Convolutional Neural Network programs modeling for automatic detection of design smells

Anonymous Author(s)

***Abstract*—TODO: To be completed**

***Keywords*—** Deep Learning, Convolutional Neural Network, Design Smells, Famix, Modeling, Software Engineering

I. Introduction

L'apprentissage profond est un sous ensemble du machine learning qui utilise des réseaux de neurones artificiels pour apprendre des données non structurées. L'apprentissage profond est devenu très populaire ces dernières années grâce à ses performances dans plusieurs domaines comme la reconnaissance d'image, la reconnaissance vocale, la traduction automatique, etc. Les réseaux de neurones artificiels sont des modèles mathématiques qui sont inspirés du fonctionnement du cerveau humain. Ils sont composés de plusieurs couches de neurones qui sont connectés entre eux. Chaque neurone est composé d'une fonction d'activation qui permet de calculer la sortie du neurone en fonction de ses entrées. Il existe plusieurs types d'architecture de réseaux de neurones artificiels. Ses architecture ont été développées pour répondre à des problèmes spécifiques tels que l'architecture de convolution pour la reconnaissance d'image, les réseaux de neurones récurrents pour la reconnaissance vocale, etc. Dans ce papier, nous nous intéressons à l'architecture de convolution car elle permet de traiter un large ensemble de problèmes comme la reconnaissance d'image, la reconnaissance de texte, la reconnaissance de séquence, etc. Par ailleurs l'architecture de convolution fait partie des architectures feed-forward qui sont une classe de réseaux de neurones artificiels qui sont composés de plusieurs couches de neurones et qui ne contiennent pas de cycles. Les réseaux de neurones feed-forward sont les plus utilisés dans l'apprentissage profond.

A. Convolutional Neural Network

Nous parlerons dans ce papier de programmes d'apprentissage profond pour désigner les programmes contenant les réseaux de neurones convolutifs. Une architecture de réseaux de neurones est de type feed-forward car elle ne contient pas de cycles. elle est composée de trois principales étapes suivantes:

Convolution: La convolution est l'action de faire passer les filtres sur une image pour extraire des patterns. Un filtre est un réseau de neurone classique qui scanne une image sous partie par sous partie afin de détecter un pattern. Le filtre scanne une fenetre (partie) de l'image puis calcule la sortie en un pixel. Ainsi le filtre navige sur toute l'image pour calculer la sortie de chaque pixel d'une nouvelle image. La nouvelle image est appelée feature map. La convolution augmente la profondeur de l'image en entrée car elle génère plusieurs feature maps tout en réduisant sa taille. On rappelle que chaque layer de la profondeur est une image (taille réduite) générée par un filtre.

Pooling: Le pooling est une opération qui permet de réduire la taille de l'image en selectionant le pixel le plus grand dans une fenetre. Il permet de réduire la consommation en calculs et en mémoire. Il permet aussi de réduire le sur l'overtfitting sur les données d'entrainement car on pert l'information sur la position des patterns détectés. On va donc utiliser le pooling sur les feature maps générées par la convolution. Chaque convolution est suivie d'un pooling et on répète cette opération plusieurs fois. Plus on va en profondeur de la convolution, plus on détecte des patterns précis susceptible de faciliter la classification.

Flatten: Le flatten est l'opération qui permet de transformer les feature maps générées à la suite des convolution et pooling en un seul vecteur. Ce vecteur est ensuite utilisé comme entrée d'un réseau de neurones classique (*dense layer*). De ce neurone sortira l'output du réseau de neurones convolutifs.

B. Odeur de conception

Le code smell est un mauvais choix de conception et d'implementation qui peut avoir un impact négatif sur la qualité du logiciel [1]. Tout comme n'importe quel programme, les programmes d'apprentissage profond peuvent contenir des odeurs de code. Nous nous intéressons dans ce papier aux odeurs de conception car ce sont des odeurs introduits tôt dans le cycle de développement du logiciel et elles peuvent avoir un impact négatif conséquent sur la performance et la qualité du logiciel. Ses odeurs sont introduites par le développeur lors de la phase de conception du logiciel.
Nikanjam et al. ont proposé une liste de 8 odeurs de conception pour les programmes d'apprentissage feed-forward [**?**]. Ils ont trouvé ses odeurs dans des revue de litterature et les plateformes open source Github et StackOverflow. Leur étude empirique a montré que les odeurs proposées sont perçu comme pertinent par les développeurs. Nous nous sommes de se fait basé sur cette liste d'odeurs pour proposer aux développeurs un systeme de détection automatique de ses odeurs.
Les 8 odeurs proposées sont classées en deux catégories principales comme suit:

