Visualizing the Evolution of Multi-agent Game-playing Behaviors

1 UNIVERSITY OF BAMBERG

2 UNIVERSITÄT DUISBURG ESSEN — Offen im Denken

Shivam Agarwal [1] @shivamlearning

Shahid Latif [1] @mrshahidlatif

Aristide Rothweiler [2]

Fabian Beck [1] @beck_fabian

✉ {shivam.agarwal | shahid.latif | fabian.beck}@uni-bamberg.de

https://s-agarwl.github.io/evolvingai

Abstract:

Analyzing the training evolution of AI agents in a multi-agent environment helps to understand changes in learned behaviors, as well as the sequence in which they are learned. We train an existing *Pommerman* team from scratch and, at regular intervals, let it battle against another top-performing team. We define thirteen game-specific behaviors and compute their occurrences in 600 matches. To investigate the evolution of these behaviors, we propose a visualization approach and showcase its usefulness in an application example.

Pommerman Game Environment:

1. In *Pommerman*, agents in two teams lay bombs to kill the enemies. Each team has two agents.
2. Three power-ups, hidden beneath wooden tiles, allows agents to:
   - kick bombs,
   - increase blast radius of the laid bombs, and
   - drop more bombs without waiting for the previous one to explode.
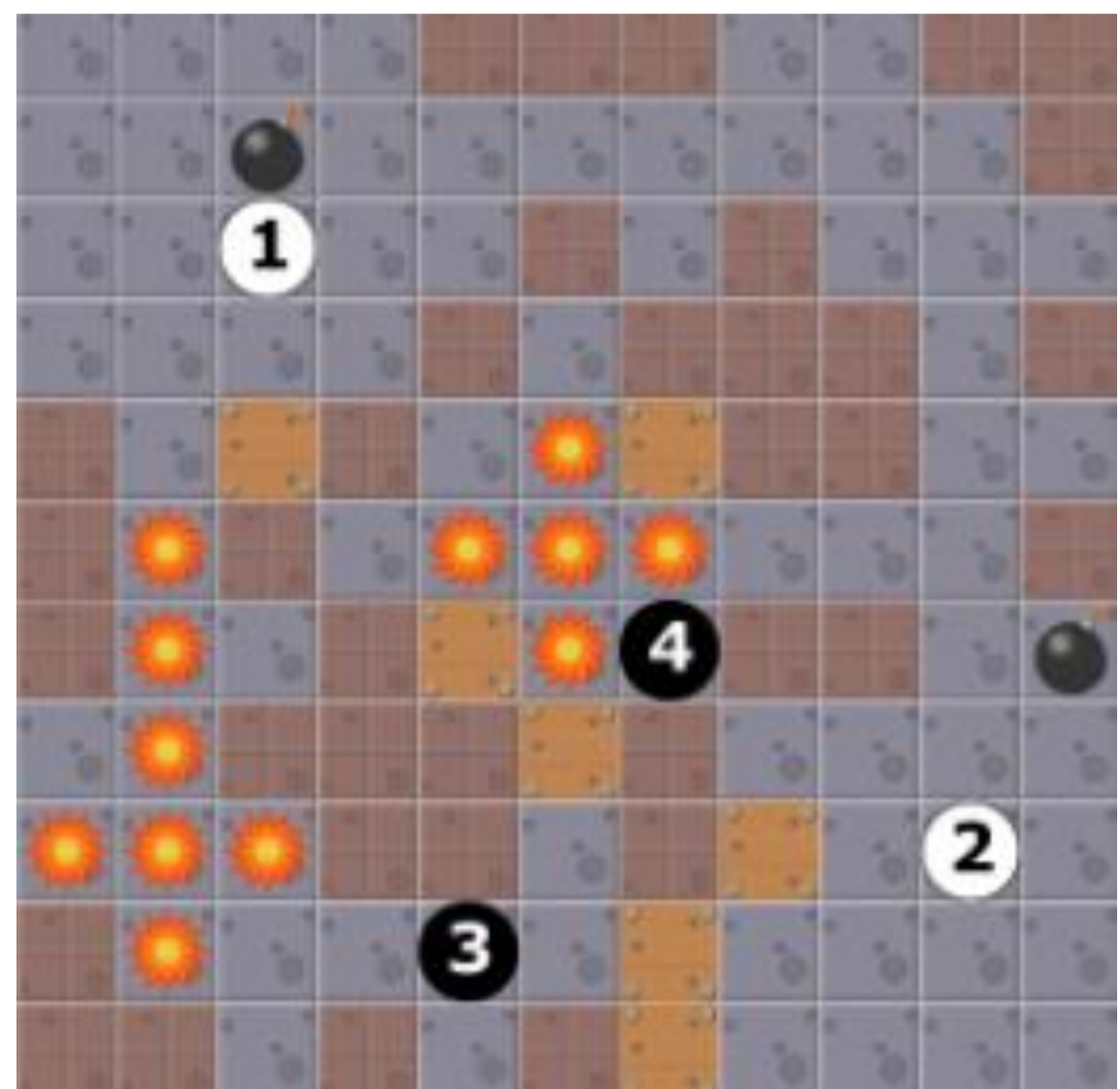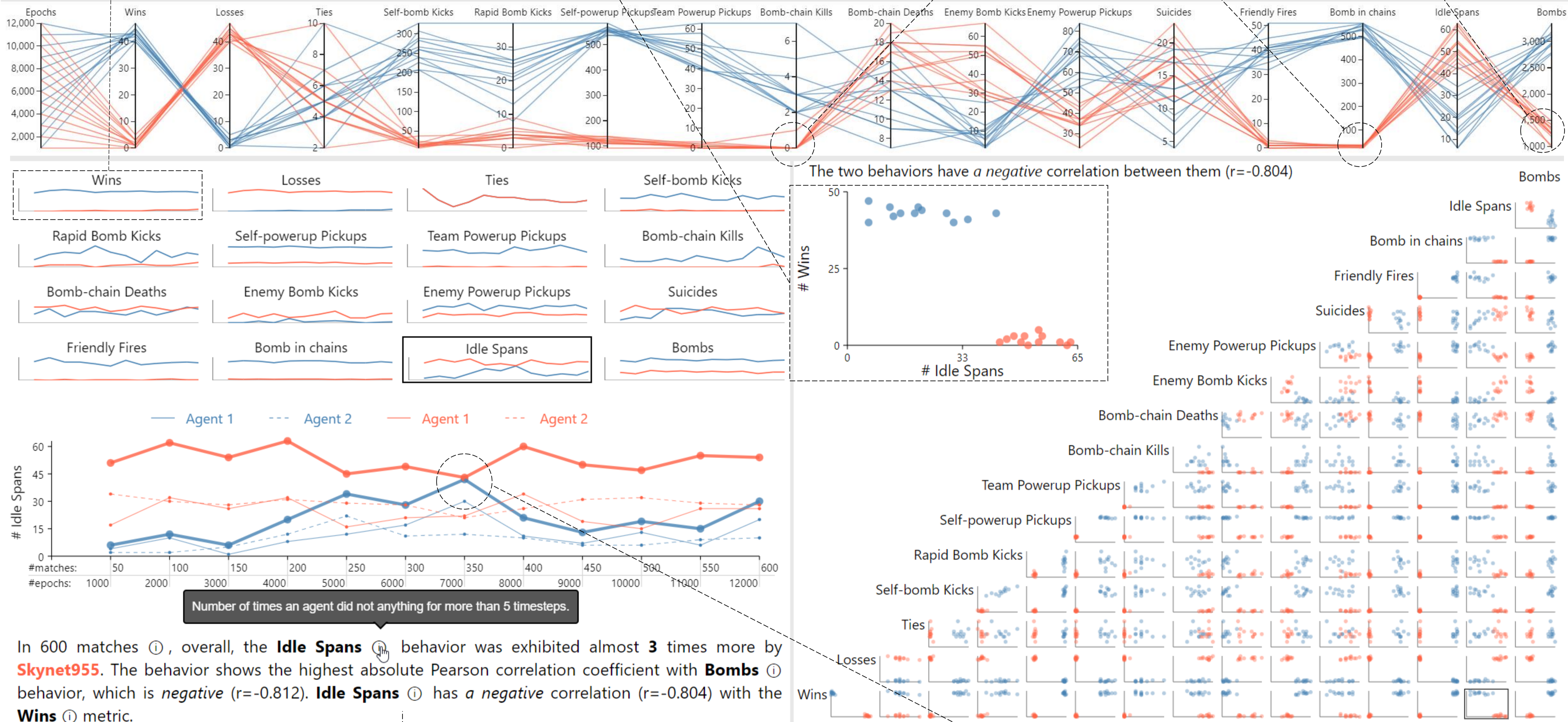
Data Collection:

