overall understanding of the Cyber Security policies of this country. Users can choose different categories and subcategories which they are interested in and the tool will display detailed information of each sentences.

For example, when we choose country as “United States”:

According to the classification result, United States do well in Organization Measures, but need to improve in Child Online Protection.

(1) Category



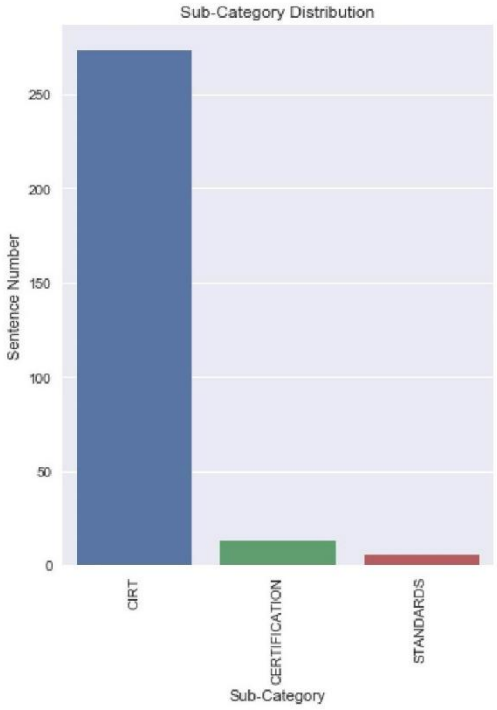

```
Country: UnitedStates
                    category  count
4    ORGANIZATION MEASURES    574
5       TECHNICAL MEASURES    291
3           LEGAL MEASURES    215
2              COOPERATION    125
0         CAPACITY BUILDING    106
1  CHILD ONLINE PROTECTION     26


 1. LEGAL MEASURES
 2. CAPACITY BUILDING
 3. TECHNICAL MEASURES
 4. COOPERATION
 5. CHILD ONLINE PROTECTION
 6. ORGANIZATION MEASURES

Choose a category you are interested in:3
Selected Category:
 TECHNICAL MEASURES
```

Figure 9. User Interaction with selected Category

(2) Sub category




```
Country:   UnitedStates
Category:  TECHNICAL MEASURES
Subcategories:
       1 . CIRT
       2 . CERTIFICATION
       3 . STANDARDS
Choose the subcategory you are interested in:3
STANDARDS
```

```
Country: UnitedStates
Category: TECHNICAL MEASURES
sub category
CERTIFICATION     13
CIRT             273
STANDARDS          5
```

Figure 10. User Interaction with selected Sub-Category

(3) Sentences of selected categories

a. All information

The final dataset used for the search engine has columns as follow, which enable user to do multiple search.

(1) Source Document
(2) Country
(3) Document ID
(4) Category
(5) Sub-Category
(6) Sentence
(7) Sentence No. in Document
(8) Sentence ID (Unique)
(9) Page No.