1) **Odeurs de conception lors de la convolution et le pooling** : Ces odeurs sont introduites lors de la formation des feature maps. Elles sont liées à la taille de la feature map ou du filtre et à la disposition des couhes. La liste des odeurs est présentée dans le tableau I.
2) **Odeurs de conception liées à l'usage des méthodes de régulation** : Ces odeurs sont introduites lors de l'utilisation de la régularisation. Elles sont liées à l'utilisation des méthodes de régularisation comme la couche dropout et la disposition de ses méthodes. La liste des odeurs est présentée dans le tableau II.

C. Meta-modeling

Thomas Kühne définit un modèle comme un artefact formulé dans un langage de modélisation, tel qu'UML, décrivant un système à l'aide de différents types de diagrammes [2]. Cette représentation abstraite d'un système est définit par un autre modèle appelé méta-modèle. Un modèle permet d'analyser, comprendre et transformer un système efficacement. Ses opérations peuvent être automatisées [3]. Il existe plusieurs types de modèles tels que les modèles de comportement (représent le comportement dynamique du système) ou de conception (représent la structure statique du systeme).
Dans ce papier, nous nous intéressons aux modèles de conception car les odeurs étudiées sont représentées sous forme statique et sont

Table I: *Odeurs de conception dans les couches de convolution et de pooling*

Design Smell	Description
Non-expanding feature map	Conserver le même nombre de caractéristiques ou le diminuer à mesure que l'architecture devient plus profonde.
Losing local correlation	Commencer par une taille de fenêtre relativement grande pour le filtrage spatial et la conservent pour toutes les couches convolutives.
Heterogeneous blocks of CNNs	Construire un modèle plus profond en empilant uniquement un ensemble de couches de convolution et de mise en commun sans configuration appropriée.
Too much down-sampling	Utilisation du pooling juste après chaque couche convolutive, en particulier pour les premières couches.
Non-dominating down-sampling	Utilisation de l'average-pooling.

Table II: *Odeurs de conception liées à la régularisation*

Design Smell	Description
Useless Dropout	Utilisation du Dropout avant les pooling layer.
Bias with Batchnorm	Conserver les valeurs de biais dans les couches lors de l'utilisation de batchnorm. Les couches d'apprentissage d'un FNN bénéficient d'un biais avec différentes initialisations.
Non-representative Statistics Estimation	Utilisation du batchnorm après le dropout.

introduit au niveau de la stucture du système.

Nous utiliserons par ailleurs les modèles **FAST** (*Famix Abstract Syntax Tree*) pour représenter les systèmes étudiés dans ce papier. Les modèles FAST sont des arbres de syntaxe issue du langage de programmation orienté objet Pharo [4], [5], [6]. Il hérite du modèle **FAMIX** qui est d'après la définition de Tichelaar et al. une représentation indépendante du langage d'un code source orienté objet. Il s'agit d'un modèle entité-relation qui modélise le code source orienté objet au niveau de l'entité du programme [7] [8]. Les modèles FAST sont développés dans l'envirennement de retroengineering Moose [9].

D. Motivation and research questions

Les développeurs ont identifié dans l'étude de Nikanjam et al. les odeurs de conception proposées, leur pertinence et donc l'interet de les éviter. Cependant, il n'existe pas de système de détection automatique de ses odeurs dans la littérature. Nous proposons dans ce papier ce système qui permettra de détecter ses odeurs dans les programmes d'apprentissage profond CNN. De plus notre système couvrira les Frameworks open source les plus utilisés sur le marché. Pour ce faire, nous avons défini les questions de recherche suivantes:

RQ1: Peut-on détecter les odeurs de conception dans les programmes d'apprentissage profond CNN avec la méta-modélisation?

