Plagiarism Detection

Derek Awender and Thomas Dougherty

Taoshi Inc.

October 2024

1 Executive Summary

Time series forecasting involves developing models to predict future conditions, with diverse applications spanning from weather forecasting to analyzing financial markets. This field includes competitive platforms that assess users based on the accuracy of their prediction models, potentially incentivizing participation or performance. These platforms are at risk of plagiarism, which can disrupt the distribution of incentive, pollute the networks with redundant information, and introduce instability.

Plagiarism occurs when users acquire the predictions of other users and then submit direct or slightly altered copies of these signals. This can deprive honest users of incentive while providing no added value to the system. With lower diversity among strategies and the inclusion of redundant information, plagiarism becomes a detriment to the overall health and stability of these platforms. Identifying and eliminating plagiarists ensures time series platforms remain fair and promotes healthy competition.

Our Plagiarism Detection System (PDS) utilizes a multi-layered approach that considers similarities in temporal behavior along with the general structure of the time series to identify plagiarism on forecasting networks. We show a proof of value for plagiarism detection on SN8, opening the door to broader applications of plagiarism detection within Bittensor.

2 Introduction

Bittensor is at the intersection of machine learning and blockchain. Each subnetwork within Bittensor is meant to host a competitive ecosystem for new tasks within machine learning, where the quality of output or predictions determines the relative rewards that should be distributed to the miner. Combining the outputs from multiple miners leads to something potentially more valuable than any individual miner, something we term the **network model**.

The strength of the network model relies on the quality and diversity of user predictions. When each user employs a unique strategy, the network becomes more resilient to single points of failure. For instance, if all users rely on the same strategy and it falters, the final output of the network greatly suffers. This is a very real scenario, as the top miners in the ecosystem may have their strategies stripped and reused by other miners attempting to stay competitive. These other miners may then submit different versions of these predictions as their own. We term this theft as **plagiarism**. Plagiarism decreases the number of unique strategies, introducing possibly harmful redundancy.

3 Problem Statement

Any incentive-based time series prediction network is vulnerable to plagiarism among users. Users can acquire the predictions of other users via unauthorized access or, in some cases, can purchase them if they are offered at a low enough price. Once the plagiarist possesses the victim's predictions, they can directly use them or modify them to avoid suspicion of copying. However, it's essential to distinguish plagiarism from the legitimate practice of building on others' strategies, which fosters healthy competition and enhances the network's overall performance.

Depending on the scoring mechanism, users ranked similarly may receive similar amounts of incentive. Plagiarists could copy top-ranked users, which would force these high-performing users to share incentive and block incentive from reaching lower ranked yet still valuable users.

Overlapping signals due to plagiarism make for a weaker product by reducing the number of failing components before system-level collapse. Plagiarism reduces the uniqueness of information from each miner, making the overall network more dependent on a smaller set of users. When one user stops producing predictions, any plagiarists would also stop producing predictions. This redundancy skews the network's output, giving undue weight to repeated signals, which could diminish the competitive edge of the final network output.

In summary, plagiarism potentially disrupts the proper distribution of incentive and threatens the stability of the network by introducing redundant information. These challenges ultimately hinder the network's ability to achieve its objectives, regardless of the specific field, be it weather forecasting, financial market analysis, or any other domain involving time series data.

4 Steps Towards a Solution

To solve the issue of plagiarism, we propose PDS, which is a multi-layered system motivated by two main ideas as to what constitutes a plagiarism event against another miner:

1. Consistently similar leverage utilization
2. Consistently similar order times, with a pattern of following

We will refer to the predicted value as leverage where the sign reflects whether the asset is predicted to increase or decrease in value (positive and negative respectively), and the magnitude represents the level of confidence. Also, the time series as a whole will be referred to as positions and changes in the leverage (e.g., at time 2000) are orders. Finally, users may be referred to as miners. Figures like the one below can help visualize the similarities between time series with time on the x-axis and leverage on the y-axis.

