Introduction à l'explicabilité dans les feedback automatisés fournis aux apprenants

Esther Félix (Première année)

Institut de Recherche en Informatique de Toulouse, Université Toulouse III
118 Route de Narbonne, 31062 Toulouse Cedex 9, France
esther.felix@irit.fr

Résumé Ce papier présente les notions de feedback et d'explicabilité dans les EIAH. Il étudie comment introduire, dans des feedback fournis automatiquement aux apprenants, des explications portant sur la manière dont ces feedback ont été générés et personnalisés. L'objectif est d'améliorer l'efficacité et l'utilité perçue des feedback en les rendant plus transparents.

Keywords: Feedback · Explicabilité · Learning Analytics · EIAH · Personnalisation

1 Introduction

La notion de feedback et son rôle dans l'éducation ont été étudiés pendant des années. De nombreuses études montrent l'impact du feedback sur l'apprentissage des élèves, grâce à son influence sur leur motivation, leur confiance, leur capacité d'auto-régulation [19,15,6]. Cet impact varie selon le type du feedback et son niveau d'élaboration, et dépend également du moment où il est délivré. Dans le cadre des EIAH, différentes approches sont proposées pour intégrer le feedback au sein des environnements de manière automatisée [17]. D'un autre côté, la communauté EIAH montre un intérêt grandissant pour la notion d'explicabilité. Ce concept récent vient du champ de l'intelligence artificielle, et peu de travaux ont jusqu'à étudié l'intérêt de l'explicabilité dans les contextes d'apprentissage informatisés. Cependant, l'analyse des apprentissages (*learning analytics*) est de plus en plus utilisée pour faire des prédictions ou prendre des décisions [11] qui gagneraient à être expliquées aux différents acteurs de l'apprentissage. Introduire de l'explicabilité dans les feedback fournis à ces acteurs pourrait augmenter leur confiance et leur compréhension du système, et donc modifier la manière dont ils l'utilisent.

Cet article propose un état de l'art des recherches portant sur les notions de feedback et d'explicabilité ainsi que sur leurs applications dans le domaine de l'éducation, et pose la question de l'impact sur le comportement des étudiants d'un feedback intégrant une composante d'explicabilité. Il propose enfin les grandes lignes d'une expérience actuellement en cours afin d'ouvrir des perspectives pour répondre à cette question.

2 Le feedback dans l'apprentissage

En 1995, Winne et Butler définissent le feedback comme une « information qui permet à un apprenant de confirmer, d'ajouter, d'écraser, d'ajuster ou de restructurer de l'information en mémoire, qu'il s'agisse de connaissances du domaine, de connaissances métacognitives, de croyances sur soi et sur les tâches, ou de stratégies cognitives » [5]. À partir de cette définition, Hattie et Timperley, dans leur revue de littérature référence sur la notion de feedback, proposent un modèle du feedback et évaluent son efficacité selon différentes perspectives [12]. Ils définissent le feedback comme une information fournie par un agent (par exemple un professeur, un parent, un pair) concernant la performance ou la compréhension d'un individu. L'objectif du feedback est de réduire l'écart entre d'une part la compréhension et la performance d'un apprenant à un moment donné (« là où il se trouve »), et d'autre part des objectifs fixés à atteindre (« là où il veut aller »). Cependant, cet objectif n'est pas facile à atteindre, car un feedback peut être inadapté, rejeté ou mal interprété et donc ne pas avoir l'impact désiré.

D'après Hattie et Timperley, le feedback, pour être efficace, doit répondre aux questions suivantes :

- Quels sont les objectifs de l'apprenant ? (« *feed-up* »). Il s'agit de comparer la situation actuelle à l'objectif visé.
- Quel est son avancement actuel ? (« *feed-back* »). C'est l'analyse de la progression qui a été effectuée.
- Quelle est la prochaine étape ? (« *feed-forward* »). La réponse à cette question doit fournir une explication de l'objectif ciblé à partir de l'état d'avancement actuel, en donnant des informations sur ce qui est compris et sur ce qui ne l'est pas. C'est ce qui mène à une adaptation de l'apprentissage, à l'élaboration de nouvelles stratégies, et aide au processus d'auto-régulation [12].