RQ2: Quels sont les odeurs de conception les plus répandu dans les programmes d'apprentissage profond CNN?

RQ3: Exist-il des lien entre les odeurs de conception dans les programmes?

Notre papier est organisé comme suit: la section II présente les travaux passé lié à notre sujet. La section III décrit la collection et le traitement des données. La section IV présente les résultats. La section V discute les résultats précedement présentées. La section VI présente les menaces à la validité et la section VII présente les limites de notre étude et les travaux futurs.

II. BACKGROUND

Dans cette section nous présenterons les précédents travaux sur la détection des ordeurs de code dans les logiciels.

L'idée de détecter des ordeurs dans les codes sources des logiciels est un sujet beaucoup étudié dans la littérature. En effet, plusieurs travaux proposent des techniques de détection depuis plus d'une décénie. Les odeurs de code sont étudiées selon plusieurs axes, comme par exemple leur présence dans des logiciels développés dans un langage de programmation spécifique [10], la sécurité dans le code source d'Infrastructure as Code [11], dans les microservices [12], ou leur impact sur la performance des applications Android [13]. Fard et al. [14] propsent JSNOSE (precision: 93% and average recall: 98%) une technique de détection des ordeurs de code du langage JavaScript. Une combinaise d'analyses statique et dynamique pour détecter des odeurs de code coté client. Chen et al. [15] proposent quant à eux Pysmell, un outil capable de détecter des odeurs de code du langage Python avec une précision moyen de 97.7%. La détection d'odeurs de code a un impact sur la qualité du logiciel, Van Emden et al. [16] font de l'assurance qualité de code Java via la détection et la visualisation d'odeurs de code. Pour ne citer que ces exemples, Moha et al. [17] proposent une approche de spéccification et de détection d'odeurs de code nommée DECOR avec une precision de 60.5%, et un recall de 100%. On trouve également dans la litterature des approches de détection d'odeurs basées sur la modélisation, comme l'approche proposée par Araujo et al. [18] pour identifier les odeurs organisationnelles dans les système multi-agents via la métamodélisation, ou celle de Khomh et al. [19] qui convertissent les règles de détection dans la litterature en un modèle probabilistique dans le but de détecter les odeurs de code. Ce dernier ouvre la voie à la détection d'odeurs de code avec l'aide de l'intelligence artificielle. On trouve de ce fait, dans les dernières années plusieurs travaux proposant des approches de détection d'odeurs de code basées sur l'intelligence artificielle [20], [21], [22]. À la lumière de ces travaux, nous nous sommes posés les questions suivantes: **Est-il possible de détecter efficacement et de manière statique les odeurs de code dans les logiciels d'apprentissage profond?**
Dans la litterature Nikanjam et al. ont initer la réponse à ces questions en proposant NeuraLint (precision: 100% and recall: 70,5%) [23], une approche basées sur modèle de détection d'odeurs de code. Cette approche utilise la métamodelisation pour atteindre son but.

Cependant la porté de NeuraLint ne s'est limité qu'aux architectures feedforward multilayer perceptron (MLP). L'apprentissage profond est un domaine en pleine expansion, et les architectures de réseaux de neurones sont de plus en plus complexes. Il est donc nécessaire de pouvoir détecter des odeurs de code d'autres architectures de réseaux de neurones. L'architecture CNN étant une des plus utilisée dans la pratique, nous nous sommes fixés dans ce papier, l'objectif de permettre la détection d'odeurs de code dans les réseaux de neurones de type CNN.

III. Study Design

In this sub-section, we describe the details of the data collection and processing approach followed to answer our different research questions.

A. Data Collection

Le développement du système de détection d'odeurs de conception s'est fait à partir des exemples de code de programmes d'apprentissage profond. Ces exemples représentent des programmes d'apprentissage profond dans lesquels on y trouve des odeurs de conceptions. Ses exemples ont été collectés par Nikanjam et al. [24] dans le cadre de leur étude empirique sur les odeurs de conception dans les programmes d'apprentissage profond. En effet, dans cette étude, ils ont présenté des exemples de programmes d'apprentissage profond contenant les ordeurs de conceptions énumérées dans l'introduction I-B. Ces exemples de code source ont été collectés à partir des plateformes StackOverflow et Github.