	Source Document	Country	Doc ID category	sub category	Sentence	Sentence_N	SentenceID	Page No.
22016	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	In the National Defense Authorization Act (NDAA) of , Congress required the Defense Department to designate a Principal Cyber Advisor to the Secretary of Defense to revi	277	8888277	77
22017	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	In addition, the Principal Cyber Advisor will govern the development of DoD cyberspace policy and strategy for the DoD enterprise.	278	8888278	77
22018	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	NATIONAL BENCHMARKING	The NDAA also stipulated that this Principal Cyber Advisor integrate the cyber expertise and perspectives of key organizations to build an intradepartmental team of key pla	279	8888279	77
22019	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	The Principal Cyber Advisor responsibilities assigned by the FY NDAA shall not be interpreted to affect the existing responsibilities and authorities of the Under Secretary of D	280	8888280	77
22020	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	ROADMAP FOR GOVERNANCE	An intradepartmental team.	281	8888281	77
22021	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	The CIMB will be a forum for synchronization, coordination, and project management.	282	8888282	77
22022	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	The PCA will work with the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics and the Joint Staff to build an intradepartmental team of DoD	276	8888276	77
22023	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	ROADMAP FOR GOVERNANCE	A senior executive forum.	277	8888277	77
22024	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	Subordinate and reporting to the CIMB, a senior executive forum will provide initial senior-level coordination on key cyber issues.	278	8888278	77
22025	UnitedStates_2015_Final_2(	UnitedStates	8B CAPACITY BUILDING	MANPOWER DEVELOPMENT	The senior executive forum will recommend courses of action to the CIMB and will coordinate with other OSD and Joint Staff governance bodies to facilitate unity of effort a	279	8888279	77
22026	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	Improve cyber budgetary management.	280	8888280	77
22027	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	DoD will develop an agreed-upon method to more transparently and effectively manage the DoD cyber operations budget.	281	8888281	77
22028	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	Today cyber funding is spread across the DoD budget, to include the Military Intelligence Program (MIP), in multiple appropriations, budget lines, program elements, and p	282	8888282	77
22029	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	In addition, the Under Secretary of Defense for Intelligence, on behalf of DoD, ensures that all National Intelligence Program (NIP) investments are aligned to support DoD	283	8888283	77
22030	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	The diffuse nature of the DoD cyber budget presents DoD with a challenge for effective budgetary management; DoD must develop a new method for managing cross-prog	284	8888284	77
22031	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	Consistent with Presidential guidance, DoD will align and simplify its cyber operations and cybersecurity policy management and identified gaps, overlaps, seams, conflicts,	285	8888285	77
22032	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	This effort will help translate national and departmental guidance and policy into tactical operations.	286	8888286	77
22033	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	It is essential to clarifying conflicts in existing documentation that currently complicate cyber operations and cybersecurity governance.	287	8888287	77
22034	UnitedStates_2015_Final_2(	UnitedStates	8B CAPACITY BUILDING	MANPOWER DEVELOPMENT	Sailors conduct an exercise at Fleet Cyber Command's headquarters in the Frank B. Rowlett Building, Fort George G. Meade, MD.	288	8888288	77
22035	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	This exercise features members of Fleet Cyber Command's Joint Force Headquarters-Cyber (JFHQ-C).	289	8888289	77
22036	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	U.S. Cyber Command will lead a comprehensive operational assessment of its posture.	290	8888290	77
22037	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	Defense regarding organizational structure, command and control mechanism, rules of engagement, personnel, capabilities, tools, and potential operational gaps.	217	8888217	77
22038	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	The goal of this posture assessment will be to provide a clear understanding of the future operational environment; key stakeholder views; as well as strategic priorities, cho	218	8888218	77
22039	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	We live in a time of growing cyber threats to U.S. interests.	239	8888239	77
22040	UnitedStates_2015_Final_2(	UnitedStates	8B TECHNICAL MEASURES	CIRT	We are vulnerable in cyberspace, and the scale of the cyber threat requires urgent action by leaders and organizations across the government and the private sector.	240	8888240	77
22041	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	NATIONAL BENCHMARKING	Since developing its first cyber strategy in , the Defense Department has made significant progress in building its cyber capabilities, developing its organizations and plans, a	241	8888241	77
22042	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	Stemming from the goals and objectives outlined in this strategy, appropriate resources must be aligned and managed to ensure progress.	242	8888242	77
22043	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	This strategy presents an aggressive, specific plan for achieving change.	243	8888243	77
22044	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	For DoD to succeed in its mission of defending the United States and its interests in cyberspace, leaders from across the Department must take action to achieve the objecti	244	8888244	77
22045	UnitedStates_2015_Final_2(	UnitedStates	8B CAPACITY BUILDING	PROFESSIONAL CERTIFICATION	They must also hold their organizations accountable.	245	8888245	77
22046	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	Because of the nature of networks and computer code, no single organization can be relied upon to do this work.	246	8888246	77
22047	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	RESPONSIBLE AGENCY	Success requires close collaboration across DoD, between agencies of the U.S. government, with the private sector, and with U.S. allies and partners.	247	8888247	77
22048	UnitedStates_2015_Final_2(	UnitedStates	8B ORGANIZATION MEASURES	POLICY	The strategic environment can change quickly.	248	8888248	77
22049	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	That is especially true in cyberspace.	249	8888249	77
22050	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	We must be dynamic, flexible, and agile in this work.	250	8888250	77
22051	UnitedStates_2015_Final_2(	UnitedStates	8B TECHNICAL MEASURES	CIRT	We must anticipate emerging threats, identify new capabilities to build, and determine how to enhance our partnerships and planning.	251	8888251	77
22052	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	As always, our women and men Â¨C both uniformed and civilian personnel Â¨C will be our greatest and most enduring strength and a constant source of inspiration.	252	8888252	77
22053	UnitedStates_2015_Final_2(	UnitedStates	8B category unknown	category unknown	By working together we will help protect and defend the United States and its interests in the digital age.	253	8888253	77