Le feedback est véhiculé par un canal qui peut prendre plusieurs formes : oral, écrit, audio, vidéo, image, schéma, etc. Le feedback a également une direction (de professeur à élève, d'élève à professeur, d'élève à élève, etc.) et une valence (positif ou négatif). De plus, le feedback peut être différencié selon plusieurs niveaux :

- Le feedback au niveau de la tâche, qui permet d'indiquer si une tâche précise a été bien accomplie ou comprise ;
- Le feedback au niveau du processus, qui donne un retour sur les stratégies de résolution qui ont été mises en place par l'apprenant ;
- Le feedback au niveau de l'auto-régulation, qui se focalise sur la régulation par l'apprenant de ses stratégies d'apprentissage ;
- Le feedback sur la personne, qui consiste en un retour sur l'apprenant mais pas sur la tâche en elle-même (par exemple : « Bravo, tu es doué »).

Les résultats obtenus par la revue de littérature de Hattie et Timperley révisés par une méta-analyse en 2020 [21] montrent que les types de feedback les plus efficaces pour améliorer les performances des apprenants sont ceux qui :

— Sont dirigés vers les tâches, les processus ou la régulation, et non vers la personne ;

— Contiennent des informations de haut niveau, c'est-à-dire qui permettent de comprendre pourquoi des erreurs ont été faites et comment les éviter, plutôt que d'indiquer simplement l'existence d'erreurs ;

— Sont délivrés selon un timing adapté en fonction du stade d'avancement dans l'apprentissage et de la difficulté de la tâche (une tâche facile nécessitant un feedback immédiat).

Il est cependant à noter que fournir un feedback n'est pas toujours souhaitable : dans certaines situations, si la tâche demandée n'a pas du tout été comprise par exemple, il vaut mieux expliquer à nouveau les consignes et s'assurer que les compétences de base sont acquises plutôt que donner un feedback sur ce que l'apprenant a effectué [12]. D'après une revue de littérature de 2008 de Valerie J. Shute portant sur le feedback formatif (défini comme les informations communiquées à l'apprenant et destinées à modifier sa pensée ou son comportement dans le but d'améliorer l'apprentissage), un feedback utile et efficace pour l'apprentissage dépend de trois éléments : la motivation (l'apprenant en a besoin), l'opportunité (l'apprenant le reçoit à temps pour l'utiliser), et les moyens (l'apprenant est capable et désireux de l'utiliser). Cependant, même en prenant en compte ces trois éléments, il existe toujours une grande variabilité des effets du feedback sur la performance et l'apprentissage [18]. De manière générale, il est souhaitable d'adapter le feedback au niveau de l'apprenant qui le reçoit. Toutes ces considérations font ressortir le besoin de personnalisation du feedback.

Dans la plupart des contextes éducatifs, fournir un feedback de haut niveau rapide et personnalisé à chaque apprenant n'est pas viable car cela exige un suivi permanent de tous les apprenants et la prise en compte des actions spécifiques de chaque individu. C'est pour cette raison que les professeurs donnent généralement des feedback au niveau de la tâche, notamment à l'aide des notes évaluatives [12]. Les EIAH offrent des perspectives intéressantes concernant cette problématique, avec la possibilité d'exploiter les *learning analytics* de manière à automatiser le feedback fourni aux apprenants, en proposant du feedback automatisé au niveau de la tâche mais aussi au niveau du processus [17]. Dans une revue de littérature de 2021, Deeva et al. ont développé un cadre permettant de classifier les systèmes de feedback automatisés (TAF-ClaF - Technologies for Automated Feedback - classification framework) en fonction de plusieurs composantes : l'architecture du système, les propriétés du feedback (timing, adaptabilité, contrôlabilité...), le contexte éducatif ainsi que la méthode d'évaluation utilisée [9]. Une des recommandations données en conclusion de l'étude est d'utiliser davantage de systèmes de feedback automatisés fondés sur les données afin de permettre une meilleure personnalisation.

Proposer du feedback dans les EIAH est d'autant plus pertinent qu'une méta-analyse récente montre qu'une partie importante des résultats obtenus par Hattie et Timperley dans des contextes d'apprentissage classiques sont transférables aux contextes d'apprentissage dans des environnements informatiques : les feedback les plus élaborés sont ceux qui mènent aux meilleurs résultats d'apprentissage [13]. Les auteurs de cette méta-analyse expliquent cependant que leurs résultats sont vérifiés pour des apprenants en études supérieures plus que pour des apprenants plus jeunes, et que les considérations concernant la forme et le timing optimaux à donner au feedback dans les environnements informatiques n'ont pas été déterminées. Ils remarquent également que la plupart des études semblent considérer que les apprenants sont nécessairement attentifs au feedback qui leur est fourni, ce qui n'est en réalité souvent pas le cas [20].

