

Raisonner avec des cas positifs et négatifs

Sujet de thèse proposé par Emmanuel Nauer et Jean Lieber (équipe \mathcal{K} du Loria),
UL, CNRS, Inria, Loria, F-54000 Nancy (prénom.nom@loria.fr)

Contexte scientifique. Le raisonnement à partir de cas (RàPC [1]) est un modèle de raisonnement lié à une base de cas BC où un cas est une représentation d'une expérience de résolution de problème, généralement sous la forme d'un couple $(pb, sol(pb))$ où pb est le problème dans un domaine d'application donné et $sol(pb)$ est une solution de pb . Par exemple, dans le domaine culinaire, un problème peut être $pb_{\text{ananas}} =$ « Je veux une recette avec de l'ananas. » et une solution $sol(pb_{\text{ananas}})$ peut être une recette d'une tarte à l'ananas. Un *cas source* est un cas $(srce, sol(srce))$ de BC.

Une session de RàPC prend en entrée un nouveau problème, le *problème cible*, noté $cible$, et se compose habituellement des étapes suivantes. Premièrement, le cas source $(srce, sol(srce))$ qui est le plus proche de $cible$ est sélectionné (étape de remémoration). $sol(srce)$ est ensuite modifiée en une solution $sol(cible)$ de $cible$ (étape d'adaptation). Après cela, le nouveau cas résultat $(cible, sol(cible))$ est proposé à l'utilisateur (étape de validation) et, si validé, stocké dans BC (étape de mémorisation). Par exemple, si un système de RàPC culinaire est interrogé avec $cible = pb_{\text{ananas}}$, une recette de tarte aux pommes peut être remémoré (s'il n'y a pas de cas source plus similaire à $cible$ que celui-ci); elle peut ensuite être adaptée en substituant les pommes par de l'ananas (en changeant les quantités – 3 pommes pour 1 ananas – et en faisant d'autres ajustements).

Pour raisonner, un système de RàPC utilise une base de connaissances généralement composée de quatre *conteneurs de connaissances* [2] : BC, l'ontologie du domaine, la similarité (i.e., connaissances de remémoration) et des connaissances d'adaptation (souvent représentées par des règles d'adaptation). Pour enrichir cette base de connaissances, des méthodes et des outils d'acquisition de connaissances (à partir de données et/ou avec des experts) ont été étudiés. Par exemple, nous avons étudié l'acquisition de connaissances d'adaptation à partir de techniques d'extraction de connaissances [3].

Ce schéma très classique du RàPC est fondé sur une hypothèse implicite : les cas sont des *cas positifs*, i.e. ils sont supposés être satisfaisants (de façon très dépendante du domaine : en cuisine, un cas positif correspondrait, par exemple, à une recette qui est appréciée par beaucoup de personnes). Cependant, il existe également *des cas négatifs*, en particulier les cas $(cible, sol(cible))$ proposés par adaptation mais rejetés lors de la validation. Ces cas ne sont pratiquement jamais utilisés par les systèmes de RàPC alors qu'ils constituent des unités de connaissances potentiellement utiles.

Objectif de la thèse. L'objectif de cette thèse est de considérer ces cas négatifs, une fois identifiés comme tels, comme des unités de connaissances à part entière pour de futures sessions de RàPC. L'idée principale est que la base de cas est partitionnée en $BC = BC^+ \cup BC^-$ et qu'un cas source est utilisé différemment selon qu'il soit positif (membre de BC^+) ou négatif (membre de BC^-).

Un nouveau schéma de RàPC doit être défini. Voici quelques idées sur comment reconsidérer les différentes étapes d'un système de RàPC en prenant en compte ces deux types de cas :

Remémoration L'idée pourrait consister à remémorer le cas source positif $(srce^+, sol(srce^+))$ et le cas source négatif $(srce^-, sol(srce^-))$ les plus proches de $cible$. Si la remémoration est implémentée via une mesure de similarité, devrait-il y avoir deux mesures ?

Adaptation Le principe pourrait être d'utiliser $(srce^+, sol(srce^+))$ tout en évitant $(srce^-, sol(srce^-))$, par exemple, à partir d'une analyse de la différence entre les solutions $sol(srce^+)$ et $sol(srce^-)$. Une autre façon de voir les choses est d'utiliser des opérations de changement de croyance : la révision des croyances est utilisée dans certaines approches d'adaptation [4] et pourrait être utilisée pour adapter $(srce^+, sol(srce^+))$, mais une autre opération de changement de croyances telle que la *contraction* pourrait être utilisée pour éviter $(srce^-, sol(srce^-))$.

