You’re Not Alone in Battle: Combat Threat Analysis Using Attention Networks and a New Open Benchmark

Soo Yong Lee†
KAIST
Seoul, Republic of Korea
syleetolow@kaist.ac.kr

Juwon Kim†
KAIST
Seoul, Republic of Korea
a80908@kaist.ac.kr

Kiwoong Park
Agency for Defense Development
Seoul, Republic of Korea
kwpark@add.re.kr

Dong Kuk Ryu
Agency for Defense Development
Seoul, Republic of Korea
dkryu@add.re.kr

Sangheun Shim‡
Agency for Defense Development
Seoul, Republic of Korea
ssheun7@add.re.kr

Kijung Shin‡
KAIST
Seoul, Republic of Korea
kijungs@kaist.ac.kr

ABSTRACT
For military commands, combat threat analysis is crucial in predicting future outcomes and informing consequent decisions. Its primary objectives include determining the intention and attack likelihood of the hostiles. The complex, dynamic, and noisy nature of combat, however, presents significant challenges in its analysis. The prior research has been limited in accounting for such characteristics, assuming independence of each entity, no unobserved tactics, and clean combat data. As such, we present spatio-temporal attention for threat analysis (SAFETY) to encode complex interactions that arise within combat. We test the model performance for unobserved tactics and with various perturbations. To do so, we also present the first open-source benchmark for combat threat analysis with two downstream tasks of predicting entity intention and attack probability. Our experiments show that SAFETY achieves a significant improvement in model performance, with enhancements of up to 13% in intention prediction and 7% in attack prediction compared to the strongest competitor, even when confronted with noisy or missing data. This result highlights the importance of encoding dynamic interactions among entities for combat threat analysis. Our codes and dataset are available at https://github.com/syleeheal/SAFETY.

CCS CONCEPTS
• Information systems → Data mining; • Computing methodologies → Machine learning.

KEYWORDS
Attention Networks; Combat Threat Analysis; Intention Prediction; Attack Prediction; Open Benchmark

†Co-first authors.
‡Co-corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:

1 INTRODUCTION
Combat threat analysis (CTA) is a critical factor for military decision-making [6]. By accurately predicting the intention and attack probability of the hostiles, friendly troops can take action to minimize the damage. Despite its importance, the complex nature of combat poses significant challenges in its analysis. Specifically, combat is interactive (entities interact with each other to make their decisions), flexible (troops can adopt new tactics that have not been observed), and noisy (combat data is inaccurate and incomplete due to sensor or communication system limitations). Accounting for these characteristics is essential for CTA.

However, prior research has been limited in accounting for such properties [1, 4, 5, 8, 16–18, 21–23, 25, 26]. For instance, their models analyze each entity independently, assuming that its intention or attack probability does not depend on the other entities. Their dataset generally are complete and noise-free, with no explicit hostile tactic defined. These assumptions, however, do not likely hold in real-world combat data, casting doubt upon their feasibility.

To tackle each challenge, we propose a novel spatio-temporal attention architecture to encode interactions among entities (the interactive property). The proposed model’s prediction, thereby, depends on specific interactions each entity has with others. Also, we test the model’s predictive capacity for unobserved tactics (the flexible property) and robustness to perturbations that reflect real-world combat data (the noisy property).

Finally, to our best knowledge, we provide the first open-source benchmark for CTA. The benchmark dataset is comprised of 1238 simulations of ground force combat, each of which is assigned one of four hostile attack tactics. The benchmark tasks are to predict (a) the intention of each squad and (b) attack probability between each squad pair within each combat simulation.

In our benchmark, our proposed model significantly outperforms the baselines and show robustness to various types of perturbations. Our main contributions are three-fold:
• Dataset: The first open-source benchmark dataset for CTA.
• Problem Formulation: A realistic formulation of CTA.
• Method: A novel spatio-temporal attention architecture for CTA, with strong performance and robustness.

2 PRELIMINARIES AND RELATED WORKS

In this section, we introduce the concepts and review the related literature. Refer to Table 1 for frequently used notations.

