

Récap cours 2023-2024

Carlos Ramisch/Manon Scolivet LIS : Recherche zen : éviter de stresser pour nos choix méthodologiques (débattables) OU Méthodologie expérimentale en informatique

Giuseppe Di Molfetta LIS : UE Quantum Algorithms: basics and principles

Stéphane Ballet I2M : Histoire de la pensée scientifique

Ctirad Klimcik I2M : la théorie conforme de champs

Olivier Dudas I2M : Théorie de Deligne--Lusztig

Carlos Ramisch/Manon Scolivet LIS : Recherche zen : éviter de stresser pour nos choix méthodologiques (débattables) OU Méthodologie expérimentale en informatique

Objectifs : Cette formation porte sur la méthodologie, les pratiques, les pièges à éviter etc. en recherche expérimentale en informatique, notamment dans des domaines liés à la science des données, IA, apprentissage, TAL... Le parti pris du cours est de s'appuyer systématiquement sur des exemples concrets, des situations réelles ou réalistes, pour ensuite aborder des notions plus abstraites de méthodologie scientifique. Chaque séance comporte des activités et exercices pratiques dont le but est de (a) rendre agréable le thème de la méthodologie scientifique, souvent considéré comme mineur ou ennuyeux, (b) justifier l'importance des notions abstraites via des exemples concrets, et (c) s'entraîner sur des compétences pratiques essentielles au travail scientifique, telles que la structuration de questions et hypothèses de recherche, la conception d'une expérience, la présentation de résultats, etc. L'objectif global du cours est de construire collaborativement un idéal de la méthodologie de recherche en science des données, et de le mettre en perspective par rapport aux pratiques actuelles, tout en nuanciant la morale binaire de la "bonne / mauvaise" recherche. Les notions et compétences développées dans ce cours devraient aider les participant.e.s à faire évoluer leurs pratiques pour tendre vers cet idéal. Nous voyons ce cours comme un moment privilégié pour étudier des notions que, d'habitude, nous utilisons sans comprendre et dans l'urgence, parce que nous n'y accordons pas assez d'importance.

Volume horaire : 24h

Approche expérimentale usuelle en recherche en informatique.

Vue idéale proposée dans ce cours de la recherche expérimentale en informatique.

1) Questions et hypothèses de recherche

Dans cette première partie du cours, nous aborderons les questions de la démarche scientifique, du rôle de l'expérience et de l'empirisme dans la recherche scientifique, pour aboutir à la définition de questions et hypothèses de recherche, point de départ de toute contribution scientifique. Cette section aborde des thèmes liés à l'épistémologie, l'histoire et la philosophie des sciences en s'appuyant sur une approche participative et sur des supports pédagogiques et accessibles, plutôt que sur des ouvrages académiques. Le but est de

susciter une réflexion critique sur la pensée scientifique, et de faire émerger l'importance des questions et des hypothèses qui orientent un travail de recherche scientifique (p.ex. thèse, article, stage).

Notions : pensée scientifique, démarche scientifique, expérience, démarche empirique, question de recherche, hypothèse de recherche, caractérisation des "bonnes" et "mauvaises" formulations des questions et hypothèses de recherche.

Compétences : développer un esprit critique sur le travail scientifique : son contour, son rôle, ses limitations. Formuler des questions de recherche pertinentes sur un sujet de son choix (p.ex. sujet de thèse) de manière claire, précise, directe et structurée, avec une ou plusieurs hypothèses associées à chaque question et sous-question.

2) Contextualisation et bibliographie

Poser clairement des questions et hypothèses de recherche ne garantit pas la qualité du travail scientifique. D'abord, il faut justifier l'intérêt et l'importance des questions de recherche

par rapport à une application, un contexte, une avancée potentielle des connaissances dans un domaine. De la même manière, il faut également justifier de la plausibilité des hypothèses associées aux questions. Souvent, pour arriver à ces justifications, il est important de savoir si d'autres personnes se sont déjà posées ces mêmes questions, ou des questions similaires, et quel est l'état actuel de la connaissance (ou "l'état de l'art") dans ce domaine. En somme, une recherche bibliographique s'impose. Il existe des outils et des techniques pour structurer ce travail, mettre en perspective des résultats, et/ou identifier des biais dans les articles.

