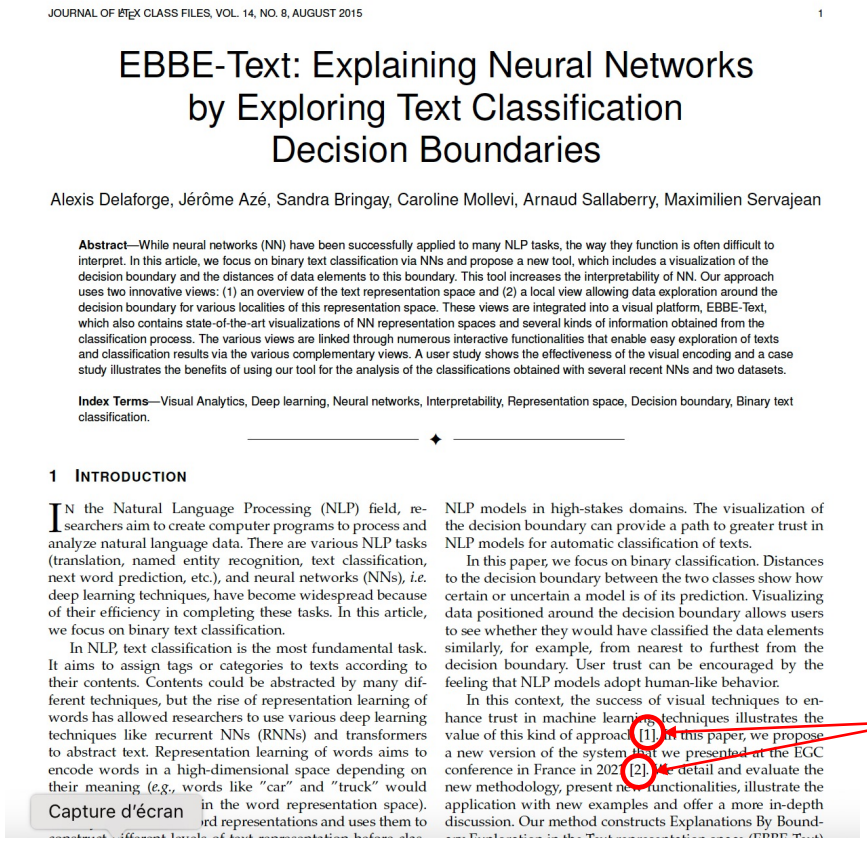
Traitement de textes

Bibliographie

Bibliographie



EBBE-Text: Explaining Neural Networks by Exploring Text Classification Decision Boundaries

Alexis Delaforge, Jérôme Azé, Sandra Bringay, Caroline Mollevi, Arnaud Sallaberry, Maximilien Servajean

Abstract—While neural networks (NN) have been successfully applied to many NLP tasks, the way they function is often difficult to interpret. In this article, we focus on binary text classification via NNs and propose a new tool, which includes a visualization of the decision boundary and the distances of data elements to this boundary. This tool increases the interpretability of NN. Our approach uses two innovative views: (1) an overview of the text representation space and (2) a local view allowing data exploration around the decision boundary for various localities of this representation space. These views are integrated into a visual platform, EBBE-Text, which also contains state-of-the-art visualizations of NN representation spaces and several kinds of information obtained from the classification process. The various views are linked through numerous interactive functionalities that enable easy exploration of texts and classification results via the various complementary views. A user study shows the effectiveness of the visual encoding and a case study illustrates the benefits of using our tool for the analysis of the classifications obtained with several recent NNs and two datasets.

Index Terms—Visual Analytics, Deep learning, Neural networks, Interpretability, Representation space, Decision boundary, Binary text classification.

1 Introduction

In the Natural Language Processing (NLP) field, researchers aim to create computer programs to process and analyze natural language data. There are various NLP tasks (translation, named entity recognition, text classification, next word prediction, etc.), and neural networks (NNs), *i.e.* deep learning techniques, have become widespread because of their efficiency in completing these tasks. In this article, we focus on binary text classification.

In NLP, text classification is the most fundamental task. It aims to assign tags or categories to texts according to their contents. Contents could be abstracted by many different techniques, but the rise of representation learning of words has allowed researchers to use various deep learning techniques like recurrent NNs (RNNs) and transformers to abstract text. Representation learning of words aims to encode words in a high-dimensional space depending on their meaning (*e.g.*, words like "car" and "truck" would [illegible] in the word representation space). [illegible]rd representations and uses them to

NLP models in high-stakes domains. The visualization of the decision boundary can provide a path to greater trust in NLP models for automatic classification of texts.

