



1. Duplicated bugs
2. Non deterministic bugs
3. Missing information
4. Environment configuration
5. Conflict and misunderstanding
6. Performance related bug
7. Memory issues
8. Lack of cooperation
9. Dependence on stakeholders
10. Security and Permission related issues
11. Network/DB/GUI related issues
12. Accessibility bugs
13. Others

Common

1. Missing information?

2. External dependencies.

3. Environment

4. Non-deterministic

5. Lack of cooperation

6. Performance

7. Memory related

8. Conflict and misunderstanding

9. Outdated

10. Inaccurate

~~IDEAS~~

}

Saeket (Focused on SD)

- ✓ 1. class/interface/method not found
- ✓ 2. Missing code
- ✓ 3. External library not found
- ✓ 4. Identifier/object not found
- ✓ 5. Too short code snippet
- 6. Misc
- ✓ 7. Database/File NI dependencies
- ✓ 8. Outdated code
- ✓ 9. Inaccurate claim
 - Not really a bug
- ✓ 10. Ill defined issues
 - non-deterministic

All changes to DottedNet

Alshangit	Zohreh	Sun	Zhang	Zhang	Thunay
- More no ans + no accepted answer	- Model Creation & data prep is challenging - Type and shape mismatch questions - Data cleaning - Questions of exceptions - Parameter selection - hyperparameter tuning - DL has more training difficulty	1. Compatibility bug 2. Variable bug 3. Documentation	- program crasher Model migration implementation training anomaly computation - sign - installation Performance	① Incorrect model parameters ② Unaligned tensors ③ Confusion with API misuse GPU comp. hyperparameter selection static graph limited debugging	Algorithm Assignment data External interface Internal interface logic non-functional Timing/optimization Config other
- 10 more times to answer		4. Performance optimization 5. Memory overflow 6. Design defect 7. Others		④ API change ⑤ API misuse ⑥ Structure inefficiency ⑦ Others	
- Lack of exports					
- DP & MD most challenging					
- DDP & FP is also challenging					
- CNN & DL is also discussed					
- TF has a higher level of difficulty					
- Most challenging, that need concept implementation					
- rotation					

Challenges

- * DP, MD, FP, Model creation, Data cleaning
 - * TF is most challenging
 - * DL^{NN} has more training difficulty
 - * Parameter & hyperparameter Selection
 - * Questions of exceptions
 - * Type and Shape
 - * Concept and Implementation
 - both are challenging
 - * Program crashes
 - * Installation
 - * Model Migration
 - * training anomaly
- 1
- ① Compatibility
 - ② Variable
 - ③ Documentation
 - ④ Performance
 - ⑤ Memory overflow
 - ⑥ Design defects
 - ⑦ API misuse and change
 - ⑧ Incorrect hyperparameter
 - ⑨ Shape mismatch
 - ⑩ Configuration

UnCommon

Variables

Shape mismatch

API misuse

A lot of
how-to

documentation

hypopatic

Lack of expertise

Common

missing info

memory

configuration

Environment

Performance

External dependencies

Common

1. Missing information?
2. External dependencies.
3. Environment
4. Non-deterministic
5. Lack of cooperation
6. Performance
7. Memory related
8. Conflict and misunderstanding
9. Outdated
10. Inaccurate