Notions : justification d'une question/hypothèse, état de l'art, recherche bibliographique, portée (ce qui est couvert et pas couvert par la recherche bibliographique), revue de la littérature (méta-analyse ou survey), lecture critique d'articles, niveau de preuve scientifique.

Compétences : contextualiser une question de recherche par rapport à un besoin, une limitation identifiée, une application potentielle, etc. Naviguer dans des bases d'articles scientifiques (p.ex. ArXiv). Faire une bibliographie, un survey sur une question de recherche. Se positionner par rapport aux aspects méthodologiques d'un travail de recherche. Manipuler des outils d'organisation de bibliographie (p.ex. Zotero).

3) Expériences

Il n'est pas rare d'arriver à la fin d'un travail de recherche (p.ex. la rédaction d'un article), et de se dire alors qu'il aurait fallu refaire toutes les expériences différemment. Ces regrets sont évitables si l'on prend le temps de bien concevoir le plan d'expériences (ou cadre expérimental), car la création de LA bonne expérience est au coeur de la recherche expérimentale. Comment créer une expérience qui permettra de vérifier les hypothèses posées ? De la conception de l'expérience aux méthodes d'évaluation, en passant par la question de l'obtention et préparation de données (annotées ou non), les différentes étapes pour bien construire les expériences sont nombreuses. À ce moment, des limitations des ressources à disposition peuvent nécessiter de revoir les questions et hypothèses de recherche. Il est important d'identifier où sont les pièges à éviter, et quelles sont les questions qu'il ne faut pas oublier de se poser. Dans ce module, nous aborderons ces questions sous l'angle de la science des données et du machine learning, sans pour autant détailler les modèles sous-jacents spécifiques à chaque travail de recherche.

Notions à transmettre : plan d'expériences (ou cadre expérimental), la notion d'expérience appliquée au machine learning, jeux de données, supervision, annotation,

préparation de données (nettoyage, découpage train/dev/test), qualité des données (accord inter-annotateur.e, histogramme), conditions expérimentales (supervisé, zéro-shot, few-shot), hyper-paramétrage, gestion des expériences, carnet d'expériences ou logbook, mesures d'évaluation des prédictions (précision, rappel, F-mesure), généralisation, biais de mesure.

Compétences : analyser une question de recherche très finement (découper la montagne). Concevoir des expériences pertinentes pour répondre aux questions et hypothèses de recherche.

Concevoir des questions et hypothèses de recherche raisonnables par rapport aux possibilités expérimentales. Récupérer, pré-traiter, adapter, étendre et/ou créer des jeux de données adaptés aux expériences prévues. Organiser et documenter ses expériences pour gérer des conditions expérimentales nombreuses et complexes. Identifier, appliquer, adapter ou concevoir des mesures d'évaluation appropriées.

4) Analyses

Dans l'absolu, le résultat d'une expérience est une métrique dont la valeur permet de valider ou d'invalider une hypothèse. En pratique, les choses sont bien plus complexes. Chaque condition expérimentale (jeu de données, modèle, hyper-paramétrage ...) constitue une sous-expérience, souvent avec plusieurs métriques associées. La combinatoire des conditions expérimentales et des métriques peut mener à des milliers de résultats dont l'analyse requiert de la patience, de la rigueur, et des outils adaptés. Cette partie du cours présente certains outils d'analyse et de visualisation de résultats ainsi que des notions statistiques utiles pour découvrir des tendances (entropie, corrélation) et pour effectuer des comparaisons (significativité, p-value). Une analyse hâtive peut mener à des conclusions fausses (dissonantes vis-à-vis des hypothèses, ou pas appuyées par les résultats). L'absence d'analyse mène à des appréciations subjectives ("la performance du modèle est bonne") ou à des résultats bruts, passant sous silence le manque de conclusions. L'objectif de ce module est de présenter des outils pour éviter ces pièges rédhibitoires et ne pas se noyer dans le chaos des résultats. Quand il n'est pas possible de tirer des conclusions (résultat "négatif"), l'analyse permet de trouver de nouvelles questions de recherche.