2.1 Concepts

Definition 2.1 (Entity, Squad, and Combat). Entities, squads, and combat describe how a battle unravels over time. An entity \( e \) refers to the smallest force unit within a combat, such as a soldier. We denote each \( i^{th} \) entity by \( e_i \), and we denote the feature vector of \( e_i \) at time \( t \) by \( E^t_i \in \mathbb{R}^d_0 \), where \( d_0 \) is the feature dimension. A squad \( S \) refers to a set of few entities. The entities within a squad are under the same force and share the same intention. The set \( S_j \) denotes the set of the entities that belong to the \( j^{th} \) squad. Combat refers to all the entities that are involved in a battle over time. We represent each \( c^{th} \) combat as a tensor \( T_c \in \mathbb{R}^{I_{max} \times d_0 \times T_{max}} \), where \( I_{max} \) and \( T_{max} \) are the numbers of entities and timestamps, respectively. The matrix \( T^t_c \in \mathbb{R}^{I_{max} \times d_0} \) describes the combat snapshot at time \( t \). That is, a combat \( T_c \) consists of the set of all entity features \( E^t_i \) at timestamp \( t \) within the same combat.

Definition 2.2 (Intention and Attack). Intention and attack probability are the targets of CTA. An intention \( Y_{int}(S_j; c) \in \mathbb{R}^n \) refers to the intended action in the combat. Each squad is assigned a \( Y_{int} \) that determines their course of action throughout the combat. Meanwhile, an attack \( Y_{atk}(S_j, S_j'; c) \in \{0, 1\} \) is assigned to each pair of squads. It indicates whether an attack between the two squads will occur or not after the observed combat.

Definition 2.3 (Tactic). A tactic refers to the overarching strategy that the hostile entities share in each combat \( T_c \), influencing their intention and attack probability (See Figure 1).

2.2 Related Works

Models for CTA. Fuzzy logic- and Bayesian networks-based algorithms have been developed for CTA [1, 4, 5, 8, 10, 21, 22]. They predict the intention or attack probability of a hostile given the entity feature and expert-given predetermined rules. Their heavy reliance on predetermined, crafted rules renders them vulnerable to changing hostile tactics and infeasible to analyze high-dimensional features or many entities.

As such, neural network-based CTA methods have been developed. Most of them use RNN variants to handle temporal features of combat [16–18, 23, 26]. By learning to analyze the combat threats in a data-driven manner, the methods may handle many variables and learn complicated rules for generalization to unobserved tactics. Despite their advantages, the prior studies consider each entity independently. That is, no prior methods have considered interactions among multiple entities within combat.

Datasets for CTA. All prior research has used synthetic datasets for their evaluation, and none were made public [1, 4, 5, 8, 16–18, 21–23, 25, 26]. Within each combat, only one hostile entity appears for its prediction. That is, the datasets do not account for entity interactions within combat (the interactive property). No research defines explicit tactics adopted by the hostiles (the flexible property). Also, they generally utilize unperturbed features, collected at a regular time interval (the noisy property).

Spatio-temporal Prediction. Recently, many models use graphical structure or attention mechanism to learn spatio-temporal patterns and tackle related problems [11–14, 24, 27, 28]. Their applications include sequential recommendation [27], query-POI (point-of-interest) matching [24], traffic forecast [11, 12, 14, 28], disease spread forecast [13], etc. However, to our best knowledge, no spatio-temporal model has been applied for CTA.

3 PROPOSED BENCHMARK

In this section, we detail our proposed benchmark dataset and tasks.

3.1 Proposed Benchmark Dataset

Here, we describe the proposed benchmark dataset for CTA. For further details, please refer to the online appendix [9].

Dataset Structure. It is a synthetic dataset based on computer simulations of ground force combats. It contains a total of 1238 combat simulations \( T_c \)’s, each with one of four tactics. Each combat \( T_c \) has 12 squads \( S_j \)’s, and each squad \( S_j \) has 5-6 entities \( e_i \)’s and an intention label \( Y_{int} \). Every pair of \( S_j \)’s within the same combat \( T_c \) has an attack label \( Y_{atk} \). Each entity has 11-dimensional features \( E^t_i \) collected at each second. Each combat \( T_c \) lasts about 1, 400 seconds on average with a standard deviation of about 230 seconds.

Dataset Semantics. The tactics, which are shared by all hostile entities within the same combat \( T_c \), include (1) Linear Advance- ment, (2) Sequential Progression, (3) Flanking Maneuver, and (4) Direct Engagement. The squad intention labels \( Y_{int} \) include (1) Tactical Encirclement, (2) Maneuvering Techniques, (3) Coordinated Rendezvous, (4) Strategic Surprise, (5) Forceful Engagement, and (6) Strategic Positioning. The attack label \( Y_{atk} \) indicates whether an attack occurs between a squad pair by the end of each combat. The entity features \( E^t_i \) contains its information about (1) Coordinates, (2) Attitude (3) Speed, (4) Force Identifier, and (5) Located Terrain Type.