Ce cadre théorique nous permet d'orienter nos recherches vers l'utilisation de feedback de haut niveau portant sur les stratégies d'apprentissage, et de faire porter nos expériences sur des apprenants en études supérieures. Nous souhaitons utiliser le potentiel des données recueillies dans les EIAH pour proposer un feedback plus efficace et personnalisé.

3 Explicabilité

Au cours des dernières années, le développement du champ de l'intelligence artificielle et du *machine learning* a fait émerger le besoin d'ajouter une couche d'interprétabilité aux modèles utilisés, de manière à expliquer les résultats des algorithmes de type « boîte noire » et ainsi comprendre leur fonctionnement. Ce champ d'étude est celui de l'explicabilité ou xAI (eXplainable Artificial Intelligence). Les explications permettent d'augmenter la transparence d'un système ainsi que la confiance des utilisateurs en ce système [14].

Utiliser le concept d'explicabilité nécessite de se poser un certain nombre de questions, notamment sur la notion même d'explication qui est rarement clairement définie [1]. En effet, une explication peut prendre de nombreuses formes qui dépendent des utilisateurs ciblés, du contexte et de l'objectif visé, et les variations possibles sont donc aussi nombreuses que les réponses aux questions « comment, qui, quoi, pourquoi ? ». De plus, les explications ne sont pas toujours voulues ni nécessairement bénéfiques pour les utilisateurs [4,10]. L'utilité des explications dépend donc elle aussi de divers facteurs, des caractéristiques des utilisateurs et du contexte [7]. Cela induit un besoin d'adaptabilité de ces explications.

Aussi, le domaine de l'explicabilité est majoritairement ciblé vers des experts en intelligence artificielle. Quelques rares auteurs ont fait des recherches visant à créer des explications à destination d'acteurs non experts en informatique, comme les utilisateurs d'un produit construit avec un algorithme opaque (par

exemple, un système d'allocation de prêt bancaire)[1,16]. Mais l'application de la notion d'explicabilité au domaine de l'éducation dans le contexte d'environnements informatiques est une perspective qui nécessite de différencier des acteurs de manière plus fine que la simple distinction expert/non expert, avec la prise en compte des besoins différents de professeurs et d'élèves à divers niveaux scolaires. Il s'agit d'une perspective à notre connaissance récente et très peu abordée [8].

Les possibilités ouvertes par l'introduction de l'explicabilité dans des applications éducatives comportent l'amélioration des *Open Learner Models* (OLM) avec un plus haut degré d'adaptabilité, de précision et d'interaction. D'autres travaux portent sur des tuteurs intelligents et visent à expliquer aux apprenants la manière dont les conseils donnés par le système ont été générés [8].

De manière générale, nous considérons l'explicabilité dans les EIAH de la manière suivante : dès lors qu'un algorithme de classification ou de prédiction est appliqué, que des traces d'apprentissage sont traitées par du *machine learning* ou qu'une décision visant à soutenir l'apprentissage est prise par le système informatique, il peut être utile de fournir à l'utilisateur une explication adaptée concernant le fonctionnement du système ou de l'algorithme. Nous nous intéressons plus spécifiquement à la possibilité d'introduire cette notion d'explicabilité dans les feedback qui peuvent être automatiquement générés par les EIAH et délivrés aux utilisateurs. Nous faisons l'hypothèse que l'explicabilité a le potentiel de rendre les feedback produits par les EIAH plus transparents et donc plus acceptables pour les utilisateurs. Dans le contexte de feedback automatiques adaptés aux utilisateurs qui les reçoivent, nous pensons que les modèles utilisés pour la génération du feedback peuvent être exploités pour fournir des explications personnalisées qui favoriseront l'acceptabilité de l'EIAH.

4 Perspectives : mise en place d'une expérience

Nous avons vu que l'explicabilité comme le feedback doivent, pour être efficaces et pertinents, répondre à un besoin d'adaptabilité. Nous pensons que fournir aux apprenants ou aux professeurs un feedback non seulement personnalisé, mais qui inclut de surcroît une explication indiquant *pourquoi* le feedback leur est adapté ou *comment* il a été conçu, pourrait améliorer la perception des feedback et leur efficacité.