Le système de détection d'odeurs de conception ainsi développé a servi sur un ensemble de programmes d'apprentissage profond collectés à partir de la plateforme Github. Ces programmes d'apprentissage profond ont été collectés selon un processus bien défini. En effet, nous avons utilisé l'API de Github et plus précisement les requêtes de recherche de type *search code*. Cette requête permet de rechercher des fichiers dans les dépôts Github contenant des mots clés spécifiques définis dans notre système. Étant donné que notre papier se concentre uniquement sur les libraires *Keras*, *Tensorflow* et *Pytorch*, nous avons utilisé les mots clés présentés dans le tableau III. Ces mots clés représentent les modules et les fonctions de ces trois librairies.

Table III: *Liste des mots clés utilisés pour la recherche de programmes d'apprentissage profond dans les dépôts Github.*

Mots clés	**Librairie**
keras.layers	Keras
keras.layers.convolutional	Keras
AveragePooling2D	Keras/Tensorflow
MaxPooling2D	Keras/Tensorflow
tensorflow.keras.layers	Tensorflow
Conv2D	Keras/Tensorflow
Convolution2D	Keras/Tensorflow
BatchNormalization	Keras/Tensorflow
import torch	Pytorch
import torchvision.models	Pytorch
torch.nn.Sequential	Pytorch
torch.nn.Conv2d	Pytorch
torch.nn.BatchNorm2d	Pytorch
torch.nn.MaxPool2d	Pytorch

La requête de recherche de type *search code* retourne un ensemble d'informations sur le repository et le fichier contenant les mots clés recherchés. Cette collecte retourne un ensemble de données brute de 9572 repositories Github diférents. Nous sélectionnons ensuite aléatoirement un sous ensemble de 10% de notre ensemble de données brutes, soit 958 repositories Github afin d'optimiser l'étape de filtrage manuel ci-après.

À partir de ce sous ensemble, nous procédons ensuite à un filtrage des répertoires selon les critères suivants: (1) nombre de commit ($\geq$ 100), (2) nombre d'étoile ($\geq$ 100), (3) dernière date du commit($\leq$ 5 years), (4) nombre de contributeurs ($\geq$ 5). Ce filtrage nous permet de ne garder que les répertoires qui sont les plus populaires et qui sont les plus actifs. Nous avons ensuite procédé à un filtrage manuel des répertoires en éliminant les projets qui ne sont pas des programmes d'apprentissage profond ou ne pas pertinente pour notre étude (n'utilise pas les librairies choisies, ou n'implémente pas de réseau de neurones). De ce fait nous nous sommes retrouvé avec 500 projets sur lesquels nous allons appliquer notre système de détection d'odeurs de conception.
Le schema **??** présente le processus de collecte des programmes d'apprentissage profond.

B. Data Processing

Le processus de collecte des programmes d'apprentissage profond nous a permis d'avoir un ensemble de programmes d'apprentissage profond sur lesquels nous allons appliquer notre système de détection d'odeurs de conception. Cependant, il est important de noter que les programmes d'apprentissage profond sont dans notre cas des programmes uniquement écrits en langage Python. Il est donc essentiel de noter que, le système de détection d'odeurs de conception développé dans ce papier est un système qui est uniquement capable de détecter les odeurs de conception dans les programmes d'apprentissage profond écrits en langage Python.

Afin de répondre à la question de recherche *RQ1*, nous avons développé un système de détection d'odeurs de conception dans les programmes d'apprentissage profond. Ce système se décompose en trois parties. La première partie est la partie relative à la modélisation des programmes d'apprentissage profond. La deuxième partie est celle relative à la détection des odeurs de conception dans les programmes d'apprentissage profond. Et la troisième est la partie relative à l'analyse des résultats de la détection.

