

# Multinational Military Exercises: Classifications and Insights

Kaylee Jensen, Tanmay Karandikar, Niko Laohoo, Jaemin Lee, Cassandra Lizardi, Dat Tran



## Abstract

Multinational military exercises (MMEs), commonly referred to as war games, are cooperative, non-combative actions undertaken by the militaries of multiple states intended to improve future military cooperation. Different types of MMEs serve different purposes; some are meant to signal a commitment or a willingness to get involved in a conflict or to deter aggression, while others are held as a coordination exercise among members of the same international institution. In this research we identify three types of exercises: warfighting, humanitarian, and peacekeeping. We then ask two research questions: (1) how can we use machine learning to classify news articles into the three MME types? (2) What can we learn about international relations from exploring MMEs over the 1980-2010 time period? To answer these questions, we structured and labeled a database of MMEs and the news stories describing them. We then used supervised machine learning to classify stories, achieving accuracies from 88 to 93.6%, and show that the multilayer perceptron outperforms the support vector machine algorithm. Substantively, MMEs involving the United States have increased from 1980-2010, and that the bulk of these exercises are for warfighting purposes. Humanitarian and peacekeeping exercises were essentially non-existent until the end of the Cold War. Finally, we only identified one exercise involving China from 1980-2001, but found 41 between 2002 and 2010. China's increase in exercises also coincides with increases by the US and Russia, although the three also participate in exercises together.

## Research Questions

### Initial core question:

- How can we construct a database that can be used for text analytics, substantive insights, and replication simultaneously?

### Text analytics:

- Given a set of text documents, how can we structure the text into data to be used in machine learning models?
- How can machine learning be used to classify articles about MMEs into warfighting, peacekeeping, and humanitarian?
- How do machine learning algorithms compare to one another for document classification purposes?

### Substantive insights for international relations:

- How have American MMEs changed over time and why?
- Why have humanitarian and peacekeeping exercises expanded since the end of the Cold War?
- How have MMEs between the major powers, specifically the US, Russia, and China changed over time? Why have they changed?

## Future Research Questions

### Text analytics:

- Will other feature selection methods improve performance?
- How might multi-classification improve performance?

### Substantive insights for international relations:

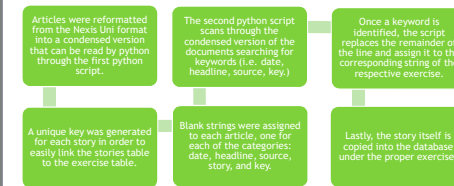
- What is the frequency of Indian MMEs with major powers?
- As all are members of NATO, how have US, French, and British MMEs been used to advance NATO interests?

## Preliminaries: Text Formatting

Articles about MMEs were originally obtained through the Nexis Uni database. The initial documents and MME data were provided by Dr. D'Orazio. To consolidate the required data from the differently-organized text files, these documents were processed using two python scripts. The first reformats all the text files into a consistent structure. The second extracts all metadata fields present in the document. The second script locates rows, identifies specific details, and copies only the pertinent information into an efficient format. Approximately 13,000 stories concerning exercises from the 1970s to 2010 were processed.

Two separate tables were formed:

- The stories table contains core points of each of the MMEs (the key, headline, author, date of publication, and the text itself)
- The exercise table contains the salient information extracted from the news stories (participants, exercise time frame, location, etc.).



## Preliminaries: Data Labeling

	WF	HU	PK	WF/HU	WF/PK	PK/HU	Unknown	Total
1980 - 1990	232	5		(-)3			5	234
1991 - 2001	222	40	44	(-)6	(-)1	(-)6	4	297
2002 - 2010	212	22	19	(-)3	(-)2	(-)2	117	363
Total	666	67	63	12	3	8	128	894

\* Note: The decline in Warfighting exercises between 2002-2010 is a result of the amount of data labeled "unknown" for that time period rather than an actual decrease. Table pertains to American MMEs.

**Warfighting:** Exercise is intended to improve states' ability to fight together in future combat, maintain individual defense capabilities, and deter any possible threats.

Example: MME #58 is a NATO-led warfighting exercise to "...test and enhance the ability of the naval units to respond to the situations likely to develop in sea battles involving the use of conventional weapons..."

**Humanitarian:** Exercises are intended to improve states' ability to coordinate for non-combat, humanitarian purposes in the future. Unlike warfighting and peacekeeping, these exercises may involve activities such as provision of medical aid.

Example: MME #1394 in the data set is an exercise with the purpose of helping build bridges and schools in the impoverished region of the Southern Philippines.

**Peacekeeping:** Exercises are intended to improve states' ability to jointly monitor or enforce peace in a zone with an ongoing or a recent conflict. This includes related peacekeeping activities such as the establishment and security of refugee camps.

Example: MME #1402 is a peacekeeping exercise with forces from the US, Mongolia, South Korea and Indonesia, in which the purpose was to improve the countries' ability to operate together in UN-sponsored peace-enforcement operations.

