

Projets Réseaux et Graphes (M2 Machine

Learning) severine.affeldt@u-paris.fr

1 Présentation

L'objectif de ce mini-projet est l'exploitation des différents outils de reconstruction et d'analyse de réseaux que vous avez vu en cours afin d'étudier un système complexe à partir de données disponibles en ligne.

Parmi les étapes importantes du projet, il vous faudra notamment :

- Etudier les articles proposés et comprendre les problématiques associées au projet choisi
- Adapter les datasets à la problématique retenue (choix des variables, discrétisation des data...)
- Effectuer une étude comparative des résultats obtenus à partir des différents outils de reconstruction (Aracne, constraint-based, search & score...)

Un rapport est attendu à la fin de ce projet. Il devra comporter une présentation générale de la problématique étudiée, des explications concernant le choix des datasets et leurs modifications éventuelles et les résultats commentés issus des reconstructions et analyses de réseaux.

2 Projets (groupe de 2 à 4 étudiants !)

2.1 Projet Tumorigénèse

De nombreux réseaux biomoléculaires assurent le bon fonctionnement des processus vitaux de l'organisme. Les mutations de gènes intervenant dans ces réseaux peuvent conduire à des dérégulations à l'origine de maladies génétiques complexes telles que les cancers. De nombreuses études cherchent à identifier ces gènes mutés, en particulier les gènes intervenant en amont des cascades d'interactions, afin de définir de nouvelles cibles thérapeutiques.

Toutefois, il peut exister une diversité importante de mutations entre les tumeurs de différents patients atteints d'un même cancer, et le tissu d'une même tumeur peut également présenter une grande hétérogénéité. En effet, les dérégulations permettant la croissance tumorale peuvent trouver leur source dans des chemins de signalisation distincts, ou à divers niveaux de ces mêmes chemins.

Malgré cette hétérogénéité intra- et inter- patients, les cellules à l'origine de la tumeur ont besoin pour se développer de franchir successivement certaines barrières communes qui induisent et maintiennent une prolifération non-contrôlée. Ainsi, il semble dans un premier temps avantageux pour une (future) cellule tumorale d'acquérir des mutations réduisant sa probabilité d'apoptose et faisant croître son taux de division. Dans un deuxième temps, les mutations favorisant la non vérification des erreurs dans le génome et/ou la non réparation de celles-ci sont également favorables au développement tumorale. Par la suite, la fixation de mutations engendrant une angiogénèse importante, ainsi qu'une insensibilité aux anti-facteurs de croissance est favorisée. Enfin, l'étape de métastase ne pourra avoir lieu que si les cellules du tissu tumoral primaire acquièrent la capacité à envahir d'autre tissus.

2.1.1 Projet 1 : Cancer du sein

Pour le cancer du sein, on distingue au moins quatre sous-types de cancer : luminal A, luminal B, basal-like et HER2. Tandis que les deux premiers sous-types de cancers sont la conséquence du dérèglement de récepteurs hormonaux (oestrogène et/ou progestérone), et peuvent bénéficier de thérapies hormonales, on ne trouve pas de dérégulations au niveau de ces récepteurs dans les sous-types basal-like et HER2. Les traitements utilisés sont alors plus lourds et parfois peu efficaces. Ce projet propose d'étudier des datasets comportant des informations relatives aux mutations identifiées dans des échantillons

de tissus tumoraux ainsi que le niveaux d'expression et le nombre de copies des gènes, ceci afin de reconstruire les chemins de mutations et/ou les interactions entre les gènes participant à la croissance tumorale dans les différents sous-types de cancer du sein.

- Articles proposés
 - [Hallmarks of Cancer: The Next Generation \(Cell 2011, Hanahan et al.\)](#)
 - [The clonal and mutational evolution spectrum of primary triple-negative breast cancers \(Nature 2012, Shah et al.\)](#)
 - [Sequence analysis of mutations and translocations across breast cancer subtypes \(Nature 2012, Banerji et al.\)](#)
 - [The landscape of cancer genes and mutational processes in breast cancer \(Nature 2012, Stephens et al.\)](#)
 - [Comprehensive molecular portraits of human breast tumours \(Nature 2012, Koboldt et al.\)](#)
- Datasets possibles
 - [cBioPortal](#) : Plusieurs datasets sur le cancer du sein avec des informations de mutations. NB : choisi les listes de gènes prédéfinis par l'interface cBioPortal.
 - [COSMIC](#) : Plusieurs datasets sur le cancer du sein avec des informations de mutations. NB : choisi les listes de top gènes via l'interface COSMIC.