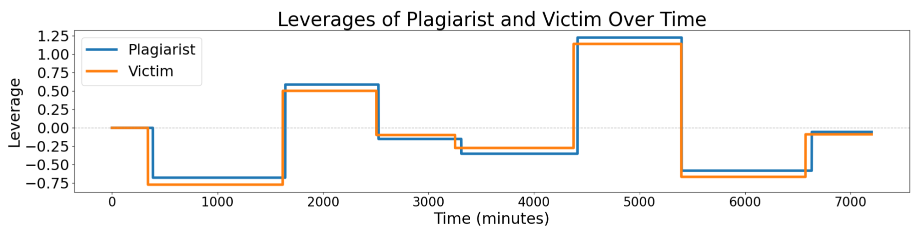

Figure 1: Visualizing Plagiarism

4.1 Simulation Environment

To track our progress and quality, we created a simulation environment for rigorous analysis and fine-tuning of plagiarism detection. This simulation randomly generates miner positions and then introduces tracked instances of plagiarism. By utilizing these cases of guaranteed plagiarism, we are able to evaluate system performance through a variety of error-tracking metrics and determine the trade-offs of each design decision.

4.2 Prototype

Before crafting a prototype, it was essential to establish a clear definition of plagiarism and identify an effective way to represent the positions of miners.

4.2.1 Structural Similarity

Analyzing time series for structural similarity involves focusing on broad trends of influence instead of scrutinizing every deviation. For example, short bursts of highly similar predictions may not indicate prolonged plagiarism. This lives in contrast to extended periods of similarity between miners, which may provide more concrete evidence for plagiarism.

Once plagiarists have acquired signals from other users, the signals could either be directly replicated or first manipulated in some way prior to implementation. The plagiarist may use the acquired signals in combination with other strategies in an attempt to create unique predictions that are more effective than the original copied strategy. They may also modify the signals, possibly with randomization, to avoid suspicion of plagiarism. As a reminder, the goal of PDS is to protect networks from instability, corrupted competition, and inequitable distributions of incentive.

Selecting and tuning the appropriate algorithm was driven by these goals of identifying pervasive patterns of similarity while minimizing the possibility of a false positive in plagiarism detection between miners.

4.2.2 Detection Algorithm

Initially, the positions of miners are represented as lists of these orders. Although compact, this representation isn't conducive to the use of similarity algorithms for time series data. Instead, we chose to represent the positions of miners as vectors of leverage, with each value representing the current leverage at a given time. Figure 1 demonstrates how the vector leverages might look when plotted. On the Y-axis, we see the leverage for the trade pairs while the X-axis represents time. Large jumps in the overall leverage values correspond to orders placed by the miner. With this representation, we were able to use a similarity algorithm to compare historical miner behaviors.

Iterating on the vector representation of positions and the similarity algorithm we chose, we introduced a method of time-shifting orders of plagiarist and victim pairs to determine whose orders are being followed. This time shifting is done by shifting one miner's signals forward in time (i.e., shifting a miner to the right in figure 1) and evaluating if this impacts the overall alignment of the orders.

This solution is simple, effective, and the heart of our current solution. However, it doesn't capture crucial insights such as information about if there is a clear history of following behavior.

4.3 Current Iteration

Building upon our analysis of structural similarity, we introduced a method of analyzing the original list of orders for similarity in order times.

4.3.1 Time Similarity

Plagiarists have to acquire the signals of other users, which will likely introduce a time lag between when the victim user submits their predictions and when plagiarists submit their predictions. While a suspicious order is likely to take place shortly after the victim's order, two things must be true for high confidence in systematic plagiarism:

1. A high percentage of the plagiarist's total orders must be suspicious.
2. A high percentage of the victim's orders must not be suspicious.

To have high confidence in culpability, the plagiarist should be consistently following the victim's orders. The victim, on the other hand, should seem to be generating unique signals. Without this distinction, frequent order submissions could be misinterpreted as plagiarism, even if both users are acting independently.

