

A Taxonomy of Self-Admitted Technical Debt in Deep Learning Systems

Federica Pepe
University of Sannio
Benevento, Italy
f.pepe8@studenti.unisannio.it

Fiorella Zampetti
University of Sannio
Benevento, Italy
fiorella.zampetti@unisannio.it

Antonio Mastropaolo
William & Mary
Williamsburg, VA, United States
amastropaolo@wm.edu

Gabriele Bavota
Software Institute
Università della Svizzera Italiana
Lugano, Switzerland
gabriele.bavota@usi.ch

Massimiliano Di Penta
University of Sannio
Benevento, Italy
dipenta@unisannio.it

背景

深層学習システムの広がり

- ・ 産業分野での利用が広がっている
 - － 金融 医療 自動運転 ソフトウェア工学 など

DLシステムの品質保証

- ・ システムの不具合が引き起こす影響は大きい
 - － 経済的な損失や人命に関わることもある

DLシステムの品質を確保することは極めて重要

DLシステムの品質

すべての問題がバグではない

- 開発者が認識している未解決の問題がある（SATD）
 - システムの誤作動を引き起こすものではない
 - ハイパーパラメータの選択など

DL特有のTD

- 従来のソフトウェアにないコンポーネントがある

 従来にない新しいタイプのTDがある

目的・アプローチ

目的

- DL特有のSATDの分類法を定義する
 - DL特有のSATDを追跡, 管理, 除去するための戦略を立てるため

アプローチ

- TensorFlow/PyTorchに依存する人気のあるPythonプロジェクトに含まれるSATDを分析

研究設計

対象データ

- GitHubでホストされるTensorFlow/PyTorchに依存する54のDLシステムから抽出した443のSATD

RQ

- DLシステムではどのような種類のSATDが見つかるか
 - 従来のシステムで発生しうるSATDとの区別する
 - DL-SATDの分類法を定性的に定義する

手順

データ収集



SATDのコメント抽出



分類法の構築

実験手順 | データ収集

プロジェクトの選定

- Pythonプログラミング言語を使用
 - DLアプリケーションに最も広く使用される言語
- 2つのDLフレームワークを使用するプロジェクト
 - TensorFlowとPyTorch
- 最終的に100のオープンソースプロジェクトを選択
 - スター順に並べ、フォークされたプロジェクトを除外
 - 手動で確認し、TensorFlowに依存する50件、PyTorchに依存する50件を選択
 - チュートリアル、コードブックなどを除外

実験手順 | SATDコメントの抽出

手順

- tensorflowまたはtorchをインポートしているファイルに注目
- ソースコードコメントを抽出
- SATDに関連するキーワードでコメントを絞り込み
 - 先行研究のキーワードを使用
- 特定のキーワードにマッチする正規表現を使用
 - TODO, FIXME, HACK, XXX
- SATDが存在しないプロジェクトを除外
- 層化ランダムサンプリングを実施
 - 443のSATDコメントを抽出

