Information-Theoretic Modeling & Markov Chain Analysis for RF-Based Drone Classification

Sujal Kumar Singh - 231010250 ECE

November 25, 2025

Abstract

This report explains how Information Theory and Markov Chains can be used to classify drones based on their RF (Radio Frequency) signals. Instead of relying only on simple metrics like mean and variance, we model the sequence and randomness of the signal to capture deeper structural patterns. By converting I/Q samples into symbolic states, building transition matrices, and analyzing entropy and divergence, we extract unique "fingerprints" for each drone type. These fingerprints improve classification, achieving **81.23%** accuracy using a Stacking Classifier.

Contents

1 Introduction

The aim of this project is to classify drones by studying how their RF signals behave over time. Different drone manufacturers use different modulation styles, timing patterns, and communication "rhythms" when their controllers talk to the drone. By treating the RF signal as a sequence of events instead of just a set of numbers, we can uncover patterns that traditional approaches miss.

Our approach uses:

- Symbolization of I/Q data (K-Means)
- Markov Chain modeling
- Shannon Entropy and Entropy Rate
- KL and Jensen–Shannon Divergence
- Feature ranking using Mutual Information

All of these combine to extract meaningful features that separate drone families.

2 Understanding the Signal

2.1 What the Drone is Actually Doing

A drone controller sends commands through radio waves. These waves constantly change to encode information like "move up," "rotate," etc. The interesting part is that each manufacturer tends to have a unique style:

- **DJI** $\rightarrow$ smooth changes in phase
- **Futaba** $\rightarrow$ sudden jumps between frequencies

These styles show up in the I/Q data.

2.2 What Are I and Q?

I/Q data represent the signal in two dimensions:

- **I (In-Phase)** = Amplitude $\times$ cos(Phase)
- **Q (Quadrature)** = Amplitude $\times$ sin(Phase)

You can imagine the signal as a clock hand rotating. The I and Q values are simply the x- and y-coordinates of the hand.

Certain drones form:

- circular clusters $\rightarrow$ meaning stable amplitude, rotating phase
- tight dots $\rightarrow$ meaning almost no variation

These shapes already reveal the modulation style.

3 Step 1: Symbolizing the I/Q Data

Entropy and Markov models work on discrete symbols, not infinite decimal values. So we convert the continuous I/Q points into a small set of symbolic "states".

3.1 Using K-Means

We cluster the I/Q data into 8 groups. Each cluster becomes a symbol from 0 to 7.

Example sequence:

$$0 \to 0 \to 1 \to 7 \to 2 \to \ldots$$

This is like converting a detailed map into regions:

- Water
- Forest
- Mountains
- City

This simplifies the signal into something we can analyze structurally.

4 Step 2: Markov Chains – The Signal's "Grammar"

Once we have a symbol sequence, we can ask:

"What is the probability of going from State A to State B?"

This gives us a **Transition Matrix**. Every drone type has a unique style in how it moves between states.

Examples:

- **DJI Matrice:** often stays in the same state → stable communication
- **Futaba:** jumps frequently between states → dynamic communication

This transition matrix becomes a very strong fingerprint.

5 Step 3: Entropy – Measuring Predictability

5.1 Shannon Entropy

This checks how uncertain we are about the drone's state at a random moment.

- **High Entropy** → drone moves across many states equally
- **Low Entropy** → drone stays mostly in one state

5.2 Entropy Rate

This measures the unpredictability of the *next* state.

Examples:

- Repetitive signals → **low entropy rate**
- Fast-changing signals → **high entropy rate**

We found:

- **Futaba** → high entropy rate
- **Matrice** → low entropy rate

6 Step 4: Divergence – Measuring Differences Between Drones

6.1 KL Divergence

KL tells us how "surprised" we would be if we assumed signal A but observed signal B.

6.2 Jensen–Shannon Divergence (JSD)

A smoother and symmetric form of KL. We used JSD to build a **dendrogram** (family tree of drones).

It correctly grouped:

- `DJI_Phantom4Pro_1` with `DJI_Phantom4Pro_2`

This confirms that our fingerprints reflect real similarities.

7 Step 5: Feature Importance and Classification

7.1 Mutual Information

We measured how useful each feature is for predicting the drone type. The most informative feature was:

`phase_circular_variance`

Because different drones handle phase modulation in different ways.

7.2 Stacking Classifier

We combined:

- Random Forest
- CatBoost

This gave the best results.

Model	Accuracy
Random Forest	80.20%
CatBoost	78.80%
Stacking Classifier	**81.23%**

Table 1: Final Classification Performance

8 Conclusion

This project shows that RF signals contain much deeper structure than simple amplitude or frequency changes. By treating the signal as a sequence and analyzing it using entropy, divergence, and Markov transitions, we discover patterns that are consistent across drone models.

These patterns act as reliable fingerprints and significantly improve classification performance. The results match expectations:

- Complex controllers → higher entropy
- Similar drone families → low JSD distance

Overall, Information Theory + Markov Modeling provides a powerful, intuitive, and efficient way to classify drone RF signals.