1) Modélisation des programmes d'apprentissage profond: La modélisation des programmes d'apprentissage profond est la première étape du processus de détection des odeurs de conception. Cette étape consiste à transformer le code source des programmes d'apprentissage profond collecté en un modèle intermédiaire qui sera utilisé pour la détection des odeurs de conception. Ce modèle intermédiaire est un modèle de type *Famix Abstract Syntax Tree* (FAST). Le modèle FAST est un modèle qui est utilisé pour représenter les programmes sous forme d'arbre syntaxique. Il hérite de la représentation abstraite de codes sources *Famix* dévéloppé dans le langage de programmation *Pharo*. Avec *Famix*, le code source est représenté sous forme d'objet et selon un méta-modèle défini pour symplifier l'analyse et la représentation de programmes complexes. Un modèle *Famix* est une représentation agnostique en terme de langage du code source d'origine et embarque des fonctions qui permettent entre autres sa navigation et sa transformation.

La transformation du code source collecté en modèle FAST se fait comme suit. Chaque projet collecté est cloné dans un répertoire local. Son code source est ensuite parsé à l'aide de la librairie *ast.py* de Python. Cette librairie permet de transformer le code source en un arbre syntaxique sous forme json. Le json obtenu est ensuite transformé en un modèle FAST à l'aide d'un importateur développé sous le patron de conception visiteur. Les visiteurs sont des fonctions qui permettent de parcourir un arbre syntaxique pour effectuer sur ses noeuds des opérations spécifiques. Dans notre cas, nous avons développé des visiteurs qui nous permettent de parcourir l'arbre syntaxique (json) et de transformer chaque noeud de cet arbre en

une entité de FAST.

Le schema **??** illustre le processus de modélisation des programmes d'apprentissage profond.

2) Détection des odeurs de conception: Le modèle FAST obtenu à partir de la méta-modélisation des programmes d'apprentissage profond est ensuite utilisé pour la détection des odeurs de conception. Cette étape consiste à parcourir le modèle FAST et à appliquer les règles de détection des odeurs de conception. Ces règles sont définies dans le langage de programmation *Pharo*. Une règle ou plusieurs règles peuvent être appliquées sur un noeud de l'arbre syntaxique afin de détecter une odeur de conception.

Le schema **??** illustre le processus de détection des odeurs de conception.

3) Analyse des résultats de la détection des odeurs de conception: Les résultats de la détection des odeurs de conception de chaque repository sont stockés sous forme *csv*. Pour chaque repository nous enregistrons le nom de chaque odeur de conception détectée, le nombre de fois que cette odeur est détectée dans un fichier du repository et le chemin de ce fichier. Ces résultats permettent de répondre aux questions de recherche *RQ2* et *RQ3*.
On calcul ensuite la répartition des odeurs de conception dans chaque repository en divisant le nombre de fois qu'une odeur de conception est détectée par le nombre total d'odeurs de conception détectées dans le repository. Cette répartition permet de répondre à la question de recherche *RQ2*. On calcul également la répartition des odeurs de conception dans l'ensemble des 500 repositories en divisant le nombre de fois qu'une odeur de conception est détectée par le nombre total d'odeurs de conception détectées dans l'ensemble des 500 repositories. Cette répartition permet aussi de répondre à la question de recherche *RQ2*.

On calcul ensuite la matrice de corrélation entre les odeurs de conception dans l'ensemble des 500 repositories. Cette matrice permet de répondre à la question de recherche *RQ3*. Une matrice de corrélation permet de calculer la corrélation entre deux variables.

Un coeficient de corrélation est calculé à l'aide de la formule suivante: $r = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$ où r est le coeficient de corrélation, $cov(X,Y)$ est la covariance entre les deux variables X et Y, σ_X est l'écart type de la variable X et σ_Y est l'écart type de la variable Y. Ce coeficient de corrélation est compris entre -1 et 1. On parle de correlation positive lorsque le coeficient de corrélation est compris entre 0 et 1. Dans ce cas, plus il est proche de 1, plus les deux variables sont corrélées positivement, et plus il est proche de 0, moins les deux variables sont corrélées. On parle de correlation négatif lorsque le coeficient de corrélation est compris entre -1 et 0. Dans ce cas, plus le coeficient de corrélation est proche de -1, plus les deux variables sont corrélées négativement, et plus il est proche de 0, moins les deux variables sont corrélées.