Notions : manipulation de résultats (p.ex. pandas), entropie, variance, covariance, corrélation, intervalle de confiance, p-value, tests de significativité statistique (paramétriques, non paramétriques, tests approximatifs par ré-échantillonnage), interprétabilité, explicabilité.

Compétences : extraire des tendances et savoir résumer des résultats à l'aide d'outils statistiques. Énoncer des conclusions appuyées par l'évidence empirique. Formuler des hypothèses de recherche pour lesquelles on peut tirer des conclusions en fonction des résultats observés. Comparer des résultats d'expérience à l'aide d'outils statistiques. Identifier des pièges récurrents dans l'analyse des données (p.ex. corrélation \neq causalité). Dédire de nouvelles (sous-)questions de recherche à partir de l'analyse des résultats d'une expérience. Valoriser des "résultats négatifs" en déduisant des nouvelles questions de recherche plus pertinentes, plus détaillées, mieux formulées.

5) Communication

La communication des résultats de recherche a lieu sous la forme d'un rite ancestral : la publication d'articles scientifiques. Il s'agit d'une étape cruciale, souvent source de stress face au spectre impitoyable de la relecture par les pairs. L'art de la communication scientifique requiert une certaine expérience, mais la maîtrise des outils peut faire gagner du

temps. Ce module aborde la structure, la forme et le contenu d'un document (article, mémoire, thèse) et d'une présentation (exposé oral, poster, vidéo) dans le contexte scientifique. Le principal défi consiste à utiliser le langage humain (imprécis) pour transmettre des notions complexes et abstraites de manière claire, précise, pédagogique et intéressante. Pour cela, on peut s'appuyer sur des outils tels que les tableaux, graphiques, schémas, équations et exemples. Ce module comportera des exercices pratiques de relecture critique et de mise en forme de résultats bruts. Des outils de gestion du temps et des astuces contre les syndromes de la page blanche et de l'imposteur complètent le répertoire d'instruments de navigation pour s'y retrouver dans un monde compétitif où ArXiv et twitter semblent remettre en question le modèle de publication scientifique. Notions : plateformes en ligne d'écriture collaborative (overleaf), structure d'un article scientifique (introduction, état de l'art, méthode, résultats, discussion, conclusions), relecture.

Compétences : extraire les contributions principales d'un travail et faire abstraction des détails. Concevoir et réaliser des tableaux, graphiques et schémas lisibles et pédagogiques dans les règles de l'art (nombre de décimales, légendes, couleurs). Formaliser une notion complexe par des équations, des symboles. Trouver des exemples illustratifs et les utiliser à bon escient pour communiquer sur sa recherche. Faire des diapositives et des posters agréables à visualiser.

6) Pratiques éthiques

Une publication marque souvent l'aboutissement d'une recherche scientifique, mais n'en est pas le point final. Aujourd'hui, une grande partie des revues et des conférences ont des pré-requis de reproductibilité (dans les mêmes conditions) et de répliquabilité (dans des conditions similaires). De plus, un travail aura plus d'impact si les données, le code-source, voire les modèles, sont mis à disposition de la communauté scientifique sous des licences libres (p.ex. Creative Commons). Dans une optique de science ouverte, ces ressources doivent être documentées (p.ex. data sheets), anonymisées, publiées dans des dépôts publics garantissant la pérennité des URLs, suivre les principes FAIR, être stockées de manière sécurisée, avec des sauvegardes, et dans le respect de la RGPD. Le travail scientifique a des aspects éthiques et n'est pas sans impact pour la société : il comporte des biais, il a un coût énergétique, et il peut influencer la politique. Une bonne prise en compte de ces aspects devrait minimiser les biais et guider le choix des sujets de recherche vers une science plus intègre et consciente.