1. We trained **S955 (Skynet955)** from scratch.
2. After every 1000 training epochs, we recorded 50 battles against **HJ (Hakozaki Junctions)**.
3. We defined 13 behaviors and quantified them from the logs.

> **HJ** has high number of **Wins** across the training of **S955**

> High number of **Idle Spans** does not seem to result in winning the battles

> Low frequencies of behaviors contributing to low performance (e.g., **Bomb-chain Kills**, **Bomb in Chains,** or laying **Bombs**)

Training **Skynet955** ⓘ team (two agents) to play against **Hakozaki Junctions** ⓘ (fully trained) in the **Pommerman** Game Environment ⓘ

The two behaviors have *a negative* correlation between them (r=-0.804)

In 600 matches ⓘ, overall, the **Idle Spans** ⓘ behavior was exhibited almost **3** times more by **Skynet955**. The behavior shows the highest absolute Pearson correlation coefficient with **Bombs** ⓘ behavior, which is *negative* (r=-0.812). **Idle Spans** ⓘ has *a negative* correlation (r=-0.804) with the **Wins** ⓘ metric.

Number of times an agent did not anything for more than 5 timesteps.

> **Idle Spans** and laying **Bombs** are negatively correlated. Hence, optimizing for one of them in the reward function might suffice

> **S955** learns to reduce the **Idle Spans** (epoch 7000), but their occurrence increases with further training

---

Behaviors:

Number of times an agent…

Behavior	Description	Behavior	Description
Self-bomb Kicks:	kicked his own bomb	**Rapid Bomb Kicks:**	kicked a bomb more than once in a short sequence
Self Powerup Pickups:	uncovered an item and then picked it up itself	**Team Powerup Pickups:**	picked up an item uncovered by an allied agent
Bomb-chain Kills:	killed others by using chained bombs	**Bomb-chain Deaths:**	got killed by bomb chains from other agents
Enemy Bomb Kicks:	kicked a bomb of an enemy agent	**Enemy Powerup Pickups:**	picked up an item uncovered by the enemy
Suicides:	got killed by a bomb placed by itself	**Friendly Fires:**	killed an ally with a bomb place by itself
Bomb in Chains:	laid bombs that were involved in a chain	**Idle Spans:**	did not do anything for more than 5 timesteps
Bombs:	laid a bomb		