

Correction automatique de copies d'étudiants au moyen d'attention croisée

TALN 2022 - Équipe LIA x LS2N

Yanis LABRAK - LIA

`yanis.labrak@univ-avignon.fr`

Philippe TURCOTTE - LIA

`philippe.turcotte@alumni.univ-avignon.fr`

Richard DUFOUR - LS2N

`richard.dufour@univ-nantes.fr`

Mickael ROUVIER - LIA

`mickael.rouvier@univ-avignon.fr`

Laboratoire informatique d'Avignon (LIA)

Laboratoire des Sciences du Numérique de Nantes (LS2N)

June 28, 2022



Sommaire

- 1 Introduction
 - Tâche
 - Corpus
 - Systemes existants
- 2 Systemes proposes
 - Vecteurs de similarites
 - Attention croisee bidirectionnelle
 - Fine-Tuning CamemBERT
- 3 Architecture de fusion
 - Systemes
 - Resultats finaux
- 4 Perspectives
- 5 References

Présentation

Contexte :

- DÉfi Fouille de Textes (DEFT) [GI22] est une campagne d'évaluation annuelle francophone.
- Deux tâches d'évaluation automatique de copies d'après une référence existante :
 - **Tâche de base** avec un corpus fini (*Traitée*).
 - **Tâche continue** où l'ont requête une base de données afin d'effectuer des prédictions itératives (*Non traitée*).

Objectifs :

- **Prédire** la note d'une réponse d'un étudiant à une question (comprise entre 0.0 et 1.0) à partir d'éléments textuels (question, réponse étudiante, réponse enseignant).
- Problème de **régression** ou de **classification**.

Corpus

Exemples de question et corrections

Id	Question	Correction (+ suggestion de notation)
1021	<p><p>Quelle fonction PHP permet d'exécuter une requête sur une base de données PostgreSQL ?
</p></p>	<p><p>pg_query</p><p>
</p><p>pas tenu compte des arguments dans la notation
</p></p>
2005	<p>À quoi sert la balise &lt;span> ?
</p>	<p>Elle permet de définir une partie de texte à laquelle on pourra par exemple appliquer un style ou un comportement particulier.
0.5 pour "La balise &lt;span> sert à définir une partie du texte. ;" 0.5 aussi si parle de partie du document et pas du texte (si pas clair que inline et pas block)
</p>
2046	<p>Quel problème d'accessibilité peut se poser lorsqu'un utilisateur agrandit la taille des caractères du texte ?
</p>	<p><p>perte de contenu, par exemple recouvrement de contenus (violation du principe 1 de wcag ie perceptible)
</p></p>

Table 1: Exemples tirés du corpus DEFT 2022 de questions et leurs corrections par des enseignants.

Corpus

Exemples de réponses et notes associées

Id	Note	Étudiant	Réponse de l'étudiant
1021	1.0	student122	<code>\$reponse=pg_query(\$connection, \$requete);! Il faut préalablement avoir effectué la connection avec la base : ! \$connection=pg_connect (\$nombase);!!</code>
2005	0.0	student10	la balise <code></code> sert a créer des sections.
2046	0.5	student102	Certains caractères peuvent dépasser du cadre de la page web.
2040	0.8	student111	<code><!ELEMENT ue (#PCDATA)> <!-toujours pareil, ici on pourrait avoir autre chose que #PCDATA->! <!ATTLIST ue id ID #REQUIRED>! <!ATTLIST ue lang CDATA></code>
2042	0.3	student92	<code><code>! <enseignant>! <nom> Snow</nom>! <prenom>John</prenom>! <bureau>IS</bureau>! <statut>dead</statut>! </enseignant>! </code></code>

Table 2: Exemples tirés du corpus DEFT 2022 de réponses d'étudiants avec leurs notes associées.

- Les données sont extraites de questionnaires électroniques de type **Moodle**.
- Énoncés en **informatique**, plus particulièrement de **programmation web**.
- **Découpage** de l'entraînement en deux sous-ensembles : entraînement (85 %) et validation (15 %).

Corpus

Distribution des classes

Pour modéliser cette tâche, nous avons opté pour une méthode de classification sur 11 classes, de 0.0 à 1.0 avec un pas de 0.1.