Notions : reproductibilité, répliquabilité, licence, dépôt de données, anonymisation, principes FAIR, RGPD, science ouverte, impact socio-environnemental de la recherche, data sheets, model sheets.

Compétences : publier les données, code-source et modèles de sa recherche sur des dépôts publics. Choisir une licence et un mode de partage pour ses ressources. Connaître et appliquer la réglementation sur la gestion des données personnelles. Documenter et décrire une expérience pour la rendre la plus reproductible possible. Avoir un regard critique sur les biais présents dans sa recherche et son impact dans la société.



Giuseppe Di Molfetta

UE Quantum Algorithms: basics and principles

Language : English

Matin : 6 séances 2H00 CM

Après-midi : 4 séances de TP de 2h chacun

Totale = 24h TD (idéalement à tenir en 2 semaines)

note: 1H CM = 1h30 (h TD)

Objectif (EN) : This is a set of lecture notes on quantum algorithms. The content will span from fundamentals to advanced topics, trying to give a large view on the current state of art in the domain.

Content (EN)

I. Quantum circuits :

1 Preliminaries

2 Efficient universality of quantum circuits

3 Quantum circuit synthesis over Clifford+T (TP 1 : Introduction Qiskit + implementation Clifford circuit)

II. Quantum algorithms for algebraic problems

4. The abelian quantum Fourier transform and phase estimation (TP 2 : Implementation of a QPE algorithms)

III. Quantum walk

5 Continuous-time and Discrete quantum walk

6 Unstructured search

7 Quantum walk search (TP 3 : Implementation Spatial Search over 2D grid)

IV. Quantum query complexity

8 Query complexity and the polynomial method

9 The collision problem

10 The quantum adversary method (TP 4 : Learning graph)

V. Quantum simulation

11 Simulating Hamiltonian dynamics. Examples from physics and chemistry

VI. Adiabatic quantum computing

12 The quantum adiabatic theorem

13 Adiabatic optimization

.....

Stéphane Ballet : Histoire de la pensée scientifique

Programme :

L'objet de ce cours est de donner des éléments de compréhension de la genèse des grands principes de la science moderne et plus généralement du processus de structuration de la science - les origines et la genèse de la science moderne - au travers son évolution du Moyenâge jusqu'à la Renaissance. Le but est d'inciter le futur chercheur à une démarche

réflexive visant à s'interroger sur la nature et la valeur des principes, des concepts, des méthodes et des résultats des sciences.

Bibliographie

[1] Gaston Bachelard. La formation de l'esprit scientifique. Bibliothèque des textes philosophiques, Vrin, 2011.

[2] Thomas Khun. La structure des révolutions scientifiques. Champs sciences, Flammarion, 2008.

[3] Alexandre Koyré. Etudes d'histoire de la pensée scientifique. Gallimard, 1973.

.....

Ctirad Klimcik : la théorie conforme de champs

La théorie conforme des champs est une discipline qui se place sur la frontière entre les mathématiques et la physique théorique et qui se distingue par son énorme richesse structurelle ainsi que pour sa beauté ; pour la maîtriser il faut avoir des connaissances importantes à la fois en géométrie, algèbre et analyse. Bien qu'il soit impossible d'exposer aux étudiants plus qu'une petite fraction de ce qui est connu dans la matière il s'agit néanmoins d'une aventure intellectuelle qui vaut la peine à suivre. La théorie conforme des champs est avant tout une représentation d'une certaine structure algèbro-géométrique assez complexe appelée l'algèbre de vertex. L'ingrédient technique important pour comprendre cette structure et le soudage conforme des surfaces de Riemann à bord. Le soudage conforme des surfaces de genre zéro avec deux bords d'orientations opposées définit une sous-structure appelée le semi-groupe de Neretin-Segal dont la théorie de représentations est intimement liée avec celle du groupe de Virasoro. Tout ce formalisme aboutit au final à la construction de fonctions de corrélation de certaines théories des champs quantiques utilisées par exemple dans la théorie des cordes.

