January 2024

PREDICTIVE AI-BASED CHANNEL SOUNDING MECHANISM

Robert Barton

Jerome Henry

Indermeet Gandhi

Federico Lovison

Doosan Jung

See next page for additional authors

Inventor(s)
Robert Barton, Jerome Henry, Indermeet Gandhi, Federico Lovison, Doosan Jung, and Qihong Shao

PREDICTIVE AI-BASED CHANNEL SOUNDING MECHANISM

AUTHORS:
Robert Barton
Jerome Henry
Indermeet Gandhi
Federico Lovison
Doosan Jung
Qihong Shao

ABSTRACT

Techniques are presented herein that support examining the channel state information (CSI) matrix that arises from a channel sounding and using the results of that examination to train a machine learning (ML)-based model that considers all of the available wireless parameters at each given measurement. Under the presented techniques, after sufficient training has been completed a CSI matrix may be predicted over short intervals (using, for example, a long short-term memory (LSTM) network) thus allowing a Wi-Fi access point (AP) to reduce the sounding frequency and, as a result, improve overall wireless performance.

DETAILED DESCRIPTION

Channel sounding is a necessary process within a Wi-Fi network environment to measure the characteristics of the wireless channel between an access point (AP) and a station (STA). Such information is crucial for optimizing the involved transmission parameters and ensuring reliable and efficient wireless communication.

Channel sounding may be used under a variety of scenarios, including when deep learning (DL)-based beamforming is being initiated or when multi-user multiple-input and multiple-output (MIMO) technologies are employed, and needs to be refreshed when the channel changes, when the STA moves, when the STA roams, when interference occurs, etc. Because channel sounding is so crucial to the operation of a modern Wi-Fi network, it is performed frequently, allowing an AP to adjust to changing channel conditions in a semi-continuous manner. However, channel sounding comes at a price – the sounding process consumes airtime and, as the relevant standard continues to evolve, the resulting feedback matrix grows larger and more and more complex.

Additionally, in a stable wireless local area network (LAN) the sounding reports that are returned by a STA may offer little new information. However, that is a fact that an AP cannot know before it receives a feedback matrix.

Some existing solutions attempt, but fail, to address the above-described challenges. For example, the CSI compression use case as articulated by the Institute of Electrical and Electronics Engineers (IEEE) Artificial Intelligence Machine Learning (AIML) Topic Interest Group (TIG) anticipates the above-described CSI feedback matrix size issue but limits its approach to just finding an efficient compression algorithm.

Techniques are presented herein that support an artificial intelligence (AI)-based mechanism that allows the information from prior sounding reports to act as a predictor for a future channel state, allowing an AP to perform channel sounding less frequently in certain situations. Among other things, the techniques fit well within the AIML TIG CSI compression use case.

The presented techniques comprise a series of steps. A first step encompasses a normally-functioning wireless LAN where an AP performs channel sounding in the usual fashion. The AP may forward each CSI feedback matrix to a learning machine, along with all of the radio frequency (RF) elements (such as a received signal strength indicator (RSSI) for each of the radio chains, a signal-to-noise ratio (SNR), and, when it is available, a STA type) that were collected by the AP.

Under a second step, channel sounding occurs in regular time intervals as well as a result of a trigger event (such as, for example, when an AP prepares to send a beamformed frame). Thus, the metrics that are recorded from the sounding reports may be viewed in time-series and may be used as input variables to a learning machine.

Accordingly, the presented techniques encompass a learning machine which records events in time-series. Within such a context, a typical AI method may comprise some form of recurrent neural network (RNN), such as a long short-term memory (LSTM) network or gated recurrent units (GRUs). A LSTM network has the advantage of maintaining a short-term memory of past events as a way to predict future events. A transformer is the latest state-of-the-art in modeling, and considering the fact that an AP is a fairly constrained device it is possible to first obtain a transformer and then leverage knowledge distillation from a teacher-student model to shrink the model size to be able to deploy a lightweight version of the same to an AP.

Under a third step, the channel sounding reports, along with all of the relevant AP metrics, may be fed in time-series as input variables to train a model. Eventually, with sufficient training the model will reach a point where it can predict the future state of a channel for a STA at a given RSSI and given the STA's previous CSI feedback matrix. Figure 1, below, depicts elements of this step.