Validation et mémorisation Le processus interactif de validation pourrait être le même : le nouveau cas $(cible, sol(cible))$ est étiqueté soit comme positif, soit comme négatif. La mémorisation stockera ce cas $(cible, sol(cible))$ soit dans BC^+ , soit dans BC^- en fonction du résultat de sa validation.

Méthodes et outils pour l'acquisition de connaissances Pour chaque conteneur de connaissances, la question se pose sur comment prendre en compte les cas positifs et négatifs :

Base de cas En plus des cas négatifs acquis lors des étapes de validation et de mémorisation, devrait-il y avoir un effort particulier pour acquérir d'autres cas négatifs ? Et si oui, ces cas devraient-ils être proches de cas positifs ?

Ontologie du domaine L'ontologie du domaine peut-être vue comme un ensemble de conditions nécessaires pour qu'un cas positif soit licite. Quel impact pourrait avoir un cas négatif sur l'ontologie ? Il pourrait être utile pour suggérer de nouvelles conditions nécessaires à ajouter à l'ontologie du domaine qui éviterait la réapparition d'un tel cas négatif dans le futur. Cela se rapproche du problème d'apprentissage avec des exemples positifs et négatifs en apprentissage automatique [5].

Similarity Comme mentionné précédemment, la remémoration du cas positif ou du cas négatif le plus proche peut être de différentes natures. Par conséquent, cette différence peut se refléter sur l'acquisition des connaissances de remémoration.

Connaissances d'adaptation L'acquisition des connaissances d'adaptation peut être réalisée par utilisation d'un processus de fouille de données sur les différences entre cas sources et en interprétant les résultats sous forme de règles d'adaptation. La question ici est comment ce principe peut être modifié en prenant en compte à la fois les cas positifs et négatifs ? Une première étude, qui donne des résultats encourageants, a été menée [6].

Contexte applicatif. Les principes proposés durant la thèse devront être validés. Deux types de validation sont envisagés : l'une sur une application jouet qui permettra de mesurer avec précision les bénéfices de l'utilisation des cas négatifs, et l'autre dans le cadre d'une application réelle développée dans notre équipe et avec nos partenaires, dans le but d'avoir une validation pratique. Cette application sera dans le domaine de la cuisine [7], du diagnostic médical [8] et/ou de la traduction automatique [9].

Références

- [1] C. K. Riesbeck and R. C. Schank. *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1989.
- [2] M. Richter and R. Weber. *Case-based reasoning : a textbook*. Springer Science & Business Media, 2013.
- [3] E. Gaillard, J. Lieber, and E. Nauer. Adaptation knowledge discovery for cooking using closed itemset extraction. In *The 8th Int. Conf. on Concept Lattices and their Applications - CLA 2011*, October 2011.
- [4] J. Cojan and J. Lieber. Applying Belief Revision to Case-Based Reasoning. In H. Prade and G. Richard, editors, *Computational Approaches to Analogical Reasoning : Current Trends*, volume 548 of *Studies in Computational Intelligence*, pages 133 – 161. Springer, 2014.
- [5] R. Michalski, J. Carbonell, and T. Mitchell. *Machine learning : An artificial intelligence approach*. Springer Science & Business Media, 2013.
- [6] T. Gillard, J. Lieber, and E. Nauer. Improving Adaptation Knowledge Discovery by Exploiting Negative Cases : First Experiment in a Boolean Setting. In *ICCBR 2018 - 26th International Conference on Case-Based Reasoning*, Stockholm, Sweden, July 2018.
- [7] A. Cordier, V. Dufour-Lussier, J. Lieber, E. Nauer, F. Badra, J. Cojan, E. Gaillard, L. Infante-Blanco, P. Molli, A. Napoli, and H. Skaf-Molli. Taaable : a Case-Based System for personalized Cooking. In S. Montani and L. C. Jain, editors, *Successful Case-based Reasoning Applications-2*, volume 494 of *Studies in Computational Intelligence*, pages 121–162. Springer, 2014.
- [8] M. B. Chawki, E. Nauer, N. Jay, and J. Lieber. Tetra : A Case-Based Decision Support System for Assisting Nuclear Physicians with Image Interpretation. In *ICCBR 2017 - Int. Conf. on Case-Based Reasoning*, volume 10339 of *Lecture Notes in Computer Science*, pages 108–122. Springer, June 2017.
- [9] Y. Lepage and J. Lieber. Case-Based Translation : First Steps from a Knowledge-Light Approach Based on Analogy to a Knowledge-Intensive One. In *ICCBR 2018 - 26th Int. Conf. on Case-Based Reasoning*, 2018.