

# Classification des interactions des utilisateurs sur GitHub via le Natural Language Inference (NLI)

De l'analyse des interactions à la modélisation de la collaboration

## Parties prenantes



## Auteur

Johnny Li  
[johnny.li@imt-atlantique.net](mailto:johnny.li@imt-atlantique.net)

## Encadrants

Cécile Bothorel  
[cecile.bothorel@imt-atlantique.net](mailto:cecile.bothorel@imt-atlantique.net)

Lucas Ribeiro Marques  
[lucas.ribeiro-marques@imt-atlantique.fr](mailto:lucas.ribeiro-marques@imt-atlantique.fr)

## Sources

- M.Fazelnia, V.Koscinski, S.Hezog, M.Mirakhori, *Lessons from the Use of Natural Language Inference (NLI) in Requirements Engineering Tasks*
- P.A. Petrova, S.I. Markov & V.V. Kachanov, *Building a Dataset for Combined Classification of Source Code Reviews*
- M. Laurer, W. Atteveldt, A. Casas, K. Welbers, *Building Efficient Universal Classifiers with Natural Language Inference*
- A.K. Turzo, F. Faysal, O. Poddar, J.Sarker, A. Iqbal, A. Bosu, *Towards Automated Classification of Code Review Feedback to Support Analytics*

## Contexte :

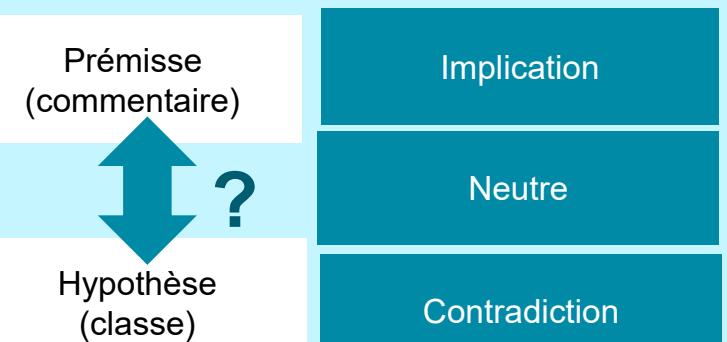
- Data2Laws** vise à étudier les mécanismes par lesquels des collectifs en ligne (communautés open source, plateformes collaboratives, etc.) passent à l'échelle et d'identifier les lois ou invariants sociotechniques qui permettent ce passage à l'échelle.
- La collaboration numérique est un objet d'étude central : massive, ouverte et distribuée, elle soulève de nombreux défis, tels que l'absence de hiérarchie formelle, les contributions asynchrones, les conflits implicites, etc.
- Les échanges entre utilisateur sur **GitHub** lors de la résolution des issues constituent une trace de cette collaboration numérique, où émerge des critiques, des débats, des solutions, des remerciements, etc.

## Axe d'étude :

Exploiter le **Natural Language Inference (NLI)** comme **outil de modélisation** de la collaboration numérique afin de caractériser l'évolution des **types d'interactions**, de mettre en évidence la **maintenance invisible**, d'analyser les mécanismes de **gestion des conflits** au sein des collectifs en ligne.

## Principe du NLI :

Tâche de traitement automatique du langage naturel (NLP) qui consiste à déterminer une **relation de bon sens** entre **deux propositions** :



Il s'agit une **tâche de classification** employé dans de nombreuses applications (question-réponse, résumé automatique, recherche d'information, etc.)

### Avantages :

- Classes définies
- Raisonnement sémantique
- Zéro-shot Learning
- Ajout de contexte via label verbalization

### Inconvénients :

- Sensible aux biais des données
- Capture difficile des relations implicites, complexes et ambiguës
- Sur-simplification de la compréhension

## Méthodes et Outils :

### Jeu de données :

10 045 commentaires de revue de code (80/20% pour le train/test split), annotés avec une taxonomie de **16 classes** puis répartis en **5 groupes** :

Code style	Discussion	Development
Style Naming	Question Response Convention Testing	Design Refactoring Functionality Roadmap Development
User Documentation Support Input/Output		Other
		Other

### Modèles :

- DeBERTa-v3 zéro-shot** : entraîné sur des tâches de NLI à partir de 33 jeux de données couvrant 387 classes
- CodeBERT** : entraîné sur des paires Natural Language/code et code seul pour la complétion de code, la génération de documentation, etc.

### Axes d'évaluation :

- Performances en **zéro-shot learning**
- Impact de la **verbalisation des labels**
- Performances après **fine-tuning**

## Résultats :

### Comparaison des métriques :

Model	Metric	Code style	Discussion	Development	User	Other	Avg
Zero-Shot	Prec.	75	46	52	0	31	41
	Rec.	33	5	96	0	16	30
	F1	46	9	68	0	21	29
Label Verb	Prec.	61	32	66	9	2	34
	Rec.	58	21	65	16	2	32
	F1	59	25	65	11	2	33
Fine-Tuning	Prec.	65	53	81	50	43	58
	Rec.	88	88	86	62	81	81
	F1	75	66	84	55	55	67
CodeBERT	Prec.	80	67	89	83	64	76
	Rec.	86	81	83	67	62	77
	F1	83	73	86	74	63	76

### Analyses :

- Les résultats **Zero-Shot** et **Label Verb** révèlent la **difficulté** du modèle à généraliser la sémantique complexe des intentions dans les revues de code.
- Le **Fine-Tuning** améliore les performances mais ne permet pas d'atteindre le niveau de performance de **CodeBERT**.
- la sémantique '**User**' est trop **implicite** ou **variée** pour être capturée efficacement par DeBERTa.