実験手順 | 分類法の構築

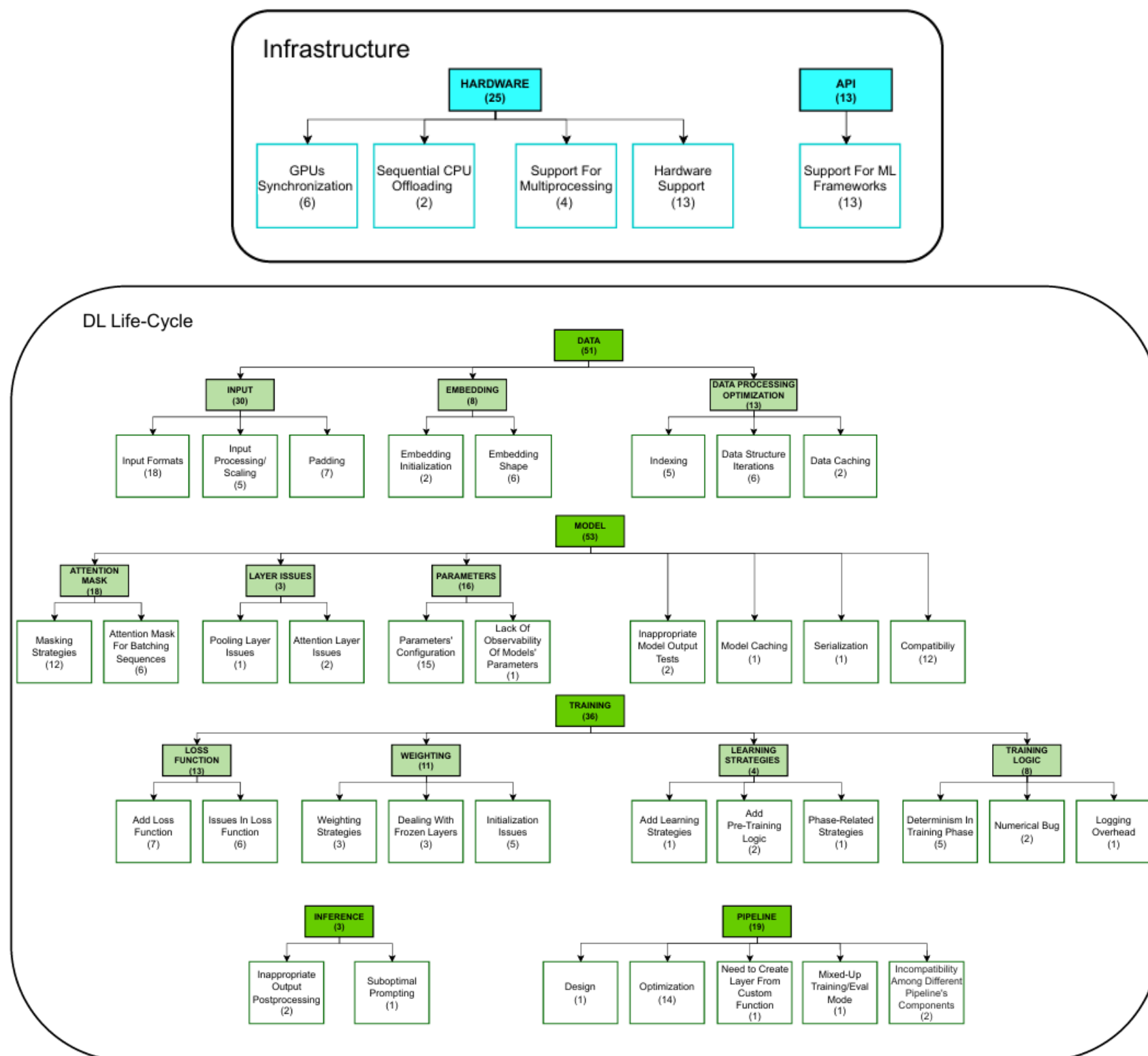
分類手順

- 4人の著者が協調的ハイブリッドカードソーティング手順を実施
- SATDを次のように分類
 - 誤検出
 - いいえ（DLコード/コンポーネントに特有ではない）
 - はい（DLコード/コンポーネントに特有）
 - “はい”の場合TDタイプを説明するラベルを付与