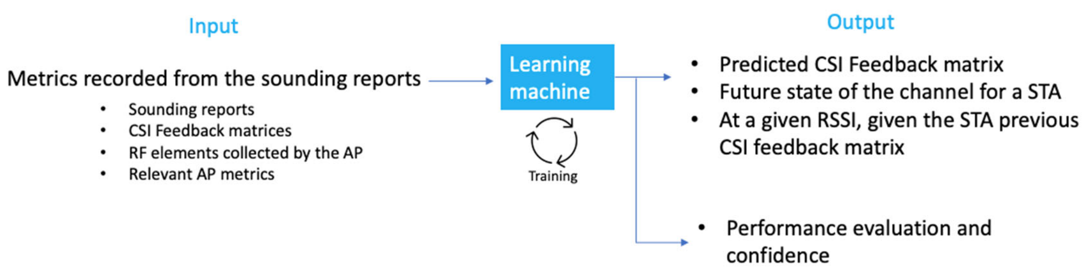

Figure 1: Illustrative Third Step

At step four, once the model has been trained an AP may send a first null data packet (NDP) sounding frame and receive a CSI feedback matrix. The learning machine may then parse the matrix and predict the next matrix. Later, the AP may send a second NDP sounding frame and receive a second CSI feedback matrix. The learning machine may then compare the predicted feedback matrix to the received feedback matrix and compute the accuracy of the prediction model. Figure 2, below, depicts elements of this step.

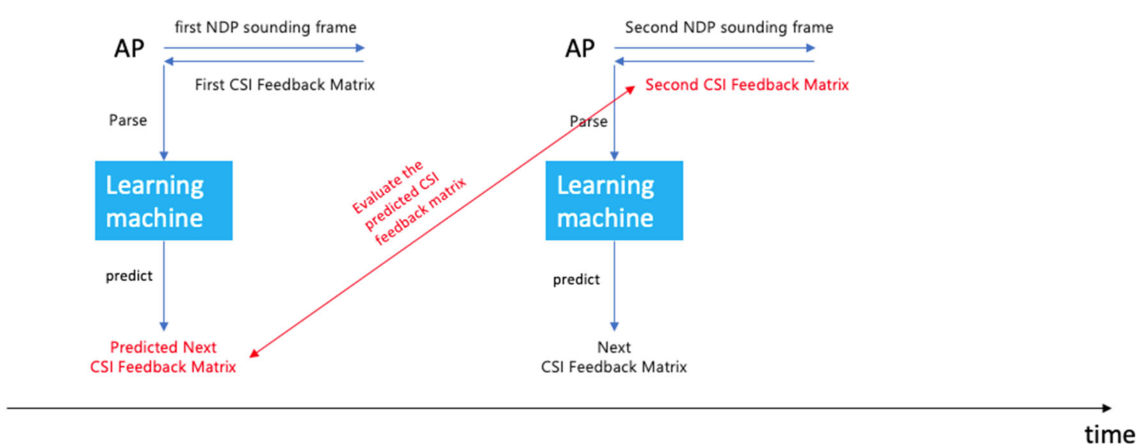

Figure 2: Illustrative Fourth Step

It is important to note that Figure 2, above, employs a simple 1+1 matrix arrangement as an example. In an actual deployment of the presented techniques, a learning machine would likely need to process n matrices before it is able to predict an n+1 matrix.

During a fifth step, in a wireless LAN (WLAN) environment, and for STAs that exhibit a high correlation between a predicted and a true channel state, it is clear that the AIML algorithm is performing well and the STA in the WLAN is stable. Under a first scenario, the AIML engine may predict an entire matrix with a high confidence. Under a second scenario, the AIML engine may correctly predict a matrix for one or more radio chains but display a poorer confidence for the prediction of other radio chains (as a result of, for example, a multipath environment or an obstacle that causes distortion on one or more of the paths in the previous matrices). Under a third scenario, the AIML engine may predict with a high confidence a subset of the matrix for a given chain but display a poorer confidence for another part of the matrix for that same radio chain (as a result of, for example, a narrow band interferer or micro-interference that may be caused by a multipath environment). Naturally, the second and the third scenarios may occur concurrently.

At a sixth step, based on a confidence level the AP may send a novel null data packet announcement (NDPA) that requests just the specific subsets that fall below a configurable target confidence level. Such a NDPA may include the radio chain index(es) and the indices of the requested tones. The NDP sounding frame may proceed as normal and the resulting CSI feedback matrix may include just the specific subset that was requested by the AP. Figure 3, below, depicts elements of this step.