Dans notre cas, les variables sont les odeurs de conception. La corrélation entre deux odeurs de conception est calculée en divisant le nombre de fois que les deux odeurs de conception sont détectées ensemble par le nombre total de fois que les deux odeurs de conception sont détectées. Cette corrélation est calculée pour chaque paire d'odeurs de conception. Plus elle est proche de 1, plus les deux odeurs de conception sont corrélées, et plus elle est proche de 0, moins les deux odeurs de conception sont corrélées.

Une corrélation positive signifie que les deux odeurs de conception sont souvent détectées ensemble. Et une corrélation négative signifie que les deux odeurs de conception sont rarement détectées ensemble. Enfin une corrélation nulle signifie que les deux odeurs de conception ne sont jamais détectées ensemble.

Le schema **??** illustre le processus de cette dernière partie.

C. Replication Package

Le code source du parse de code source en arbre de Syntax *json* est sur Github à l'adresse https://github.com/aurpur/parserPythonToJson et celui du système de détection d'odeurs est disponible sur Github à l'adresse https://github.com/aurpur/famixPythonImporter. Le système de collection de données est aussi disponible sur Github à l'adresse https://github.com/aurpur/ms-github-data-collection.

IV. Case Study Results

Dans cette section nous présentons les résultats de notre étude. Nous présentons d'abord les résultats relative à la détection des odeurs de conception dans les programmes étudiés. Puis, nous présentons les résultats relative de l'analyse des résultats. **TODO: To be completed**

A. Détection des odeurs de conception

Ici nous présenterons les chiffres sur la détection des odeurs de conception dans l'ensemble des repositories tel que la répartition des odeurs de conception dans le dataset (Le nombre de repositories par odeurs de conception), ou la profondeur (Nombre d'occurence d'une odeur de conception dans un repository).

Cela soutiendra l'hypothèse selon laquelle les odeurs de conception peuvent être détectées sur les programmes CNN à travers un model FAST dans Pharo. Et de ce fait, on peut donc détecter les odeurs de conception à l'aide de la modélisation.

B. Analyse des résultats de la détection

Ici nous présenterons les chiffres sur l'analyse des résultats de la détection des odeurs de conception dans les programmes étudiés. Le but est de présenter le classement des odeurs de conception, et de présenter les chiffres sur les possibles correlation entre elles.

Cela soutiendra l'hypothèse selon laquelle il y a des odeurs de conception qui sont plus répandu que d'autres dans les programmes CNN (dans le référentiel de notre étude) et que certaines odeurs de conception présentent des correlations entre elles.

V. Discussion

Dans cette section, nous discutons les résultats de notre étude et répondons à nos questions de recherche. Nous discutons également des conséquences de nos résultats dans le contexte de la recherche et du développement de logiciels. **TODO: To be completed**

A. RQ1: Peut-on détecter les odeurs de conception dans les programmes d'apprentissage profond CNN avec la méta-modélisation?

Ici nous discuterons des résultats de notre première question de recherche. Nous discuterons de la technique de détection des odeurs de conception proposée et de sa performance. Nous calculerons l'exactitude, la précision et le rappel de la technique proposée. On discutera également du temps d'exécution du système proposée.

Nous parlerons de l'objectif derrière la question (Motivation), rappelerons la méthode utilisée et les trouvailles.

B. RQ2: Quels sont les odeurs de conception les plus répandu dans les programmes d'apprentissage profond CNN?

Ici nous discuterons des résultats de notre deuxième question de recherche. Nous discuterons de la répartition et la profondeur des odeurs de conception dans l'ensemble de données.

Nous parlerons de l'objectif derrière la question (Motivation), rappelerons la méthode utilisée et les trouvailles.

C. RQ3: Exist-il des lien entre les odeurs de conception dans les programmes?

Ici nous discuterons des relations entre les odeurs de conception dans les programmes d'apprentissage profond.

Nous parlerons de l'objectif derrière la question (Motivation), rappelerons la méthode utilisée et les trouvailles.

D. Implications

Ici nous parlerons de l'impact de notre étude sur la recherche et le développement de logiciels. Nous parlerons aussi de l'impact de nos résultats.

VI. Threats to validity

Notre étude est sujet de plusieurs menaces à la validité. Dans cette section nous présonterons ces menaces et les stratégies que nous avons utilisé pour les mitiger.