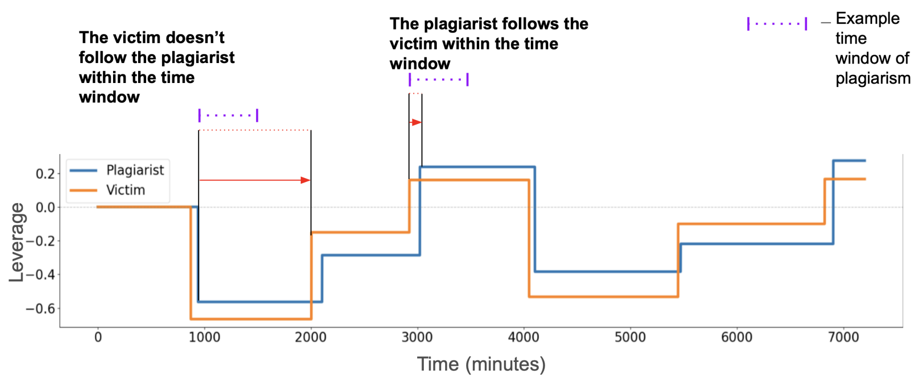

Figure 2: Time Analysis Diagram

4.4 State of Progress

Iterative improvements from the detection system are tracked against our simulation ecosystem in figure 3. This plot highlights the potential trade-off between false positive and true positive rates. *False positives* are innocent miners being detected as plagiarists and *true positives* are plagiarists being correctly identified.

Experimentation and tuning between iterations helped us refine our approach by incorporating techniques to detect anomalies in order timing and structural similarity. These improvements are tracked as measured progress in the final ROC curve from figure 3. Measuring quality as the area under each ROC curve (AUC), we compare each technique against a random control to evaluate the state of our progress.

The first iteration improves upon the control group by increasing AUC from 0.56 to 0.96. Further iteration on our second iteration sees a marked improvement to AUC and false positive rate, with our final model clocking in at an AUC value of 0.98.

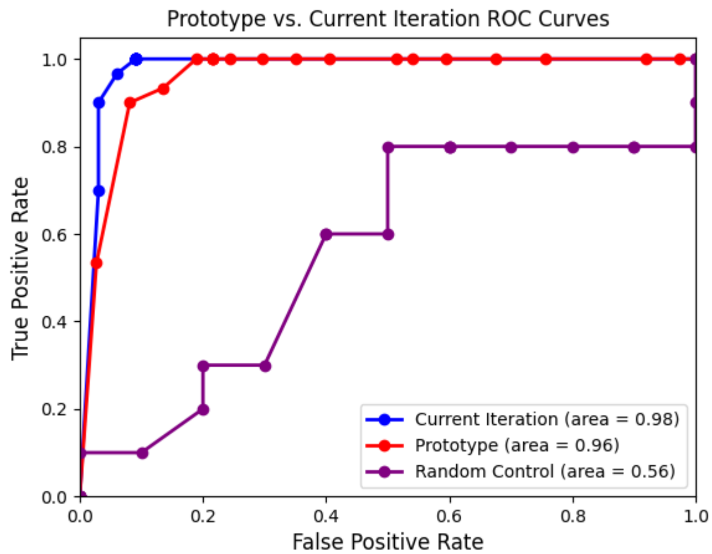

Figure 3: Comparison of Versions of PDS

5 Application to SN8

We will apply PDS to subnet 8 to help protect the signal integrity of our miners and ensure equitable distribution of incentive for each unique strategy.

5.1 Impact on Network Output

Each of the miner's positions acts as a weighted input, contributing to the overall network output. By setting larger weights for miners with higher performance, the network output generally remains balanced without being too dependent on one strategy. However, plagiarism increases the risk of the network output being overly reliant on one or a few strategies due to redundant information.

For example, consider a network with two miners, Miner A and Miner B. If Miner A contributes at a higher level than Miner B, Miner A will receive greater weight. This greater weight will shift the network output to more closely mirror Miner A's predictions. This is true even if Miner B is also contributing value, as long as the value contribution is lower than that from Miner A.