In this paper, we focus on binary classification. Distances to the decision boundary between the two classes show how certain or uncertain a model is of its prediction. Visualizing data positioned around the decision boundary allows users to see whether they would have classified the data elements similarly, for example, from nearest to furthest from the decision boundary. User trust can be encouraged by the feeling that NLP models adopt human-like behavior.

In this context, the success of visual techniques to enhance trust in machine learning techniques illustrates the value of this kind of approach [1]. In this paper, we propose a new version of the system that we presented at the EGC conference in France in 202[illegible] [2]. [illegible] detail and evaluate the new methodology, present new functionalities, illustrate the application with new examples and offer a more in-depth discussion. Our method constructs Explanations By Bound-

Capture d'écran

Liste des références en fin de document

References

[1] A. Chatzimparmpas, R. M. Martins, I. Jusufi, K. Kucher, F. Rossi, and A. Kerren, "The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations," *Computer Graphics Forum*, vol. 39, no. 3, pp. 713–756, 2020.

[2] A. Delaforge, J. Azé, A. Sallaberry, M. Servajean, S. Bringay, and C. Mollevi, "Ebbe-text : Visualisation de la frontière de décision des réseaux de neurones en classification automatique de textes," *Revue des Nouvelles Technologies de l'Information*, vol. EGC, RNTI-E-37, pp. 169–180, 2021.

[3] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau, "Visual analytics in deep learning: An interrogative survey for the next frontiers," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2674–2693, 2019.

[4] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys*, vol. 51, pp. 1–42, 2018.

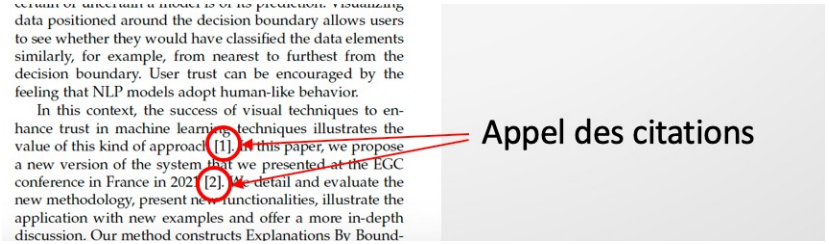
[5] Y. Goyal, A. Mohapatra, D. Parikh, and D. Batra, "Towards transparent ai systems: Interpreting visual question answering

Appel des citations

Appel de références

- Citer un ouvrage / un article dans le texte avec un appel de référence :
  - *D'après Alon et al. [AKP08], les résultats …*
  - *Comme l'affirme Hallay[2] dans son ouvrage …*
  - *La natalité (McDiarmid, 2005) diminue depuis …*

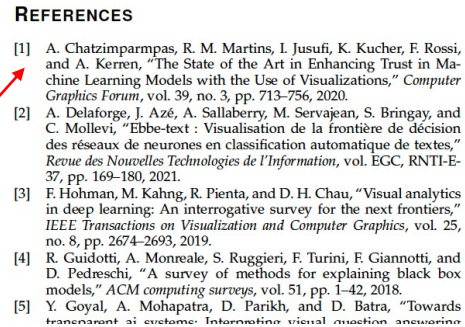
Plusieurs présentations possibles de cet appel de référence



Liste de références

- **Fin du document : récapitulatif des références bibliographiques**

Cohérence de la présentation



REFERENCES

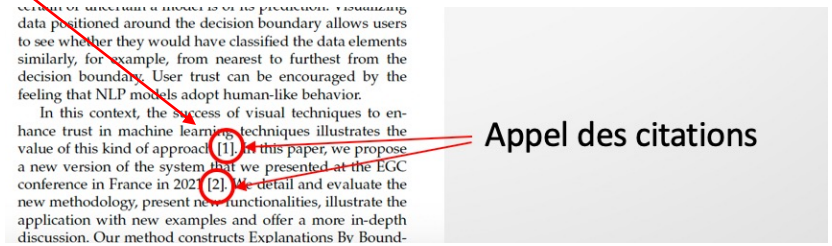
[1] A. Chatzimparmpas, R. M. Martins, I. Jusufi, K. Kucher, F. Rossi, and A. Kerren, "The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations," *Computer Graphics Forum*, vol. 39, no. 3, pp. 713–756, 2020.