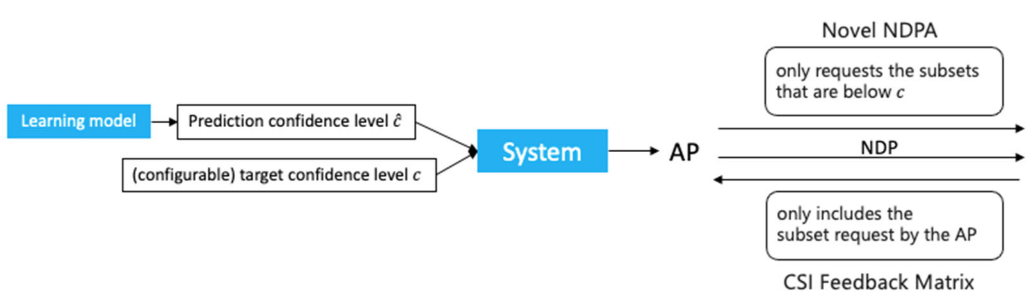

Figure 3: Illustrative Sixth Step

During a seventh step, as the above-described predictions become more accurate it will become less necessary to perform a channel sounding. Accordingly, an AP may gradually begin to reduce the channel sounding interval. At first, the AP may skip every nth sounding exercise to determine how well its prediction algorithm is performing. Over time, if the WLAN continues to be stable, and a prediction proves to be accurate, the AP may continue to further reduce the sounding interval until it reaches some minimum level. The effect of such reduced sounding activity in a busy WLAN will lead immediately to a higher overall performance of the WLAN. Thus, the AP may skip the usual sounding requests and only issue a request every, for example, nth sounding exchange. Alternatively, the AP may occasionally perform a sounding exchange but ask a STA to only return a sample matrix (e.g., a random set of tones in a random number of radio chains, the tones that have the lowest confidence along with a random set of tones that have a greater confidence, etc.). Figure 4, below, depicts elements of such an approach and highlights aspects of such a dynamic sounding exercise scheduling approach.

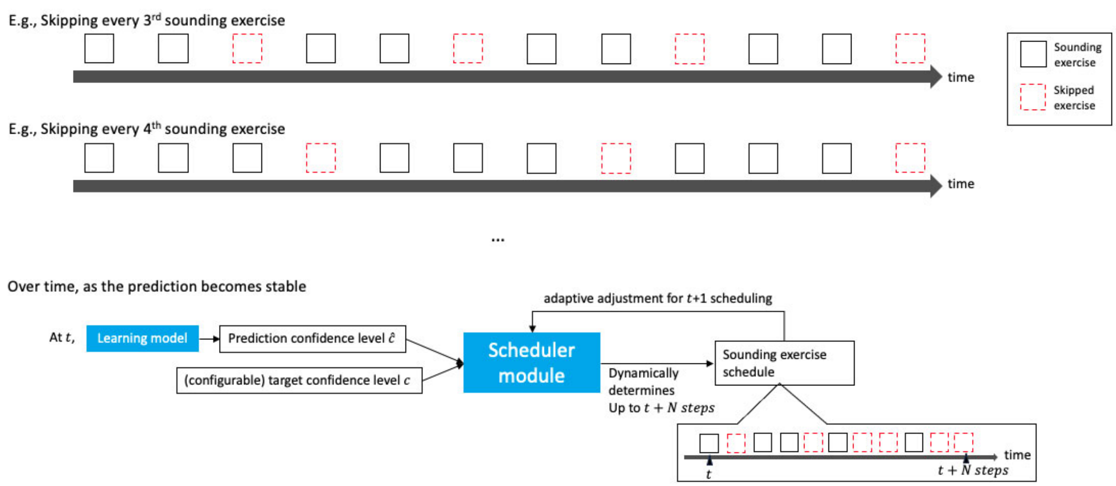

Figure 4: Illustrative Sixth Step

Under an eighth step, in cases where a prediction becomes poor (meaning that either the model has not been sufficiently trained or a WLAN is simply too stochastic or unstable), the channel sounding frequency may increase until it returns to its normal intervals (i.e., there is no reduction in normal channel sounding). Later, when a prediction becomes more accurate, the sounding interval may then be gradually reduced (as articulated in the preceding steps).

As described and illustrated above, the presented techniques encompass a learning machine. Such a learning machine may comprise a lightweight yet powerful model that leverages recent technological advances that facilitate small, but well-performing, models in resource constrained devices. Such advances may include model pruning, quantization, knowledge distillation through a teacher-student framework, low-rank matrix factorization, or any combination of one or more of these methods. Figure 5, below, provides an illustrative depiction of several lightweight, yet powerful, machine learning frameworks.