## Text Analytics

**Aim:** Design an algorithm to predict the exercise type (warfighting, peacekeeping, or humanitarian) with only the article given as input.

### Data Formatting:

- Data was prepped by consolidating only the keys, articles, and exercise types.
- Punctuation and special characters were removed from the articles. Stop words (i.e. a, and, or, what, etc.) were removed using the Natural Language Toolkit (NLTK).
- Text was vectorized to allow the algorithm to use it as input, through Python's Scikit-Learn package, "CountVectorizer," which converts text documents to a matrix of token counts.
- 70% of the entries were randomly selected and used for training data, leaving the remaining 30% as test data.

### Methodology:

- Two Natural-Language Processing (NLP) machine learning models were implemented: Multi-Layer Perception (MLP), and Support Vector Machine (SVM).
- The models classified documents into separate categories: Warfighting, Peacekeeping, and Humanitarian.
- Each model was paired with the three different categories separately and categorized the exercises as warfighting and not warfighting for example.
- Resulting in six sub-models (2 models and 3 scenarios for each)
- A confusion matrix was constructed for each of the six sub-models. A confusion matrix describes our models' performance of classification and mis-classification respectively for the different categories. This matrix provides a visualization of our classification, and the quality of the models' performance. The diagonal entries describes the number of the correctly identified exercises with the off-diagonal entries representing the exercises that were incorrectly identified.

### Results:

**SVM-Warfighting**  
Accuracy: 88.0%

	Actual no	Actual yes
Predicted no	28	34
Predicted yes	20	368

**SVM-Humanitarian**  
Accuracy: 90.7%

	Actual no	Actual yes
Predicted no	401	3
Predicted yes	39	7

**SVM-Peacekeeping**  
Accuracy: 93.1%

	Actual no	Actual yes
Predicted no	405	11
Predicted yes	20	14

**MLP-Warfighting**  
Accuracy: 88.4%

	Actual no	Actual yes
Predicted no	18	44
Predicted yes	8	380

**MLP-Humanitarian**  
Accuracy: 90.2%

	Actual no	Actual yes
Predicted no	404	0
Predicted yes	44	2

**MLP-Peacekeeping**  
Accuracy: 93.6%

	Actual no	Actual yes
Predicted no	413	3
Predicted yes	26	8

- The accuracy of all of the sub-models are higher than the Naive-Bayes classifier for all warfighting (the modal category). This classifier demonstrated 82.01% for the classified stories (baseline model for comparison).

## Substantive Insights

### USA MME Types:

- The total number of MMEs steadily increase through all three time periods with Warfighting exercises constituting the majority at 666 out of 894 total USA MMEs.
- Peacekeeping MMEs were not conducted by the USA until the post-Cold War periods.
- The largest number of Humanitarian oriented MMEs conducted by the USA occur during the "most peaceful" of the three time periods: post-Cold War and pre-9/11 War Effort.

### Why Humanitarian?

MME #427, *Balikatan*, is a 1991 exercise involving the US and the Philippines. The exercise "...concentrated on helping victims of last June's Mount Pinatubo volcanic eruption"

- This transition from military exercises having the reserved purpose of improving warfighting capabilities to also serving humanitarian purposes started in 1991. This exercise is significant because it appears to have acted as a catalyst for militaries around the world adding an additional role of humanitarian aid and peacekeeping to combat exercises.
- #427 is highlighted because it precedes a significant increase in the humanitarian role/exercises for militaries after its occurrence.

MME #503 represents the first joint operation between U.S. and Russia since WWII. The purely humanitarian nature of the exercise indicates that Humanitarian exercises, exclusive or in addition, act as symbols of good faith or at least an end in tensions.

### Rival Comparisons:

To explore shifts in system polarity, the total number of MMEs conducted by Russia and China, two other major powers in these time periods, were counted.

	RUS	CHN
1980-1990	40	
1991-2001	55	1
2002-2010	94	41

- US outnumbered Russian MMEs during the Cold War period 29:5
- Neither rival has a total number of MMEs comparable to the US, but the dramatic increase in China MMEs between 1991-2001 and 2002-2010 presents an interesting possibility for inquiry.

### Cooperation Comparisons:

- 1980-1990: No combinations of the three
- 1991-2001: US/RUS exercises (10 exercises); CHN/RUS/USA (1 exercise) pre-Taiwan tensions
- 2002-2010: US/RUS (13 exercises); RUS/CHN (4 exercises); US/CHN (2 exercises)

## References

The coding language that was employed was Python. We uploaded our code to a GitHub repository to maintain a record of it and for future reference. The Pandas software library was utilized to clean and structure our spreadsheet data. SKLearn was employed to run machine learning algorithms and analyze our text-based stories. Google Collaboratory was employed to allow for sharing and editing of code between members. Excel commands and functions were employed to check for maintained data accuracy across different subsets of the database.

This research project was under the faculty supervision of Vito D'Orazio. Following UTD Match Day 2019, research took place during the Spring 2020 semester.