2.1.2 Projet 2 : Glioblastome

Le glioblastome est la tumeur primitive du cerveau la plus fréquente et la plus agressive. Les traitements actuels ne permettent pas la guérison, et l'espérance de vie à cinq ans de cette maladie a peu évolué ces trente dernières années. Ce projet propose d'étudier des datasets comportant des informations relatives aux mutations identifiées dans des échantillons des tissus tumoraux ainsi que le niveaux d'expression et le nombre de copies des gènes, ceci afin de reconstruire les chemins de mutations et/ou les interactions entre les gènes participant à la croissance tumorale de ce type de cancer.

- Articles proposés
 - [Hallmarks of Cancer: The Next Generation \(Cell 2011, Hanahan et al.\)](#)
 - [Comprehensive genomic characterization defines human glioblastoma genes and core pathways \(Nature 2008, McLendon et al.\)](#)
 - [The somatic genomic landscape of glioblastoma \(Cell 2013, Brennan et al.\)](#)
- Datasets possibles
 - [cBioPortal](#) : Plusieurs datasets sur le glioblastome avec des informations de mutations. NB : choisi les listes de gènes prédéfinis par l'interface cBioPortal.
 - [COSMIC](#) : Plusieurs datasets sur le glioblastome avec des informations de mutations. NB : choisi les listes de top gènes via l'interface COSMIC.

Attention : les datasets ont peu d'échantillons, pensez à les combiner !

2.2 Projet 3 : Survie post opératoire des patients atteints d'un cancer du poumon

Le cancer du poumon est une des plus importantes cause de mortalité en France. Il est notamment causé par le tabagisme passif ou actif (9 cas sur 10) et est en général très agressif. L'une des options de traitement est l'ablation chirurgicale de la partie touchée du poumon ; elle permet théoriquement de stopper la propagation du cancer, et d'éviter, si l'intervention a lieu suffisamment tôt, que le cancer ne produise des métastases. Retirer une partie du poumon constitue cependant une prise de risques importante, et le bénéfice en terme de survie engendré par une telle procédure se doit par conséquent d'être avéré.

Ce projet propose donc d'étudier un jeu de données rassemblant les observations réalisées sur 470 patients ayant subi ce type de chirurgie entre 2007 et 2011. L'enjeu majeur de sa création était la prédiction, selon les variables définies, de l'espérance de vie des patients après l'intervention. Cependant, étant donné le nombre de variable observées par patient ainsi que leur relative simplicité,

la reconstruction d'un réseau mettant en jeu tous ces paramètres peut s'avérer pertinente notamment pour comprendre les relations indirectes entre eux.

2

- Articles proposés
 - Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients (Applied Soft Computing 2014, Zieba et al.)
- Datasets possibles
 - [UCI Machine Learning Repository: 1 dataset](#) (d'autres articles de référence sont également indiqués)

Pour ce projet, il est également proposé d'analyser le réseau causal inféré via une méthode bayésienne exacte : [REBEL](#). L'article correspondant est disponible en ligne : [Advances in Exact Bayesian Structure Discovery in Bayesian Networks](#).

2.3 Projet 4 : Prise en charge des patients diabétiques

Les protocoles d'analyse des taux glycémie de patients présentant des symptômes liés au diabète dans les centres de soins intensifs permettent une détection rapide et une prise en charge efficace des malades. Toutefois, les centres hospitaliers non spécialisés suivent des protocoles d'admission classiques pouvant conduire à la non détection et/ou à une mauvaise prise en charge des patients diabétiques, impactant significativement leur état de santé et augmentant leur chance de réadmission. Une modification des protocoles traditionnels d'admission dans les centres non spécialisés permettrait une prise en charge plus efficace des patients diabétiques.

Ce projet cherche à identifier les facteurs induisant une mauvaise prise en charge des malades, et donc une réadmission rapide, afin d'établir des protocoles efficaces de détection et de suivi des patients diabétiques.

- Articles proposés
 - [Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records](#) (Strack et al. 2014, Biomed Res Int)
- Datasets possibles
 - [UCI Machine Learning Repository: 1 dataset](#) (d'autres articles de référence sont également indiqués)