Premièrement, notre étude est sujet de la menace de validité de construction. En effet le modèle FAST généré par notre approche pourrait ne pas être correct et conforme au métamodèle défini ou encore mal représenter le programme d'apprentissage profond en entrée. Pour mitiger cette menace, nous avons mis en place des tests unitaires pour chaque méthode visiteur. Ces tests unitaires nous permettent de vérifier que les méthodes implémentées dans l'importateur son correctement implémentées et que le modèle FAST généré est conforme au métamodèle défini. De plus, nous avons implémenté pour un ensemble de données de test (des exemples synthétiques) des tests fonctionnels (Smock Testing) pour vérifier que le modèle FAST généré représente bien le programme d'apprentissage profond en entrée.

Deuxièmement, notre étude est sujet de la menace de validité interne. Il subsiste un risque que les conclusions obtenues ne soient pas causées par le code source en entrée. Pour mitiger cette menace, nous avons en plus de tester notre système sur des exemples synthétiques, nous avons utilisé des projets réels collectés dans Github pour valider le fonctionnement de notre système.

Troisièmement, notre étude est sujet de la menace de validité externe. Il y a un risque que les résultats obtenus ne soient pas généralisables. Pour mitiger cette menace, nous avons collecté aléatoirement des projets dans Github afin de représenter différente application de l'architecture CNN.

Quatrièmement, notre étude est sujet de la menace de validité de fiabilité. Il y a enfin un risque que les résultats obtenus ne soient pas reproductibles. Pour mitiger cette menace, nous avons en plus d'utilisé des projets de différentes applications de l'architecture CNN, utilisé des projets implémentées dans les trois librairies open source les plus utilisées dans l'industrie (Tensorflow, Pytorch et Keras).

VII. Limitations and Future Work

1) Limitations: Notre étude s'est fait dans un contexte limité que nous allons décrire dans cette section. Quand nous parlons de programmes d'apprentissage profond, nous faisons référence aux programmes contenant un réseau de neurones à convolution (CNN). Il est donc important de noter qu'il existe d'autres types de programmes d'apprentissage profond qui ne sont pas couverts par notre étude. De plus, il existe plusieurs types de librairies qui permettent de créer des programmes d'apprentissage profond. Notre étude se limite aux programmes créés à partir des librairies TensorFlow, Keras et PyTorch. Et nos données proviennent exclusivement de référentiel open source Github ou StackOverflow (pour les exemples). Enfin, notre système de détection d'odeurs de conception est limité à la détection des 8 ordeurs de conception décrit plus haut.

2) Future Work: Les travaux futurs peuvent aller dans plusieurs directions. En effet il est possible d'appliquer notre approche à d'autres types de programmes d'apprentissage profond, d'autres librairies, d'autres odeurs de conception, d'autres langages de programmation et d'autres types de référentiels (non open source par exemple). Nous avons fait, une analyse statique des programmes d'apprentissage profond, il est également envisageable de faire une analyse dynamique en simulant l'exécution des programmes à travers de la modélisation. En plus de la détection via la modélisation, d'autres champ de recherche sont la détection via l'apprentissage automatique et la correction automatique des odeurs de conception. Ces derniers sont des sujets de recherche très prométeurs qui pourront avoir un impact très important sur la qualité des programmes d'apprentissage profond futur.

VIII. Conclusion

TODO: To be completed

References

[1] M. Fowler and K. Beck, "Refactoring: Improving the design of existing code," in *11th European Conference. Jyväskylä, Finland*, 1997.

[2] T. Kühne, "Matters of (meta-) modeling," *Software & Systems Modeling*, vol. 5, pp. 369–385, 2006.

[3] C. A. González and J. Cabot, "Formal verification of static software models in mde: A systematic review," *Information and Software Technology*, vol. 56, no. 8, pp. 821–838, 2014.

[4] A. P. Black, O. Nierstrasz, S. Ducasse, and D. Pollet, *Pharo by example*. Lulu. com, 2010.

[5] A. Bergel, D. Cassou, S. Ducasse, and J. Laval, *Deep Into Pharo*. Lulu. com, 2013.

