Page 4

They utilize different frequency communication signals to transmit and receive information.

RF communication is ideal for long-range coverage, penetration through obstacles, and reliable performance in various weather conditions. FSO communication offers high data rates, low interference, and enhanced security, making it suitable for high-speed data transmission. A hybrid system can leverage the advantages of both channels. This hybrid system is often used in satellite, airplane and infrastructure etc.

Page5

As we all known, just like the figure shown, The channel attenuation keep increasing along with the increasing of precipitation intensity. In this study, the weather factors or features involves hunmidity,temperture,wind ...

Page 8

IN the second part，we first introduce regression Random Forest which is an ensemble algorithm comprising a number of decision trees. Each decision tree is known as a weak estimator and it sampled from the population using bootstrap methods, which involves random sampling with replacement. The forest aggregates the outputs of all estimators to provide the final predictive result. The generic，specific and hybrid model are all designed based on Random Forest。

Page9

The generic model, trained on the entire training dataset, can predict channel attenuation under all weather conditions. In contrast, specific models are trained for different weather conditions, each model can predict attenuation only under a specific weather scenario. We will develop seven specific models, each corresponding to a different weather condition.

The hybrid model introduces a key innovation: it utilizes the predicted channel attenuation from one model as a new feature to predict another channel's attenuation. This approach comprises two methods: M1 uses the predicted RFL\_Att to forecast FSO\_Att, while M2 does the reverse, using FSO\_Att to predict RFL\_Att.

Page 12

RMSE provides a measure of the model’s error in the same units as the target variable, while R2 evaluates the proportion of variance in the target variable explained by the model. PCC that means person corralation coefficient and Mutual Information are primarily used to assess whether the hybrid model can capture the relationship between RFL\_Att and FSO\_Att.

Page 15

The establishment of model actrually is the process of hyperparameter tuning. This study designs it by coarse tuning and fine tuning, the numeric range can be shown in the table.

Page 16

The establishment of model actrually is the process of hyperparameter tuning. This presentation uses the FSO model training as an example to illustrate the process of generic model development.The learning curve in the coarse tunning indicates that as the number of estimators increases, the RMSE decreases and the R² increases. However, when the number of estimators reaches approximately 130, further increases do not improve RMSE and R². Additionally, with the same number of estimators, a maximum depth of 30 yields the best model performance, which was indicated by the block dash line. And the fine tuning can identify the optimized hyperparameters by a more precise way. specific and hybrid model all adopted the same tunning strategy, so I will not go into details here

In here, I use the wrapper method to conduct feature importance and pruning, the method evaluate the importance of features by iteratively removing features and assessing the model's performance. This approach helps in selecting the most relevant features in RFL AND FSO generic model respectively. In the two figures, the features located to the left of the green dashed line are unimportant features. And the table indicates unimportant features do not affect the model's performance even if they are deleted.

Page 17

The process of coarse tuning helps us locate the approximate optimal point for the model, while fine tuning can identify the optimized hyperparameters by a more precise way. The fine tuning figure tell us When the number of estimators reaches 130 and the maximum depth reaches 30, the FSO model achieves locally optimal hyperparameters.

Page 19

The feature importance figure can quantitatively show which features are more important for model development. We can see that distance is the most important feature for the FSO model, indicating that distance significantly determines FSO attenuation. Similarly, rain intensity, distance, and absolute humidity are the three most important features for the RFL model.

We can also observe that some other features contribute less to the model development. This project uses wrapper methods to evaluate and simplify input features.

Page 20

The wrapper method for random forests involves using the model to evaluate the importance of features by iteratively removing features and assessing the model's performance. This approach which called feature pruning helps in selecting the most relevant features to improve the model's accuracy and efficiency.In the two figures, these features located to the left of the green dashed line do not affect the model's performance even if they are deleted.

Page 22

in this study, there are 7 different weather conditions. The 7 specific models have different performance in prediction and important predictors. Let us take the foggy and Drizzle condition for example. In the RFL\_foggy\_MODEL, relative humidity are the most significant factors, while in the RFL\_drizzle\_MODEL, absolute humidity takes precedence. obviously, the RFL\_drizzle\_MODEL requires more important features to attain a low RMSE than, RFL\_foggy\_MODEL. I summarize The frequency of important features in both channels as the histogram shown.

The frequency of important features in both channels was illustrated in histograms. Absolute humidity appeared in all RFL-specific models, while temperature and distance appeared in 6 FSO-specific models.

In the left figure, the dashed lines are generally below the corresponding solid lines, indicating that the generic model has a lower RMSE compared to the specific models under the same weather conditions. This suggests that the generic model performs better than the specific models.

Page 23

The frequency of important features in both channels was illustrated in histograms. Absolute humidity appeared in all RFL-specific models, while temperature and distance appeared in 6 FSO-specific models. Based on the findings above and the comparison of model metrics between with and without feature pruning, we can conclude that Feature Pruning can quantitatively determine feature importance under different weather conditions, effectively reduce model complexity and avoid overfitting.

Page 24

Through the comparison of results, it's evident that while RFL-specific models may perform better under clear and snowy conditions, in most cases, generic models in both channels exhibit superior prediction performance. Considering the versatility and effectiveness of the generic model, conclusion 2 can be drawn: The generic model excels in predicting channel attenuation across diverse weather conditions.