Figure 11. Dataframe of all information

b. Sentences of selected Category

```
Selected Country:   UnitedStates
Selected Category:  TECHNICAL MEASURES
Number of Sentences:  291
```

	SentenceID	Sentence	sub category
20166	8686561	The way business is transacted,government operates,and national defense is conducted have changed.	CIRT
20167	8686562	These activities now rely on an interdependent network of information technology infrastructures called cyberspace.	CIRT
20168	8686563	TheNational Strategy to Secure Cyberspaceprovides a framework for protecting this infrastructure that is essential to our economy,security,and way of life.	CIRT
20172	8686630	Priority II: A National Cyberspace Security Threat and Vulnerability Reduction Program .	CIRT
20173	8686631	Priority III: A National Cyberspace Security Awareness and Training Program .	CIRT
20177	8686602	Cyberspace is composed of hundreds of thousands of interconnected computers,servers, routers,switches,and fiber optic cables that allow our critical infrastructures to work.	CIRT
20178	8686603	Thus, the healthy functioning of cyberspace is essential to our economy and our national security.	CIRT
20186	8686589	Minimize damage and recovery time from cyber attacks that do occur.	CIRT
20199	8686624	Public-private engagement is a key component of our Strategy to secure cyberspace.	CIRT
20200	8686625	This is true for several reasons.	CIRT
20201	8686626	Public-private partner- ships can usefully confront coordination problems.	CIRT
20202	8686627	They can significantly enhance information exchange and cooperation.	CIRT

Figure 12. Sentences of selected category

c. Sentences of selected Sub-category

```
Selected Country:   UnitedStates
Selected Category:  TECHNICAL MEASURES
Number of Sentences:  5
```

	SentenceID	Sentence	sub category
20632	8686128	processes and procedures that diminish the possibil- ities oferroneous code,malicious code,or trap doors that could be introduced during development.	STANDARDS
20686	8686084	Colleges and universities are encouraged to secure their cyber systems by establishing some or all ofthe following as appropriate:( ) one or more ISACs to deal with cyber attacks and vulnerabilities;( ) model guidelines empowering ChiefInformation Officers (CIOs) to address cybersecurity;( ) one or more sets ofbest practices for IT security;and,( ) model user awareness programs and materials.	STANDARDS
20856	8686054	providers of information technology products and services,and other organizations can make it easier for home users and small businesses to secure their systems.	STANDARDS
21412	8787176	As such, DHS is examining and comparing different approaches and will seek industry consensus on approaches to be brought forward for consideration by standards organizations.	STANDARDS
21625	8787085	Following on this research, international packet switching network standards were developed in collaboration with entities in other countries under the auspices of the ITU.	STANDARDS

Figure 13. Sentences of selected sub-category

Appendix

1. Category percentile

CAPACITY BUILDING	6.27%
AGENCY CERTIFICATION	0.10%
MANPOWER DEVELOPMENT	3.42%
PROFESSIONAL CERTIFICATION	1.41%
STANDARDISATION DEVELOPMENT	1.34%
CATEGORY UNKNOWN	41.30%
CATEGORY UNKNOWN	41.30%
CHILD ONLINE PROTECTION	4.43%
INSTITUTIONAL SUPPORT	0.38%
NATIONAL LEGISLATION	1.28%
REPORTING MECHANISM	2.48%
UN CONVENTION AND PROTOCOL	0.29%
COOPERATION	2.92%
INTRA-AGENCY COOPERATION	0.21%
INTRA-STATE COOPERATION	0.07%
PUBLIC SECTOR PARTNERSHIP	2.64%
LEGAL MEASURES	14.55%
CRIMINAL LEGISLATION	11.46%
REGULATION AND COMPLIANCE	3.09%
ORGANIZATION MEASURES	17.75%
NATIONAL BENCHMARKING	5.80%
POLICY	3.60%
RESPONSIBLE AGENCY	1.80%
ROADMAP FOR GOVERNANCE	6.55%
TECHNICAL MEASURES	12.77%
CERTIFICATION	0.74%
CIRT	11.54%
STANDARDS	0.49%