[2] A. Delaforge, J. Azé, A. Sallaberry, M. Servajean, S. Bringay, and C. Mollevi, "Ebbe-text : Visualisation de la frontière de décision des réseaux de neurones en classification automatique de textes," *Revue des Nouvelles Technologies de l'Information*, vol. EGC, RNTI-E-37, pp. 169–180, 2021.

[3] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau, "Visual analytics in deep learning: An interrogative survey for the next frontiers," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2674–2693, 2019.

[4] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys*, vol. 51, pp. 1–42, 2018.

[5] Y. Goyal, A. Mohapatra, D. Parikh, and D. Batra, "Towards transparent ai systems: Interpreting visual question answering

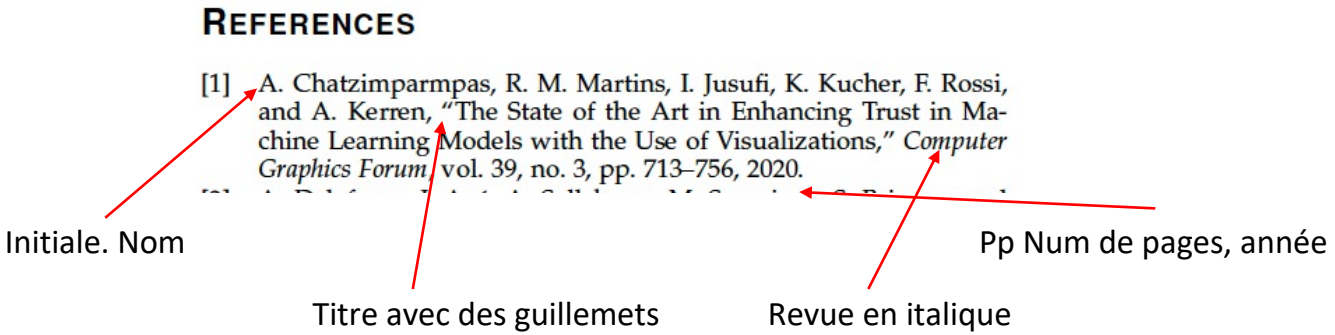


data positioned around the decision boundary allows users to see whether they would have classified the data elements similarly, for example, from nearest to furthest from the decision boundary. User trust can be encouraged by the feeling that NLP models adopt human-like behavior.

In this context, the success of visual techniques to enhance trust in machine learning techniques illustrates the value of this kind of approach [1]. In this paper, we propose a new version of the system that we presented at the EGC conference in France in 2021 [2]. We detail and evaluate the new methodology, present new functionalities, illustrate the application with new examples and offer a more in-depth discussion. Our method constructs Explanations By Bound-