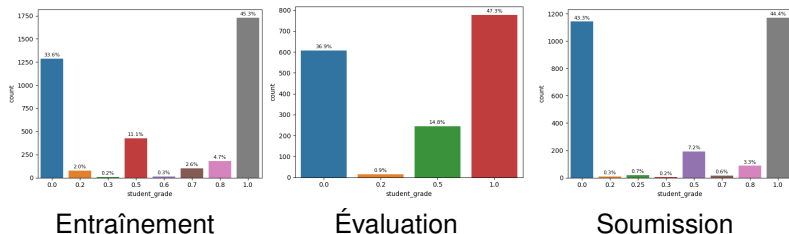


Figure 1: Distribution des notes sur les différents corpus.

Traitements textuels :

- 1 Transformation de la casse en minuscule.
- 2 Suppression des espaces en début et/ou fin de texte ainsi que les espaces successifs.
- 3 Retrait des caractères spéciaux `\n` et `\t`.
- 4 Normalisation des suggestions de notation avec l'ajout de "`__PTS__`" pour chaque note.

Traitements du code :

- 1 Suppression des balises HTML superflues liés à la mise en forme *Moodle* (`<p>` ou `
`).
- 2 Décodage des entités HTML, telles que "` `", "`<`" ou "`>`".
- 3 Ajout d'un espace avant et après toute ponctuation ou tous symboles à l'exception des nombres décimaux. (e.g : "`< a >`")

Méthodes explorées pour DEFT 2021

Représentations textuelles :

- TF-IDF
- Bags-of-words
- Word2Vec [Mik+13]
- FastText [Boj+16]
- CamemBERT [Mar+20]
- SentenceBERT [RG19]

Classification :

- Par seuils
- SVM
- Random Forest

Mesures de similarités :

- Cosine
- Jaro-Wikler
- Damerau-Levenshtein
- Monge-Elkan
- Euclidienne
- Recouvrement
- Sorensen-Dice
- Ratcliff-Obershelp
- Burrows-Wheeler
- *Soft Cardinalité* [JGG15]

Systemes proposés

Les trois systemes :

- 1 Vecteurs de similarités
- 2 Attention croisée bidirectionnelle
- 3 Fine-tuning CamemBERT

Systèmes proposés

Évaluation des combinaisons

Comparaison des performances de chaque combinaison en utilisant notre classifieur basé sur les plongements de mots CamemBERT [Mar+20] et un MLP sur le corpus de validation :

Combinaison	Précision	Rappel	F1-score
$R_{\text{etudiant}} + R_{\text{enseignant}} + Q$	59,26%	60,77%	59,94%
$R_{\text{etudiant}} + R_{\text{enseignant}}$	66,08%	65,57%	65,30%
$R_{\text{etudiant}} + Q$	52,24%	46,35%	45,53%

Table 3: Où R_{etudiant} est la réponse de l'étudiant, $R_{\text{enseignant}}$ est la réponse de l'enseignant et Q est le texte de la question.

La métrique principale utilisée pour l'évaluation et le classement est ici la *précision*.

Systemes proposés

Vecteurs de similarités (Baseline)

Étapes :

- 1 **Extraction** des plongements de mots CamemBERT
- 2 Transformation en vecteurs unidimensionnelles grâce à un **mean-pooling**
- 3 Construction des **vecteurs de distances de similarités** basé sur les 11 méthodes de calculs de distances suivantes :
 - a) Cosine
 - b) Euclidean
 - c) Braycurtis
 - d) Canberra
 - e) Chebyshev
 - f) Cityblock
 - g) Correlation
 - h) Minkowski
 - i) Squeclidean
 - j) Jensenshannon
- 4 **Classification** en l'une des onze classes grâce à un perceptron multicouche (MLP)

Systemes proposés

Vecteurs de similarités - Résultats

Classe	Précision	Rappel	F1-Score
0,00	58,32 %	85,51 %	69,34 %
0,10	0,00 %	0,00 %	0,00 %
0,20	0,00 %	0,00 %	0,00 %
0,30	0,00 %	0,00 %	0,00 %
0,40	0,00 %	0,00 %	0,00 %
0,50	0,00 %	0,00 %	0,00 %
0,60	0,00 %	0,00 %	0,00 %
0,70	0,00 %	0,00 %	0,00 %
0,80	0,00 %	0,00 %	0,00 %
0,90	0,00 %	0,00 %	0,00 %
1,00	95,89 %	60,65 %	74,30 %
Précision			66,91 %
Macro Moyenne	39,55 %	36,54 %	35,91 %
Moyenne pondéré	86,43 %	66,91 %	73,05 %

Table 4: Résultats des vecteurs de similarités sur le corpus de validation.