Les thèmes abordés dans le cours seront :

- 1) Les champs quantiques
- 2) Symétrie de Weyl et symétrie conforme
- 2) Les identités de Ward version Eguchi-Ooguri
- 3) Quantification radiale et reconstruction d'Osterwalder-Schrader
- 4) Operator product expansion, fusion,
- 5) L'algèbre de Virasoro, ses représentations
- 6) Les surfaces de Riemann, le soudage conforme, les axiomes de Segal, algèbres de vertex, relation de Borchers

.....

Olivier Dudas : Théorie de Deligne--Lusztig

Cette théorie, développée à la fin des années 70 a pour but de construire et d'étudier les représentations linéaires d'une certaine classe de groupes finis appelés "groupes réductifs finis" ou "groupes finis de type de Lie" (par exemple $GL_n(q)$, $SL_n(q)$, $Sp_{2n}(q)$, ..., $E_8(q)$)

Pourquoi ces groupes ? D'après la classification, les groupes finis simples se répartissent en 4 familles :

- les groupes cycliques d'ordre premier,
- les groupes alternés,
- 26 groupes sporadiques (inclassables !),
- les groupes finis simples de type de Lie (très proches des groupes réductifs finis).

Cela montre que la "majorité" des groupes simples sont de type de Lie. De nombreux résultats fondamentaux sur les groupes finis dépendent de la classification et donc de notre compréhension de cette classe de groupe, en un sens la plus difficile.

Pourquoi la théorie de Deligne--Lusztig ? Les méthodes algébriques pour construire les représentations des groupes réductifs finis atteignent très vite leur limite (déjà pour $SL(2,q)$...). Néanmoins, si on reconnaît que ces groupes sont les équivalents finis des groupes algébriques réductifs, alors on peut construire leurs représentations à partir des objets géométriques standard en théorie de Lie (espaces projectifs, grassmanniennes, variétés de drapeaux...). C'est ce point de vue, dû à Deligne et Lusztig, que nous adopterions tout au long de ce cours.

Le but serait de donner une introduction à la théorie en partant des exemples concrets de $SL(2,q)$ et $GL(n,q)$. Le plan suivi serait :

- Variétés de drapeaux et variétés de Schubert,
- Variétés de Deligne--Lusztig (propriétés géométriques élémentaires et descriptions explicites pour quelques cas dans des groupes linéaires),
- Cohomologie ℓ -adique, formule des traces,
- Caractères unipotents et lien avec les algèbres de Hecke par la formule du caractère,
- Calculs explicites de cohomologie, notamment pour SL_2 ,
- Conjectures pour les éléments réguliers, et lien avec la conjecture du défaut abélien.

En plus de ces aspects standards contenus dans plusieurs ouvrages de référence, j'aimerais en profiter pour digresser sur certaines notions fondamentales mais très modernes en théorie géométrique des représentations, par exemple :

- les formules de comptage de point à la Lefschetz
- les méthodes de catégorification d'algèbres (de Hecke) via les faisceaux et leur convolution
- l'utilisation de l'algèbre homologique (catégories dérivées) en théorie des représentations

Les prérequis pour le cours sont :

- théorie des représentations des groupes finis (niveau maîtrise)
- bases de géométrie algébrique (variétés affines et projectives)
- bases des groupes algébriques (si possible groupes réductifs).