

Stage Master 2 :

Fouille visuelle de données expérimentales biologiques par évolution interactive.

Contexte

Le développement de méthodes informatiques de visualisation de données dans les laboratoires de recherche pose de nombreuses questions de recherche difficiles du fait de la complexité, de la nature et de la quantité de données qui peuvent être récupérées au cours des expériences. L'obtention des données est coûteuse en temps et en ressources, faciliter leur exploitation est une question cruciale. L'objet de cette étude est d'élaborer de nouvelles méthodes d'exploration visuelle de jeux de données très multi-dimensionnelles, telles celles modélisées dans l'équipe « MALICES » ou celles générées par l'équipe « EcoMic » : relativement peu d'échantillons, mais beaucoup de mesures (physicochimiques, microbiologiques) sur un écosystème microbien fromager, y compris des mesures dynamiques et des métadonnées génomiques issues de technologies de séquençage à haut débit. Ces méthodes seront développées dans le cadre d'une collaboration entre une équipe de l'INRIA (AVIZ) spécialiste de visualisation et deux équipes INRA de l'UMR782, l'équipe « MALICES » spécialiste de modélisation agro-alimentaire et biologique et l'équipe « EcoMic », de biologistes expérimentateurs travaillant sur les mécanismes de fonctionnement de l'écosystème fromager. L'approche proposée repose sur l'emploi d'une technique d'optimisation stochastique (l'algorithmique évolutionnaire interactive) comme moteur d'exploration pour guider la construction de vues pertinentes au sein d'un outil de visualisation de données. Cette démarche s'appuie sur des travaux prospectifs récents ayant prouvé l'intérêt de ce schéma (logiciel EvoGraphDice, voir figure 1 et <http://www.aviz.fr/EVE>).

Algorithmes évolutionnaires, algorithmes génétiques

L'évolution artificielle consiste à formuler des problèmes d'optimisation sous la forme d'un espace offert à l'exploration d'une population artificielle, et à simuler une évolution Darwinienne sur cette population (<http://evelyne-lutton.fr/EvolutionaryAlgorithms.html>). Ces techniques d'optimisation stochastique permettent d'aborder des problèmes d'optimisation très irréguliers et difficiles, souvent hors de portée des méthodes d'optimisation classiques. Un des avantages de ces techniques est qu'aucune propriété de continuité ou de régularité n'est imposée aux quantités optimisées, ce qui permet de prendre facilement en compte des jugements subjectifs issus d'interactions-utilisateurs au cours des processus d'optimisation (Evolution interactive).

Visualisation interactive évolutionnaire

Dans le cadre de ce stage, ces techniques d'évolution artificielle interactive seront utilisées dans une optique d'exploration et non plus d'optimisation. Les questions importantes posées au cours de ce travail sont les suivantes :

- Comment faciliter les pratiques d'analyse de données expérimentales en utilisant la visualisation ?

- Comment interconnecter de nouveaux outils de visualisation avec les outils classiques d'analyse statistique ?
- Comment évaluer l'efficacité d'un outil de visualisation dans le contexte des labos de l'unité GMPA ?
- Comment mesurer et prendre en compte l'intérêt de l'utilisateur de façon fluide et non invasive ?