[6] O. Zaitsev, S. Ducasse, and N. Anquetil, "Characterizing pharo code: A technical report," Ph.D. dissertation, Inria Lille Nord Europe-Laboratoire CRIStAL-Université de Lille; Arolla, 2020.

[7] S. Tichelaar, S. Ducasse, and S. Demeyer, "Famix and xmi," in *Proceedings Seventh Working Conference on Reverse Engineering*, 2000, pp. 296–298.

[8] S. Demeyer, S. Ducasse, and S. Tichelaar, "Why famix and not uml," in *Proceedings of UML'99*, vol. 1723, 1999.

[9] S. Ducasse, M. Lanza, and S. Tichelaar, "Moose: an extensible language-independent environment for reengineering object-oriented systems," in *Proceedings of the Second International Symposium on Constructing Software Engineering Tools (CoSET 2000)*, vol. 4, 2000.

[10] A. Saboury, P. Musavi, F. Khomh, and G. Antoniol, "An empirical study of code smells in javascript projects," in *2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER)*. IEEE, 2017, pp. 294–305.

[11] A. Rahman, M. R. Rahman, C. Parnin, and L. Williams, "Security smells in ansible and chef scripts: A replication study," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 30, no. 1, pp. 1–31, 2021.

[12] T. Cerny, A. Al Maruf, A. Janes, and D. Taibi, "Microservice anti-patterns and bad smells. how to classify, and how to detect them. a tertiary study," *How to Classify, and How to Detect Them. A Tertiary Study*.

[13] G. Hecht, N. Moha, and R. Rouvoy, "An empirical study of the performance impacts of android code smells," in *Proceedings of the international conference on mobile software engineering and systems*, 2016, pp. 59–69.

[14] A. M. Fard and A. Mesbah, "Jsnose: Detecting javascript code smells," in *2013 IEEE 13th International Working Conference on Source Code Analysis and Manipulation (SCAM)*, 2013, pp. 116–125.

[15] Z. Chen, L. Chen, W. Ma, and B. Xu, "Detecting code smells in python programs," in *2016 international conference on Software Analysis, Testing and Evolution (SATE)*. IEEE, 2016, pp. 18–23.

[16] E. van Emden and L. Moonen, "Java quality assurance by detecting code smells," in *Ninth Working Conference on Reverse Engineering, 2002. Proceedings.*, 2002, pp. 97–106.

[17] N. Moha, Y.-G. Gueheneuc, L. Duchien, and A.-F. Le Meur, "Decor: A method for the specification and detection of code and design smells," *IEEE Transactions on Software Engineering*, vol. 36, no. 1, pp. 20–36, 2010.

[18] P. Araujo, S. Rodríguez, and V. Hilaire, "A metamodeling approach for the identification of organizational smells in multi-agent systems: Application to aspecs," *Artificial Intelligence Review*, vol. 49, pp. 183–210, 2018.

[19] F. Khomh, S. Vaucher, Y.-G. Guéhéneuc, and H. Sahraoui, "A bayesian approach for the detection of code and design smells," in *2009 Ninth International Conference on Quality Software*. IEEE, 2009, pp. 305–314.

[20] S. Alawadi, K. Alkharabsheh, F. Alkhabbas, V. Kebande, F. M. Awaysheh, and F. Palomba, "Fedcsd: A federated learning based approach for code-smell detection," *arXiv preprint arXiv:2306.00038*, 2023.

[21] R. Sandouka and H. Aljamaan, "Python code smells detection using conventional machine learning models," *PeerJ Computer Science*, vol. 9, p. e1370, 2023.

[22] A. Alazba, H. Aljamaan, and M. Alshayeb, "Deep learning approaches for bad smell detection: a systematic literature review," *Empirical Software Engineering*, vol. 28, no. 3, p. 77, 2023.

[23] A. Nikanjam, H. B. Braiek, M. M. Morovati, and F. Khomh, "Automatic fault detection for deep learning programs using graph transformations," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 31, no. 1, pp. 1–27, 2021.

[24] A. Nikanjam and F. Khomh, "Design smells in deep learning programs: an empirical study," in *2021 IEEE International conference on software maintenance and evolution (ICSME)*. IEEE, 2021, pp. 332–342.