Problem statement

Problem statement is unclear. The motivation for RL in the first paragraph is very confusing. I think there is more convincing motivation than "we can't collect labels" and "define the reward function without any explanations of how to optimize it". The second phrase is incorrect. Any task (even in classical ML) doesn't contain explanations of how to optimize.it. "Trials could be errors"?. Agent doesn't develop some strategy. It tries to find the opimal one.

Also, i'd recommend to add more links in this section. Optionally, you can add picture representing the problem in clear way.

Section 2

What algorithm? Mb, the main goal is to create a algorithm which ...

Exploration can be implemented in the following levels: Step-level, Experiment-level, Episode-level ...

There are following strategies for exploration

- Informed switching
- Blind switching
- Combining

For 2.3 there are no mathematical formulas explaining provided algorithms.

Related work

In this section I would recommend to add more references. Exporaliation/Explotations is one of the main problem of Reinforcement learning. So, there are many papers solving this problem.

Data

Add more description of dataset, add examples from this dataset.

Recommendation

Fix latex, language mistakes

Experiment

- The experiment protocol is reasonable.

Results

- There are no results presented in the paper

Code

`requirements.txt` contains error

I was unable to reproduce experiments because I have to manually specified all paths.