Systèmes proposés

Attention croisée bidirectionnelle

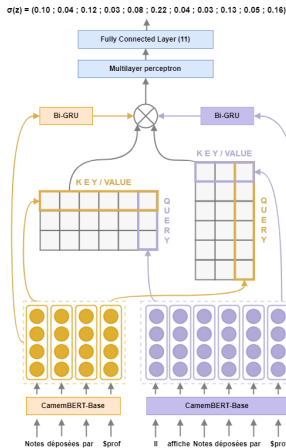


Figure 2: Attention [Vas+17] croisée bidirectionnelle + Bi-GRU.

Systemes proposés

Attention croisée bidirectionnelle - Résultats

Classe	Précision	Rappel	F1-Score
0,00	58,15 %	64,30 %	61,07 %
0,10	0,00 %	0,00 %	0,00 %
0,20	0,00 %	0,00 %	0,00 %
0,30	0,00 %	0,00 %	0,00 %
0,40	0,00 %	0,00 %	0,00 %
0,50	4,10 %	12,20 %	6,13 %
0,60	0,00 %	0,00 %	0,00 %
0,70	0,00 %	0,00 %	0,00 %
0,80	0,00 %	0,00 %	0,00 %
0,90	0,00 %	0,00 %	0,00 %
1,00	61,70 %	55,05 %	58,18 %
Précision			51.28 %
Macro Moyenne	17.71 %	18.79 %	17.91 %
Moyenne pondéré	52.35 %	51.28 %	51.56 %

Table 5: Résultats attention croisée sur le corpus de validation.

Systemes proposes

Fine-Tuning CamemBERT

$\sigma(\mathbf{z}) = (0.10 ; 0.04 ; 0.12 ; 0.03 ; 0.08 ; 0.22 ; 0.04 ; 0.03 ; 0.13 ; 0.05 ; 0.16)$

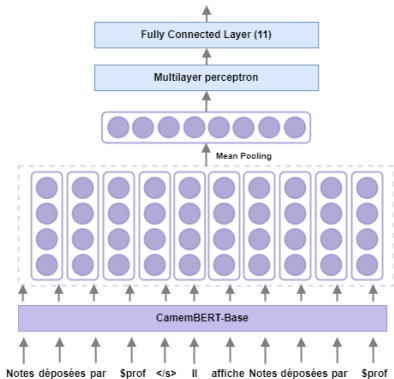


Figure 3: Architecture Fine-Tuning classifieur CamemBERT

Systemes proposes

Fine-Tuning CamemBERT - Resultats

Classe	Précision	Rappel	F1-Score
0.00	76.57 %	76.44 %	76.50 %
0.10	0.00 %	0.00 %	0.00 %
0.20	0.00 %	0.00 %	0.00 %
0.30	0.00 %	0.00 %	0.00 %
0.40	0.00 %	0.00 %	0.00 %
0.50	33.95 %	22.54 %	27.09 %
0.60	0.00 %	0.00 %	0.00 %
0.70	0.00 %	0.00 %	0.00 %
0.80	0.00 %	0.00 %	0.00 %
0.90	0.00 %	0.00 %	0.00 %
1.00	74.97 %	84.32%	79.37 %
Précision			71,47 %
Macro Moyenne	46,37 %	45,83 %	45,74 %
Moyenne pondéré	68,79 %	71,47 %	69,83 %

Table 6: Resultats Fine-Tuning CamemBERT sur le corpus de validation.

Architecture de fusion

Les trois systèmes :

- 1 Vecteurs de similarités
- 2 Attention croisée bidirectionnelle
- 3 Fine-tuning CamemBERT

Architecture de fusion des systèmes :

- 1 Run 1 : Vecteurs de similarités
- 2 Run 2 : CamemBERT + Attention croisée bidirectionnelle
- 3 Run 3 : Vecteurs de similarités + Attention croisée bidirectionnelle

Architecture de fusion

Vue d'ensemble

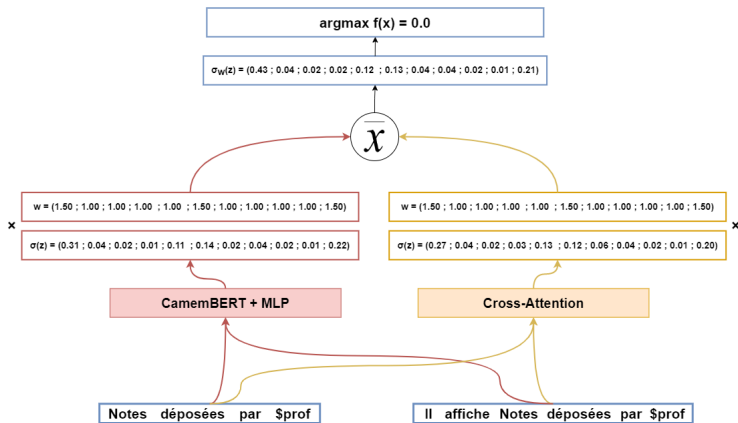


Figure 4: Architecture de fusion.