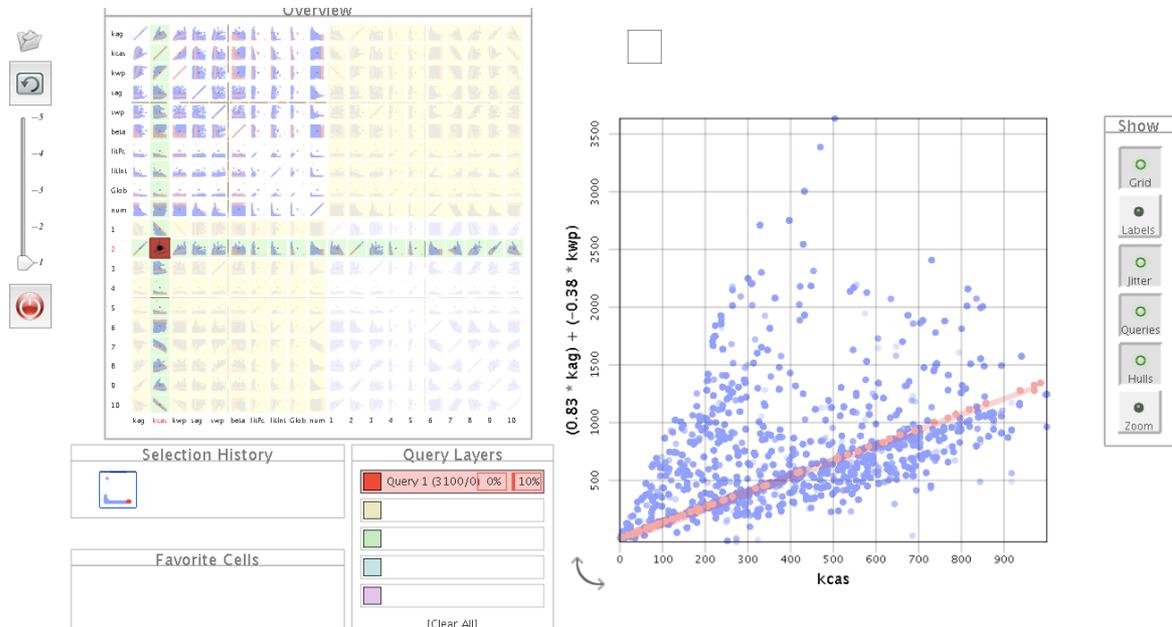


Figure 1. Exemple d'utilisation du logiciel EvoGraphDice sur un jeu de données issu de la simulation d'un modèle de gel lacté (Lutton et al, 2013). L'ensemble des vues possibles est résumé dans la matrice « overview » (à gauche de la figure), dans laquelle les 10 dimensions du jeu de données sont représentées en ligne et en colonnes. Un jeu additionnel de dimensions composites (numérotées 1 à 10) est proposé par l'algorithme évolutionnaire qui « apprend » progressivement les préférences de l'utilisateur, en capturant son évaluation de chaque vue zoomée par le biais d'un curseur de notation qu'il peut manipuler (en haut à gauche, 1 = peu intéressant, 5 = très intéressant). La vue zoomée (à droite de la figure) est repérée dans la matrice par l'intersection de la ligne et de la colonne surlignée en vert. Dans ce zoom, nous voyons une représentation visuelle d'une dépendance remarquable entre une variable composée (la No 2 = $0.83 \text{ kag} - 0.38 \text{ wp}$) et une variable primaire (kcas). Différents outils de sélection, de sélection de vues favorites, de gestion des historiques, et des transitions animées permettent une navigation fluide et intuitive dans le jeu de données de dimensions 10 + 10. La navigation dans un espace de plus grande dimension (>20) pose des problèmes de lisibilité de la matrice des résumés, de placement dans la fenêtre de visualisation, et de surcharge cognitive pour l'utilisateur. L'interconnexion avec des outils d'analyse statistique ou des outils de modélisation pose encore d'autres problèmes. Les solutions envisageables dépendent du domaine d'application et des contraintes de travail des utilisateurs ciblés.

Buts du stage

Le logiciel EvoGraphDice servira de base pour le développement d'un prototype d'exploration évolutionnaire interactive d'espace multidimensionnel de grande dimension, les fonctionnalités suivantes seront étudiées :

- suggestion de points de vues non triviaux par combinaison non linéaires d'axes,
- calculs de différents résumés de données (statistiques ou non),
- sélection et visualisation de sous-ensembles pertinents de dimensions,
- visualisation de données non numériques (catégorielles, symboliques),
- prise en compte de données manquantes, d'incertitudes.

La validation des méthodes repose sur le développement de logiciel et sur le test sur données réelles fournies par les équipes partenaires : Malices, EcoMic (sous la responsabilité de Françoise Irlinger), et d'autres utilisateurs partenaires qui pourront s'associer au cours du projet.