Dataset Quality. The simulations are carefully crafted based on expert military knowledge, ensuring the realism of the combat situations represented. For data quality control, we further conduct statistical analyses of data distribution [9].

Dataset Visualization. We visualize the features for each tactic in Figure 1. Clearly, the hostile entities under different tactics have distinct trajectories. The t-SNE [19] also shows that the entity features are well separated by tactics.
3.2 Proposed Benchmark Task
Given combat, we aim to predict the intention $Y_{\text{int}}$ and attack probability $Y_{\text{atk}}$ among squads $S_j$’s. In real-world combat data, entity features are most likely perturbed (the noisy property). First, features are often collected sporadically, resulting in irregular time intervals. Second, unobserved features at a given timestamp $t$ lead to missing values. Lastly, features are likely to be contaminated with noise due to limitations in sensor and communication systems. Therefore, we train and test models on the dataset collected at irregular time intervals, with either noisy or missing features.

Also, the hostilies can always adopt new, unobserved tactics to counterfeiT their intention and attack probability (the flexible property). Therefore, we test the model performance on the combats $T_c$’s with unobserved (untrained) tactics.

Please refer to Section 5.1 for details on how we implement this problem formulation.

4 PROPOSED MODEL: SAFETY
In this section, we introduce our proposed model, Spatio-temporal Attention For the EsTimation anAlysis (SAFETY), comprised of three main components: a spatio-temporal attention network; squad aggregation; a classifier. The attention network encodes spatio-temporal interactions among entities within combat via attention mechanism. Then, the entity features within each squad are mean aggregated to obtain squad features. Finally, a classifier layer outputs probability distributions over intention $\hat{Y}_{\text{int}}$ and/or attack $\hat{Y}_{\text{atk}}$ (See Figure 2).

4.1 Spatial Attention.
First, the input features are augmented with positional encoding [20]. To encode spatial interactions, we adopt self-attention mechanism [15, 20]. Specifically, we update entity features with spatial attention to other entity $i'$ within the same combat $T_c$ by

$$a_{ij}^{t} = \text{softmax}\left(\frac{Q_i^{t} K_i}{\sqrt{d}}\right), \quad H_i^{t} = \sigma(S_i^{t} + \sum_{j} a_{ij}^{t} V_i^{t}),$$

where $a_{ij}^{t}$ is an attention coefficient between entities $i$ and $i'$ at timestamp $t$, $H_i^{t} \in \mathbb{R}^{d}$ is the entity $i$’s updated features at timestamp $t$, $d$ is the hidden dimension, and $\sigma$ is activation function ELU [3].

In addition, $Q_i^{t} = E_i^{t} W_1$, $K_i^{t} = E_i^{t} W_2$, $S_i^{t} = E_i^{t} W_3$, and $V_i^{t} = E_i^{t} W_4$, where $W$s $\in \mathbb{R}^{d \times d}$ are the learnable parameters. That is, spatial interactions that each entity has at each timestamp are encoded as the aggregation of other entities’ features by their attention.

4.2 Temporal Attention.
The output features from the spatial attention network $H_i^{t}$ are fed into the temporal attention network. The same attention mechanism is applied to encode temporal interactions. However, for each entity $i$ at a given time $t$, it computes attention across different timestamps $t'$. Formally, we learn temporal attention with

$$\tilde{a}_{it}^{t'} = \text{softmax}\left(\frac{\tilde{Q}_i^{t} \tilde{K}_i^{t'}}{\sqrt{d}}\right), \quad \tilde{H}_i^{t} = \sigma(S_i^{t} + \sum_{t'} \tilde{a}_{it}^{t'} \tilde{V}_i^{t'}),$$

where $\tilde{a}_{it}^{t'}$ is an attention coefficient between timestamp $t$ and $t'$ of entity $i$. In addition, $\tilde{Q}_i^{t} = H_i^{t} \tilde{W}_1$, $\tilde{K}_i^{t'} = H_i^{t} \tilde{W}_2$, $\tilde{S}_i^{t} = H_i^{t} \tilde{W}_3$, and $\tilde{V}_i^{t'} = H_i^{t} \tilde{W}_4$, where $\tilde{W}$s $\in \mathbb{R}^{d \times d}$ are the learnable parameters.