Appel des citations

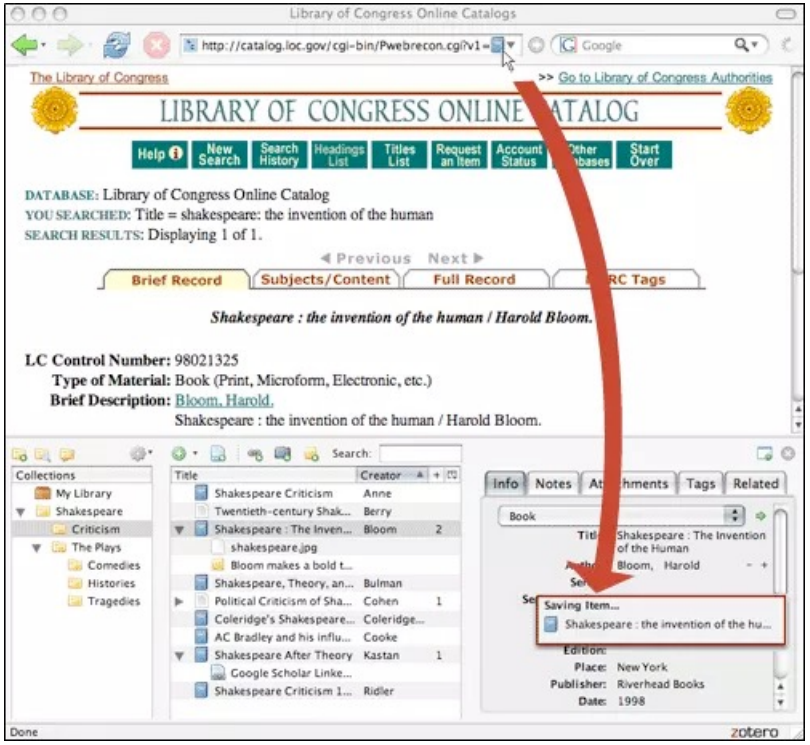
Importance de l'automatisation



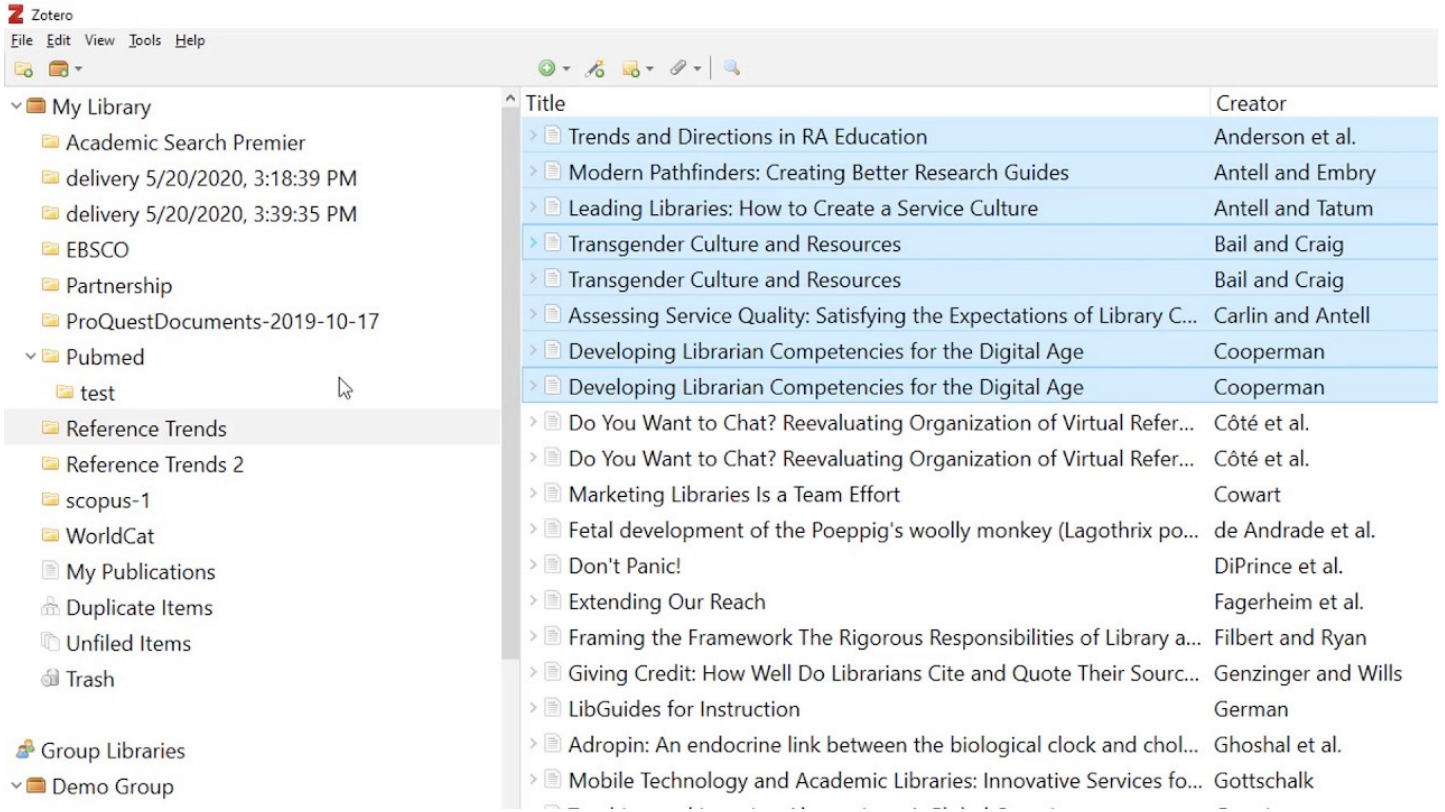
- **Problèmes possibles sans automatisation du processus :**
  - Ne pas être homogène dans la présentation
  - Oublier des ouvrages dans la bibliographie
  - ...

Zotero

Récolter automatiquement des références



Structurer les références



Insérer une référence dans le document