Reason

My Findings from the Analysis

AIShangiti → ML Pipelines

DP, MD most challenging

no-accept, no-answer

152 Posts.

challenges

how
shape

lack of expertise

lack of information

lack of communication /
cooperation

duplicate

missing info

explanation

Topics

Updated software

↳ Guidelines to tackle the challenges

lack of cooperation

-

too many questions

specify configuration

documentation

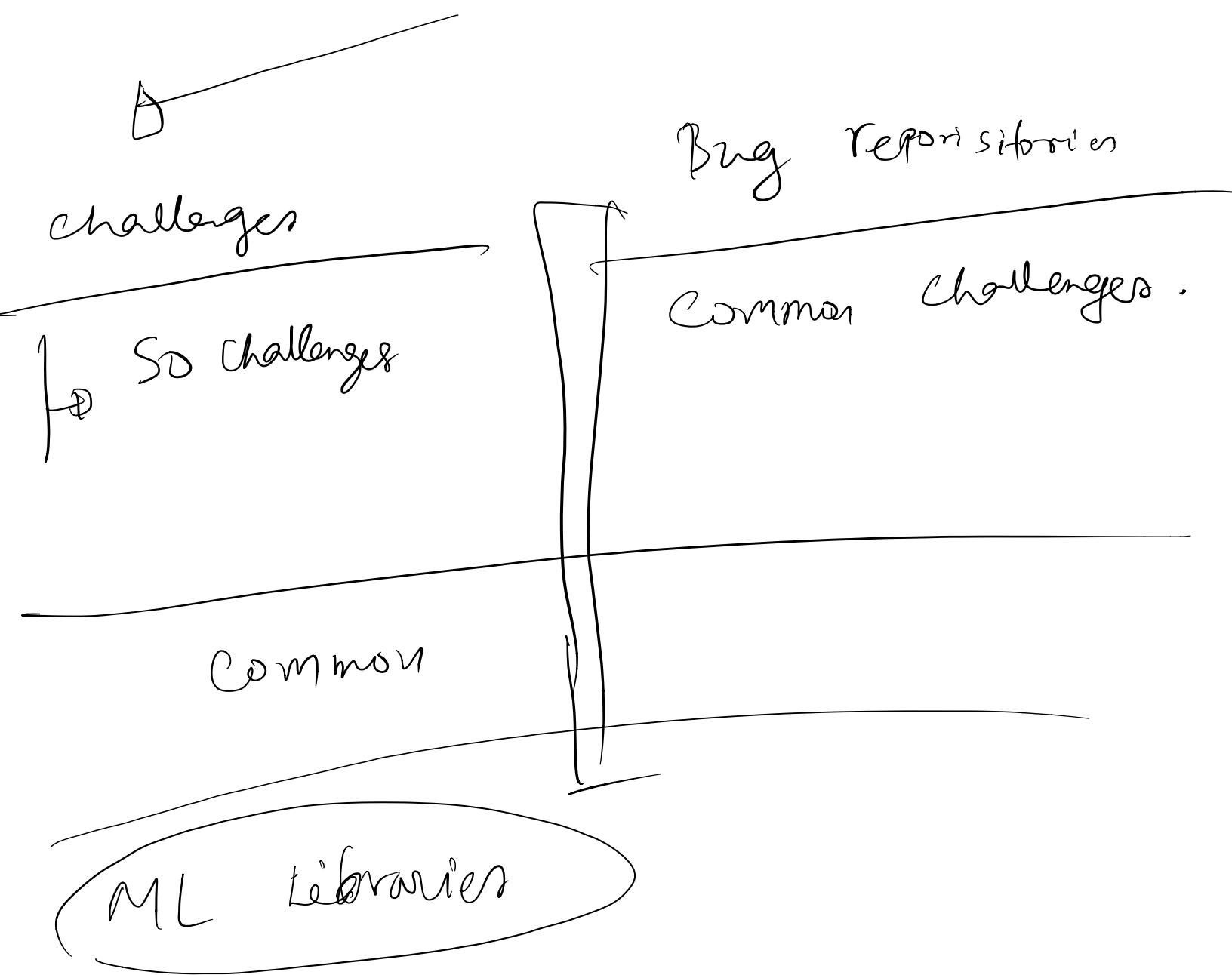
tags

No exact solution (no-one-on)

individual variety

Lack of Expertise

- Bugs/Issues posted in SE suffer from no acceptable answers
- Stats
- Reason: Reproduce the issue



challenges

Bugs

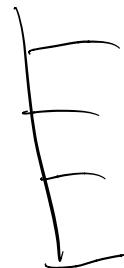
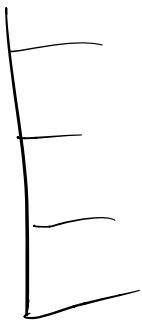
reproduce challenges

Common Challenges

ML Common Challenges

Common

uncommon



What I did with a dataset



How to solve the problem