結果

全体

- 大きく 2 つに分類
- 7 つのカテゴリ
- 41 の葉



インフラストラクチャカテゴリ

ハードウェア（25件）

- ・ **ハードウェアコンポーネント**に関するSATD
- ・ モデルのトレーニング時や本番環境での推論時に使用される

外部APIとの相互作用に関連（13件）

- ・ DLフレームワークの使用と統合に関連するSATD
- ・ DLフレームワークがサポートしていない機能に関連

DLライフサイクルカテゴリ

データ（51件）

- ・モデルのトレーニングに使用される **データの形式** に関連するTD

モデル（53件）

- ・DL **モデルの設計と設定** に関するTD
- ・アテンションマスク，レイヤ，その他の関連パラメータの観点

トレーニング（36件）

- ・適切な損失関数の選択
- ・ **異なる設定** を実験する時間の不足がある場合がある

DLライフサイクルカテゴリ

推論（3件）

- ・ 生成される結果の不適切な後処理に関連
- ・ 最適でないプロンプティングに関連

パイプライン（19件）

- ・ DLパイプラインの設計と最適化の設定に関連

考察

DL SATDとDLバグの違い

- システムへの影響の種類の違い
 - バグ：システムがタスクを達成できない
 - SATD：トレーニング時間や収束性の観点でパフォーマンスが低下

静的解析ツールの適合性

- DL特有のTDを捉えられない
- 特殊なアプローチが必要

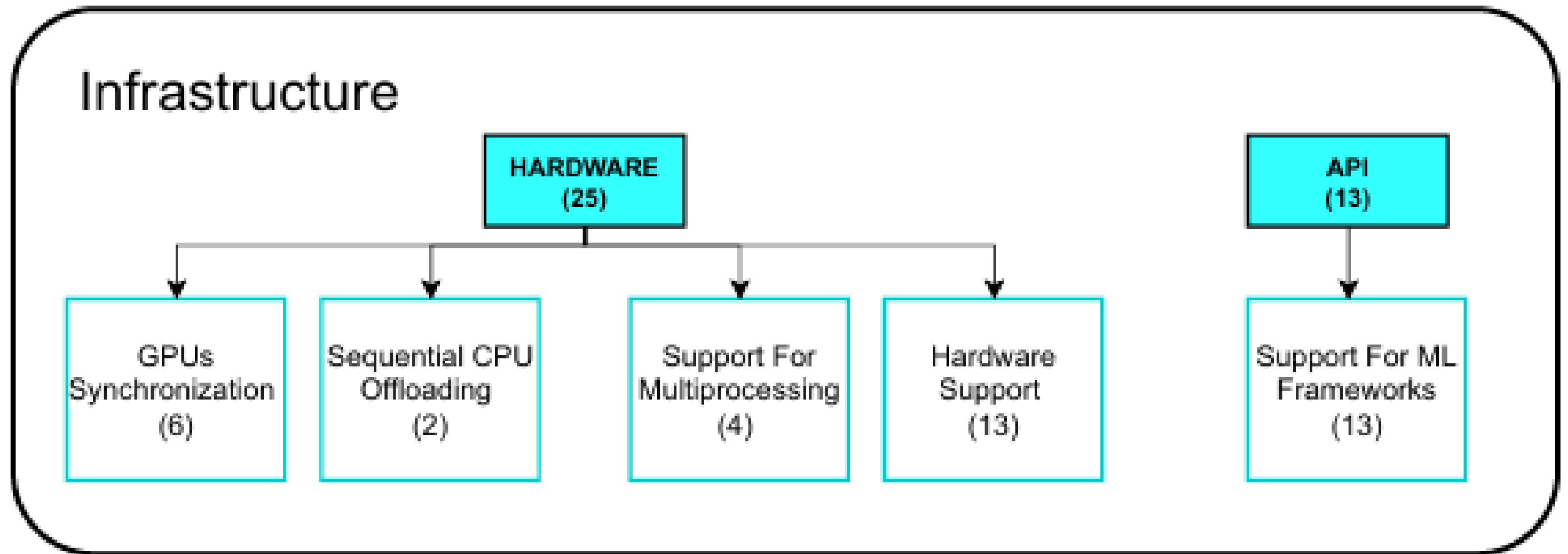
総括

| DL特有のSATDの分類を行った

| 7つのカテゴリに整理された

| 既存の静的解析ツールはSATDに対処できない

補足資料



補足資料

DL Life-Cycle