In a world without plagiarism detection, this may cause an over-reliance on the predictions from Miner A. For example, if a plagiarist enters the network and mimics Miner A's signals, the output would further skew towards Miner A. The figure below depicts the possible impact of plagiarism on the output of the network under this scenario.

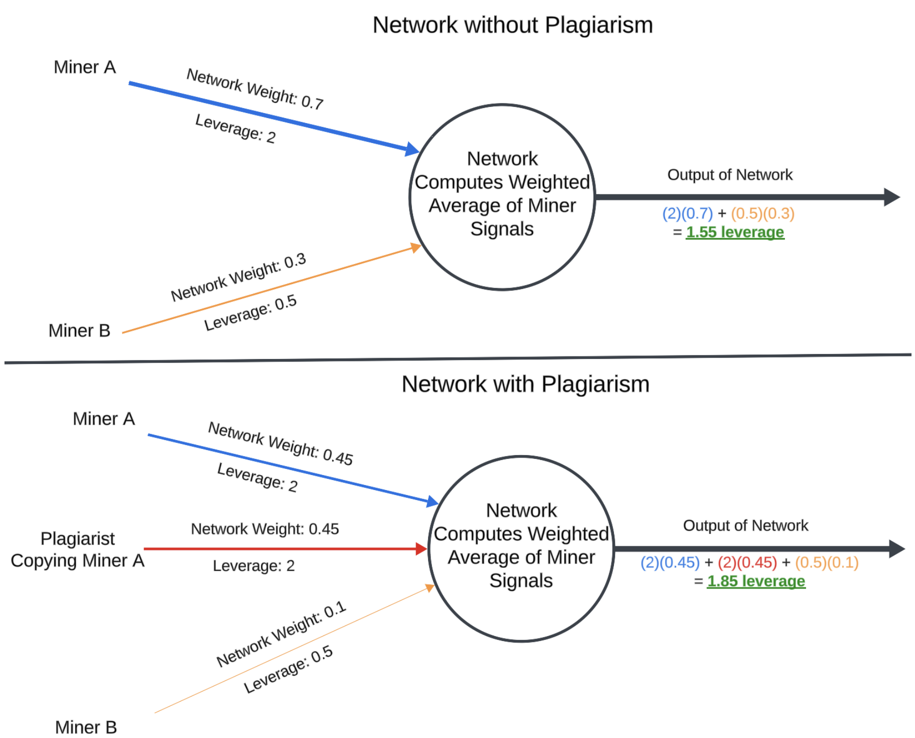

Figure 4: Network Model with and without Plagiarism

5.2 Impact on Distribution of Incentive

Generally, miners with comparable scores receive similar quantities of incentive, amplifying the potential harm caused by plagiarism. However, on SN8, miners must first pass a challenge period to demonstrate the potential value of their strategies before receiving any incentive. PDS may be able to identify and eliminate plagiarism while these plagiarists are still attempting to pass this challenge period. This would safeguard the rightful earnings of honest miners, ensuring the integrity of the incentive system and mitigating disruption.

6 Future Development

Multiple versions of PDS have been developed and improved upon, with this most recent iteration being the most advanced. As with any system, further improvements can be made as outlined below.

6.1 Combination Plagiarism Detection

We could broaden our approach to address combination plagiarism, the act of copying and combining signals from multiple miners rather than just one. This could possibly be achieved by extending our structural similarity and time analysis methods. However, a new algorithm designed to decompose a plagiarist's signals into their original sources may offer a more precise method of detection.

6.2 Dynamic Adjustments to External Variables

In financial market prediction, if numerous honest users' predictions converge during a specific timeframe, it could signal strong convergence between strategies based on a powerful external factor. The detection system could be refined by explicitly considering such scenarios to minimize false positives.

7 Conclusion

PDS offers a robust solution to the pervasive threat of plagiarism in time series forecasting. By maintaining fairness, promoting uniqueness of information, and safeguarding the integrity of incentive mechanisms, PDS fosters a healthy and competitive environment. Through continuous development and refinement, PDS will remain a crucial tool in upholding the standards and advancing the field of time series forecasting in SN8 and Bittensor as a whole.