4.3 Squad Aggregation and Prediction.
In order to predict the intention and/or attack probability of squads $S_j$’s, we need to aggregate the final entity features $H_j^{t}$’s. First, the entity features are mean aggregated across all timestamps $t$’s and for each squad $S_j$ to obtain the final squad features $Z_j$. Formally,

$$Z_j = \frac{1}{|S_j|} \sum_{e_i \in S_j} \frac{1}{|t|} \sum_{t \epsilon T} \tilde{H}_i^{t}.$$  

Thereby, the final squad features $Z_j$, a classifier layer computes the probability distributions for intention $\hat{Y}_{\text{int}}$ and/or attack $\hat{Y}_{\text{atk}}$. 

Figure 1: Data Visualization. (a) shows 3D trajectories of entities $e_i$ from combat $T_c$ assigned with each tactic. The red and blue represent hostile and friendly entities, respectively. (b) presents a t-SNE visualization of entity features $E_i^{t}$ from each tactic.

Figure 2: An overview of SAFETY. It takes combat $T_c$ as the input, computing the probability distributions for intention $Y_{\text{int}}$ and attack $Y_{\text{atk}}$. While the figure depicts two squads, SAFETY has the capability to handle any number of squads.
5 EXPERIMENTS

In this section, we delineate the experimental outcomes.

5.1 Experimental Settings

**Baselines.** We compare SAFETY’s performance against static and temporal prediction models. The static models are k-Nearest Neighbors (kNN), XG-Boost (XG) [2], and Multi-Layer Perceptron (MLP). The temporal models are LSTM [7] and state-of-the-art models in CTA, including BiGRU-Att [18] and PCLSTM [23].

**Realistic Perturbations.** We make the following perturbations to reflect the flexible and noisy properties of real-world combat data.

- **Irregular Time Intervals:** 20 random timestamp $t$’s are sampled from each combat $T_c$ as the model input.
- **Feature Noise:** For continuous feature values, noise sampled from a Gaussian distribution $N(0, \sigma^2)$ was added, where $\sigma^2$ is an experimental hyperparameter (See the x-axis in Figure 3(a)). For binary feature values, zeros were changed to ones and vice versa with probability $\sigma^2$.
- **Missing Features:** We remove (spec., mask corresponding features with 0s) (1) random features of random entities in random snapshots (Feature), (2) random entities at random snapshots (Entity), (3) random squads at random snapshots (Squad), or (4) random snapshots (Time), by a given probability (See the x-axis in Figure 4).
- **Test on Unseen Tactic:** At each trial, we use all combats $T_c$’s with 3 randomly sampled tactics as the train set. All the combats $T_c$’s with the remaining tactic serve as the test set.

**Details.** Further details can be found in the online appendix [9].

5.2 Model Performance

**CTA under Feature Noise.** In Figure 3(a), we present the model performance over increasing feature noise. In both attack and intention prediction, the performances of all models decline over increasing noise. However, SAFETY always maintains higher performance over the baselines, demonstrating its stronger resistance to noise. Specifically, when considering the presence of noise, SAFETY outperforms the second-best model, BiGRU-Att, by at most 6% in attack prediction and 8% in intention prediction (F1-Micro). In both predictions, we find statistical significance (t-test; $p<0.001$) in the performance difference between SAFETY and BiGRU-Att across multiple noise levels.

**CTA under Missing Features.** Figure 3(b) describes model performance over increasing masking probability for Feature. We further show model performance with different masking types, described above, but with a fixed masking probability of 50% in Figure 4. We show SAFETY outperforms all baselines for all masking probabilities and types, with a particularly large margin in intention prediction. Specifically, when accounting for missing data, SAFETY outperforms the second-best model, BiGRU-Att, by at most 7% and 13% in attack and intention prediction (F1-Micro), respectively. Again, in both predictions, we find statistical significance (t-test; $p<0.001$) in the performance difference between SAFETY and BiGRU-Att across multiple masking levels and types. Notably, methods that do not account for temporal dynamics (MLP, XG, and kNN) demonstrate considerable degradation in performance when confronted with missing data.

6 CONCLUSIONS

In this study, we address the limitations of the prior literature in CTA by addressing the central characteristics of combat data (interactive, flexible, and noisy). We propose a novel framework for CTA, with the first open-source benchmark dataset, realistic problem formulation, and a novel spatio-temporal prediction model, SAFETY. The strong performance of SAFETY highlights consideration of interactions among entities, along with temporal dynamics, is significant w.r.t. CTA. In conclusion, our benchmark and model represent a significant advancement in CTA, offering valuable insights for military decision-making and situational awareness.

**Acknowledgements.** This work was supported by Agency for Defense Development (ADD) grant funded by the Korea Government (No. UI2100072D, Technique Analysis and Model Prototyping for the Elements Identification of Enemy Behavior and Threat).
REFERENCES