Page25

The figure shows the structure of the hybrid model. Specifically, in the first stage, the training and hyperparameter tuning of Random Forest 1 is the same as in the separate model. The biggest difference is in the second stage, where the first predicted channel attenuation is used as part of the training set for Random Forest 2. The purpose of hybrid model is to leverage the intrinsic relationship between the FSO and RFL channels, along with weather factors, to predict channel attenuation.

Page 26

The left figure shows the actual PCC and NMI between the RFL and FSO channels, and the right figure illustrates that the predicted PCC and NMI align with this distribution, indicating that the hybrid model effectively captures the correlation between the two channels.

Specifically, in dusty environments, (Dian),there is a moderate negative linear correlation between the attenuation of the two channels, meaning that a decrease in attenuation in one channel corresponds to an increase in the other. This serves as a good indicator for the hybrid model.

Additionally, under dust, fog, and shower conditions, (dian) the NMI value is higher than the PCC, suggesting that the correlation between the two channels includes a significant non-linear relationship. These conditions also serve as good indicators for the hybrid model.

However, the left figure indicates that the predicted attenuation of one channel contributes just 4.8 percent to feature importance, suggesting that one channel's predicted attenuation doesn’t significantly impact predicting the other channel's attenuation in hybrid model. (dian)The right table and figure further demonstrate that the hybrid model performs similarly to the generic model, and its performance is even slightly lower in certain weather conditions.

Now, we compare all models together. I trained the hybrid model under different weather conditions, similar to the specific models. We can see that the hybrid models perform better than the specific models in almost all weather conditions. This is because the hybrid model uses the attenuation of an additional channel as a predictor.

However, except for the hybrid model under foggy conditions(dian) and the specific RFL model(dian) under snowy conditions, the generic model performs better in all other scenarios.

Next, we explore the linear relationship between the FSO and RFL channels in the Hybrid Model. In the left figure, the PCC of the two channels shows a moderate negative correlation under dusty conditions. This means that when the communication system encounters dusty weather, the FSO channel suffers significant attenuation, while the RFL channel may provide more stable communication with less attenuation. Additionally, the two channels exhibit strong or moderate positive correlation under foggy, snowy, and rainy conditions, indicating that their performance improves or worsens simultaneously, making it difficult to determine which channel has better communication performance.

Furthermore, this study explores the PCC of the FSO channel and RFL channel at different frequencies. The right figure shows that the 83.5 GHz RFL channel has a strong negative correlation with the FSO channel, while the 73.5 GHz RFL channel exhibits only a weak negative correlation with the FSO channel. This indicates that the hybrid model can accurately capture the linear relationship between the two channels under different weather conditions. Specifically, in dusty conditions, the 83.5 GHz RFL channel may have better communication quality than both the 73.5 GHz RFL and the FSO channel

Page 27

The PCC can only reveal the linear correlation between the two channels, but it cannot uncover nonlinear relationships. Mutual information can effectively address this limitation. The left figure illustrates the joint distribution probability of the pair of RFL and FSO attenuation, where the highlighted points represent higher probabilities.

The joint distribution probability can be used to calculate the mutual information, as shown in the right figure. From the figure, it can be concluded that the hybrid model designed in this project can effectively capture the mutual information between the two channels. The mutual information under dusty, foggy, snowy, and shower conditions is relatively higher than that under clear and drizzle rain conditions, indicating that the hybrid model may better utilize the mutual information of the two channels to predict channel attenuation under certain weather conditions.

Page 28

However, the left figure indicates that the predicted attenuation of one channel contributes just 4.8 percent to feature importance, suggesting that one channel's predicted attenuation doesn’t significantly impact predicting the other channel's attenuation. The comparison of model metrics between the generic and hybrid models directly illustrates that the generic model has better prediction performance than the hybrid model, even though the hybrid model utilized the mutual information between the two channels.

The predicted attenuation of one channel has less feature importance for predicting another channel under all wheather conditions, resulting in a prediction performance that is slightly worse than that of a generic model.

The generic model excels in predicting channel attenuation across diverse weather conditions.

In FSO generic model, Distance, Visibility, Temperature, ParticulateMin and VisibilityMax are the most significant predictors, while AbsoluteHumidity, RainIntensity, Distance, and RainIntensityMax are the most important features in RFL generic model.

Firstly, feature pruning can effectively reduce model complexity and identify the most significant weather features for each model. Specifically, in the FSO generic model, Distance and Visibility are the most significant predictors, while Absolute Humidity and Rain Intensity are most important in the RFL generic model.

（点）In FSO-specific models, Temperature and Distance are common features, whereas Absolute Humidity is predominant in RFL-specific models.

Additionally, （点）the study identifies key predictors under different weather conditions. For example, in the RFL channel, Relative Humidity is the most significant predictor on foggy days, while Absolute Humidity is more important on drizzly days.

Considering both model performance and complexity, generic models perform best across both RFL and FSO channels, providing more accurate predictions under most weather conditions. Hybrid models can effectively capture the correlation information between RFL and FSO channel attenuation, allowing them to outperform specific models. In conclusion, generic models are the best, followed by hybrid models, with specific models ranking last.

I also conducted exploratory data analysis to understand the distribution and correlation among different predictor variables, which provided an initial understanding of various features. For example, Temperature and Absolute Humidity are highly correlated, while Particulate Matter and Rain Intensity show a strong correlation. This suggests that these paired features may have a special impact in model building.