Pour vérifier cette hypothèse, nous nous appuyons sur la plateforme Lab4CE (Laboratory for Computer Education) [3]. Cet outil permet à chaque élève, grâce à des technologies de virtualisation, d'avoir accès à son propre laboratoire virtuel dont les caractéristiques peuvent varier en fonction de la matière étudiée. En nous appuyant sur cet outil, nous avons conçu une expérience en cours avec des étudiants de l'Institut Universitaire Technologique de l'Université Toulouse III dans le contexte de l'apprentissage de l'informatique. La plateforme Lab4CE

est accessible en ligne pendant toute la durée de l'expérience, et les étudiants ont une séance de Travaux Pratiques hebdomadaire de programmation Shell pendant cinq semaines. À la fin de chaque semaine, l'analyse des productions des apprenants à travers un algorithme de classification permet de classer les élèves selon trois profils de programmation [2] qui révèlent l'efficacité des stratégies de programmation des élèves. Par exemple, le fait de soumettre différentes versions d'un programme très rapidement, sans prendre un temps de réflexion après chaque erreur (profil « essai-erreur »), est révélateur d'une mauvaise stratégie. Le résultat de cette classification est ensuite utilisé par le système de génération de feedback, qui attribue un certain feedback à un apprenant en fonction du profil dans lequel il a été classifié (par exemple, les élèves classifiés dans le profil « essai-erreur » obtiennent un feedback leur conseillant spécifiquement de modifier cette stratégie).

Avant le début de l'expérience, nous avons constitué des groupes de niveau équivalent et contenant un nombre égal d'élèves. Ces groupes déterminent le type de feedback reçu à la fin de chaque semaine par les élèves. Les élèves du premier groupe reçoivent un feedback de haut niveau sous la forme d'un conseil fondé sur leur profil et accompagné d'une explication venant appuyer le conseil par des statistiques (par exemple, un faible temps moyen écoulé entre deux soumissions de code permet d'expliquer la nécessité de prendre un temps de réflexion avant d'écrire ou modifier son programme). Le deuxième groupe reçoit uniquement le conseil personnalisé sans l'explication associée. Le troisième groupe est un groupe contrôle ne recevant pas de feedback. Ce groupe permet d'avoir une base de comparaison commune aux deux types de feedback, et de vérifier si la présence de feedback a un effet ou non dans notre contexte d'apprentissage (les effets du feedback pouvant en effet être variables, comme le montrent les résultats inégaux de la littérature sur le sujet). Cette expérience en cours vise à comparer l'évolution des comportements au fil des séances de TP, les résultats à l'examen de cette matière ainsi que la perception des élèves confrontés à des feedback personnalisés de haut niveau (portant sur les stratégies d'apprentissage) avec et sans explications.

Références

1. Arya, V., Bellamy, R.K.E., Chen, P.Y., Dhurandhar, A., Hind, M., Hoffman, S.C., Houde, S., Liao, Q.V., Luss, R., Mojsilović, A., Mourad, S., Pedemonte, P., Raghavendra, R., Richards, J., Sattigeri, P., Shanmugam, K., Singh, M., Varshney, K.R., Wei, D., Zhang, Y. : One Explanation Does Not Fit All : A Toolkit and Taxonomy of AI Explainability Techniques. arXiv :1909.03012 [cs, stat] (Sep 2019), `http://arxiv.org/abs/1909.03012`, arXiv : 1909.03012
2. Bey, A., Pérez-Sanagustín, M., Broisin, J. : Unsupervised Automatic Detection of Learners' Programming Behavior. In : Scheffel, M., Broisin, J., Pammer-Schindler, V., Ioannou, A., Schneider, J. (eds.) Transforming Learning with Meaningful Technologies. pp. 69–82. Lecture Notes in Computer Science, Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-29736-7_6