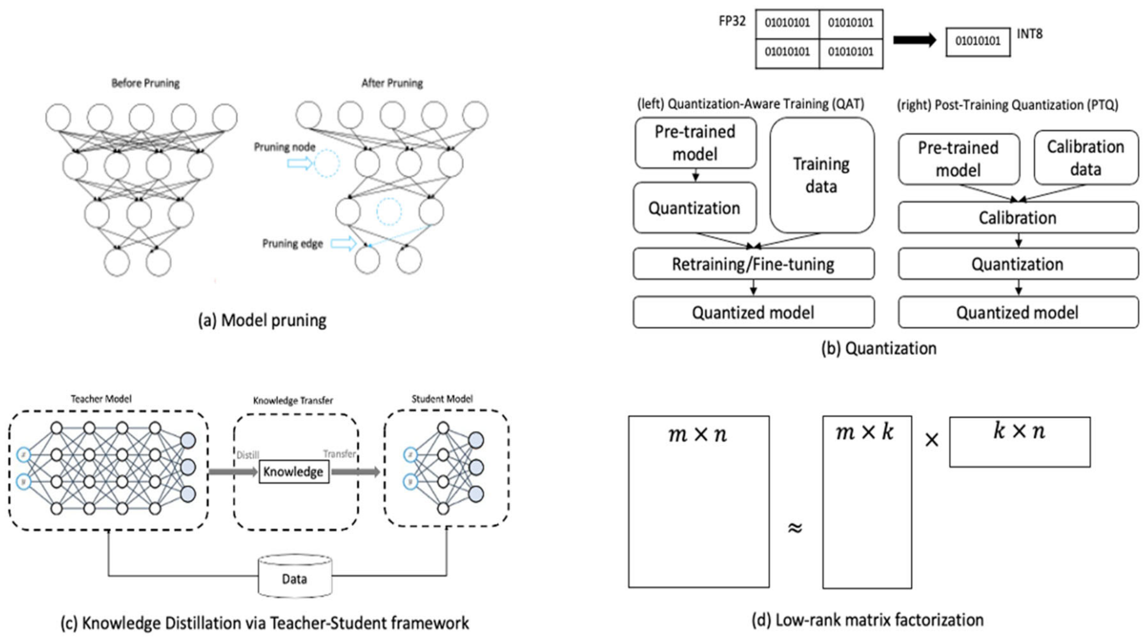

Figure 5: Illustrative Machine Learning Frameworks

Within the above-described context, a CSI matrix may be viewed as a three-dimensional (3D) digital image and a time series of four-dimensional (4D) CSI tensors may be viewed as a flow of digital images. By formulating the instant problem as one of generating a next image (or frame) prediction, it is possible to leverage existing spatiotemporal predictive and generative learnings (such as a convolutional LSTM network, a variational autoencoder (VAE), a generative adversarial network (GAN), or video diffusion techniques) and engineer the above-described models for deployment to resource-constrained devices. In such a way, an AP may be equipped with all of the benefits that arise from powerful models such as transformer architecture-based models.

The presented techniques also encompass a scheduler module. Such a module may operate on a variety of elements including, for example, a learning model's performance and prediction confidence level, a target confidence level (which may be configurable), and, if it exists, a previously-created sounding exercise schedule. The scheduler module's principal task is the creation of the next N-1 sounding exercises, which may be based primarily on the identified differences between a model's confidence level and a target confidence level. Importantly, if a previously-created schedule exists then the scheduler module does not necessarily need to create a new schedule from scratch but may, instead, adaptively change the previous schedule based on the current confidence level.

The above-described scheduling activity may leverage different sequential machine learning models (such as, for example, one-to-many or many-to-many RNNs, a Seq2Seq model, or sequential transformer models) in order to generate a future sequence (which is, in effect, the next sounding exercise schedule).

In summary, techniques have been presented herein that support examining the CSI matrix that arises from a channel sounding and using the results of that examination to train a ML-based model that considers all of the available wireless parameters at each given measurement. Under the presented techniques, after sufficient training has been completed a CSI matrix may be predicted over short intervals (using, for example, a LSTM network) thus allowing a Wi-Fi AP to reduce the sounding frequency and, as a result, improve overall wireless performance.