Prérequis

- Bonnes compétences en programmation (java, C, C++, Matlab, R), apprentissage, et développement d'interfaces.
- Un intérêt pour la recherche (ce stage pourra éventuellement déboucher sur une thèse).
- Une connaissance des méthodes évolutionnaires et de visualisation sera un plus.
- Il n'est pas nécessaire d'avoir des connaissances en biologie pour aborder ce sujet, mais l'étudiant aura à collaborer avec des biologistes.

Lieu de travail

UMR GMPA, Thiverval-Grignon.

Encadrement

Evelyne Lutton, *Equipe MALICES*.

Françoise Irlinger et Eric Dugat-Bony, *Equipe ECOMIC*.

Pour candidater

Envoyez CV et lettre de motivation à Evelyne.Lutton@grignon.inra.fr

Quelques références pour commencer

- Waldo Cancino, Nadia Boukhelifa, Anastasia Bezerianos, and Evelyne Lutton. Evolutionary visual exploration: Experimental analysis of algorithm behaviour. In *VizGEC 2013, Workshop on Visualisation Methods in Genetic and Evolutionary Computation. Genetic and Evolutionary Computation Conference, GECCO 2013.*, July 2013.
- Nadia Boukhelifa, Waldo Cancino, Anastasia Bezerianos, and Evelyne Lutton. Evolutionary visual exploration: Evaluation with expert users. In *EuroVis 2013, 15th annual Visualization Symposium*, June 2013. June 17-21, Leipzig, Germany.
- Evelyne Lutton, Alberto Tonda, Sebastien Gaucel, Julie Fouquier, Alain Riaublanc, and Nathalie Perrot. Food model exploration through evolutionary optimization coupled with visualization: application to the prediction of a milk gel structure. In *From Model Foods to Food Models. DREAM Project's International Conference*, June 2013.
- Waldo Cancino, Nadia Boukhelifa, and Evelyne Lutton. Evographdce: Interactive evolution for visual analytics. In *IEEE Congress on Evolutionary Computation, June 10-15, 2012*. June 10-15, Brisbane, Australia.
- Evelyne Lutton and Jean-Daniel Fekete. Visual analytics of ea data. In *Genetic and Evolutionary Computation Conference, GECCO 2011*, 2011. July 12-16, 2011, Dublin, Ireland.
- Thomas, J., Cook, K. (2005): Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE Computer Society
- Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (eds.) (2010): Mastering the Information Age - Solving Problems with Visual Analytics. Eurographics
- von Landesberger, T., Schreck, T., Fellner, D. W., & Kohlhammer J. Expanding the Frontiers of Visual Analytics and Visualization - Chapter: Visual Search and Analysis in Complex Information Spaces - Approaches and Research Challenges Editors: J. Dill, R. A. Earnshaw, D. J. Kasik, J. A. Vince, & P. C. Wong; Springer 2012
- von Landesberger, T., Bremm, S., Andrienko, G. & Andrienko, N., Tekusova, M., Visual Analytics Methods for Categorical Spatio-Temporal Data, IEEE Visual Analytics Science and Technology, to appear 2012
- Commission SFSTP (2011), Le cahier de laboratoire électronique (CLE), STP PHARMA PRATIQUES - volume 21 - N° 6 - novembre-décembre 2011.
- http://www.dsimb.inserm.fr/~debrevrn/PUB/072_SFSTP_ELN_2011/Amiard_STP_PharmaPratique_2011.pdf
- E. Guettala (2013), "VizAssist : un assistant utilisateur pour le choix et le paramétrage des méthodes de fouille visuelle de données", thèse de l'université de Tours, Polytech Tours - Département Informatique, 5 septembre 2013.
- Packham, I.S.J., Rafiq, M.Y., Borthwick, M.F. and Denham, S.L. (2005), Interactive visualisation for decision support and evaluation of robustness—in theory and in practice, *Advanced Engineering Informatics*, Volume 19, Issue 4, October 2005, Pages 263-280,