3. Broisin, J., Venant, R., Vidal, P. : Lab4CE : a Remote Laboratory for Computer Education. International Journal of Artificial Intelligence in Education **27**(1), 154–180 (2017). https://doi.org/10.1007/s40593-015-0079-3, https://hal.archives-ouvertes.fr/hal-01530319
4. Bunt, A., Lount, M., Lauzon, C. : Are Explanations Always Important ? A Study of Deployed, Low-Cost Intelligent Interactive Systems. Proceedings of the 2012 ACM international conference on Intelligent User Interfaces p. 10 (2012)
5. Butler, D., Winne, P. : Feedback and Self-Regulated Learning : A Theoretical Synthesis. Review of Educational Research - REV EDUC RES **65**, 245–281 (Sep 1995). https://doi.org/10.2307/1170684
6. Campos, D.S., Mendes, A.J., Marcelino, M.J., Ferreira, D.J., Alves, L.M. : A multinational case study on using diverse feedback types applied to introductory programming learning. In : 2012 Frontiers in Education Conference Proceedings. pp. 1–6 (Oct 2012). https://doi.org/10.1109/FIE.2012.6462412, iSSN : 2377-634X
7. Conati, C., Porayska-Pomsta, K., Mavrikis, M. : AI in Education needs interpretable machine learning : Lessons from Open Learner Modelling. ArXiv (2018)
8. Conati, C., Barral, O., Putnam, V., Rieger, L. : Toward personalized XAI : A case study in intelligent tutoring systems. Artificial Intelligence **298**, 103503 (Sep 2021). https://doi.org/10.1016/j.artint.2021.103503, https://www.sciencedirect.com/science/article/pii/S0004370221000540
9. Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., De Weerdt, J. : A review of automated feedback systems for learners : Classification framework, challenges and opportunities. Computers & Education **162**, 104094 (2021). https://doi.org/https://doi.org/10.1016/j.compedu.2020.104094, https://www.sciencedirect.com/science/article/pii/S036013152030292X
10. Ehrlich, K., Kirk, S., Patterson, J., Rasmussen, J., Ross, S., Gruen, D. : Taking advice from intelligent systems : the double-edged sword of explanations. In : Proceedings of the 16th international conference on Intelligent user interfaces. pp. 125–134 (Jan 2011). https://doi.org/10.1145/1943403.1943424
11. Elias, T. : Learning Analytics : Definitions, Processes and Potential. Learning Analytics p. 23 (2011)
12. Hattie, J., Timperley, H. : The Power of Feedback. Review of Educational Research **77**(1), 81–112 (Mar 2007). https://doi.org/10.3102/003465430298487, http://journals.sagepub.com/doi/10.3102/003465430298487
13. Kleij, Eggen, T., Veldkamp : Effects of feedback in a computer-based assessment for learning - ScienceDirect, https://www.sciencedirect.com/science/article/pii/S0360131511001783
14. Kraus, S., Azaria, A., Fiosina, J., Greve, M., Hazon, N., Kolbe, L., Lembcke, T.B., Müller, J.P., Schleibaum, S., Vollrath, M. : AI for Explaining Decisions in Multi-Agent Environments. Proceedings of the AAAI Conference on Artificial Intelligence **34**(09), 13534–13538 (Apr 2020). https://doi.org/10.1609/aaai.v34i09.7077, http://arxiv.org/abs/1910.04404, arXiv : 1910.04404
15. Martinez-Argüelles, M., Plana-Erta, D., Hintzmann, C., Batalla-Busquets, J.M., Badia-Miró, M. : Usefulness of feedback in e-learning from the students' perspective. Intangible Capital **11**, 627–645 (Oct 2015). https://doi.org/10.3926/ic.622
16. Ribera Turró, M., Lapedriza, A. : Can we do better explanations ? A proposal of User-Centered Explainable AI. In : Explainable Smart Systems 2019 Conference (Mar 2019)

17. Serral, E., De Weerdt, J., Sedrakyan, G., Snoeck, M. : Automating immediate and personalized feedback taking conceptual modelling education to a next level. In : 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS). pp. 1–6 (Jun 2016). https://doi.org/10.1109/RCIS.2016.7549293, iSSN : 2151-1357

18. Shute, V.J. : Focus on formative feedback. Review of Educational Research **78**(1), 153–189 (2008). https://doi.org/10.3102/0034654307313795, https://doi.org/10.3102/0034654307313795

19. Tanes, Z., Arnold, K.E., King, A.S., Remnet, M.A. : Using Signals for appropriate feedback : Perceptions and practices. Computers & Education **57**(4), 2414–2422 (Dec 2011), https://www.learntechlib.org/p/50819/, publisher : Elsevier Ltd

20. Timmers, C., Veldkamp, B. : Attention paid to feedback provided by a computer-based assessment for learning on information literacy. Computers & Education **56**(3), 923–930 (Apr 2011). https://doi.org/10.1016/j.compedu.2010.11.007, https://www.sciencedirect.com/science/article/pii/S0360131510003283

21. Wisniewski, B., Zierer, K., Hattie, J. : The Power of Feedback Revisited : A Meta-Analysis of Educational Feedback Research. Frontiers in Psychology **10**, 3087 (2020). https://doi.org/10.3389/fpsyg.2019.03087, https://www.frontiersin.org/article/10.3389/fpsyg.2019.03087