Architecture de fusion

Résultats

Nous avons proposé **une baseline** et **deux systèmes** issus de la fusion :

Méthode	Validation				Évaluation
	Taux de succès	Précision	Rappel	F1-Mesure	Précision
Best DEFT-2021 [Sui+21]	-	-	-	-	68,20 %
Best DEFT-2022	-	-	-	-	75,60 %
Run 1 : Similarities Features	65,63 %	86,63 %	65,63 %	71,80 %	60,60 %
Run 2 : CamemBERT + Attention croisée bidirectionnelle	71,90 %	73,72 %	71,90 %	72,69 %	72,60 %^{1*}
Run 3 : Similarities Features + Attention croisée bidirectionnelle	66,06 %	87,48 %	66,06 %	72,17 %	64,90 %
Attention croisée bidirectionnelle	53,22 %	48,98 %	53,22 %	50,64 %	54,70 %
CamemBERT	71,47 %	68,79 %	71,47 %	69,83 %	69,70 %

Table 7: Tableau des résultats.

^{1*} résultats différents sur le classement officiel dû à une erreur d'ordonnement

Perspectives

Pistes d'amélioration :

- Représenter individuellement le code (CodeBERT [Fen+20]) et le texte (CamemBERT).
- Inclure la "Soft Cardinalité" [JGG15] dans notre meilleur système.
- Concevoir un système unifié avec perceptrons multicouches.
- Favoriser les approches de bout-en-bout par réseaux neuronaux.
- Utiliser une métrique plus adaptée à la tâche.

Section 5

References

References I

- [Boj+16] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. “Enriching Word Vectors with Subword Information”. In: *arXiv preprint arXiv:1607.04606* (2016).
- [Fen+20] Zhangyin Feng, Daya Guo, Duyu Tang, Nan Duan, Xiaocheng Feng, Ming Gong, Linjun Shou, Bing Qin, Ting Liu, Daxin Jiang, and Ming Zhou. *CodeBERT: A Pre-Trained Model for Programming and Natural Languages*. 2020. DOI: 10.48550/ARXIV.2002.08155. URL: <https://arxiv.org/abs/2002.08155>.
- [GI22] Cyril Grouin and Gabriel Illouz. “Notation automatique de réponses courtes d’étudiants : présentation de la campagne DEFT 2022”. In: *Actes de DEFT*. Avignon, France, 2022.
- [JGG15] Sergio Jimenez, Fabio A. Gonzalez, and Alexander Gelbukh. “Soft Cardinality in Semantic Text Processing: Experience of the SemEval International Competitions”. en. In: *Polibits* (June 2015), pp. 63–72. ISSN: 1870-9044. URL: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-90442015000100010&nrm=iso.

References II

- [Mar+20] Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benot Sagot. “CamemBERT: a Tasty French Language Model”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020. DOI: [10.18653/v1/2020.acl-main.645](https://doi.org/10.18653/v1/2020.acl-main.645). URL: <https://doi.org/10.18653%2Fv1%2F2020.acl-main.645>.
- [Mik+13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. *Efficient Estimation of Word Representations in Vector Space*. 2013. DOI: [10.48550/ARXIV.1301.3781](https://arxiv.org/abs/1301.3781). URL: <https://arxiv.org/abs/1301.3781>.
- [RG19] Nils Reimers and Iryna Gurevych. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. 2019. DOI: [10.48550/ARXIV.1908.10084](https://arxiv.org/abs/1908.10084). URL: <https://arxiv.org/abs/1908.10084>.

References III

- [Sui+21] Philippe Suignard, Alexandra Benamar, Nazim Messous, Clément Christophe, Marie Jubault, and Meryl Bothua. “Participation d’EDF R&D à DEFT 2021 (EDF R&D Participation to DEFT 2021)”. French. In: *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier DÉfi Fouille de Textes (DEFT)*. Lille, France: ATALA, June 2021, pp. 72–81. URL: <https://aclanthology.org/2021.jeptalnrecital-deft.8>.
- [Vas+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*. 2017. arXiv: 1706.03762 [cs.CL].

Questions?