Rebonjour à tous. Merci d'être ici aujourd'hui pour la réunion CSI.
Je suis en thèse sous la direction de Nicolas, Karim et Rémi. Mon sujet de thèse est Transport optimal pour l'apprentissage par transfert entre les espaces
Dans cette réunion, je vais vous présenter ce que j'ai fait depuis l'année dernière.

Je vais commencer avec une intro rapide de la distance de GW. Plus précisement, on s'intéresse au graphe pondéré qui est muni d'une matrice d'adjacence et d'un histogramme associé aux noeuds du graphe. Etant donné 2 graphes pondérés, on définit la distance de GW entre eux comme le minimum d'un problème d'optimisation. Où on cherche un plan de transport qui minimise cette fonction d'objective. Intuitivement, on veut que les clusters dans le graphe de source soient alignés aux autres dans le graphe de cible, qui se ressemblent le plus (entre guillemets).

On peut aussi associer un feature vecteur à chaque noeud du graphe et on obtient un graphe pondéré labelisé. Maintenant, la FUGW entre 2 graphes pondérés labelisés est définit par le minimum de ce problème. On a 4 termes dans sa fonction objective. Le 1er, c'est le terme de GW qui préserves la géometrie globale de la structure sous accente. Le terme de Wasserstein en bleu prend en compte des alignements de source et de target features. Le terme d'unbalanced permet de ne pas transporter la masse à certaines positions entre les source et cible. Le terme d'entropie n'a que pour l'intérêt d'accélerer l'optimisation.

On a colaboré avec l'equipe MIND à l'INRIA et appliqué FUGW sur les données neuro-imaging. Dans cet expériment, on cherche à aligner les cerveaux de 2 humains et à apprendre une template du cerveau, qui est juste un barycentre. Ici, le cerveau est représenté comme un mesh et la matrice d'adjacence est la distance géodésique sur le mesh. le feature est un vecteur dans R300 dont les coordonées sont les functional activation maps.

On a obtenu des résultats très favorables dans ces expé et notre papier a été accepté au Neurips 2022. En ce moment, on est en train de tester FUGW en grande échelle et on a obtenu des résultats préliminaires assez encourageants.

Maintenant je vais parler du 2è papier accepté l'année dernière. D'abord, on définit la notion de sample-feature space. Ici, on a 2 espace mesuré compact, l'espace de sample Xs muni d'une measure mu_s et l'espace de feature Xf muni d'une mesure mu_f. On introduit aussi l'intéraction xi, qui est juste une fonction intégrable à valeur réelle définit sur l'espace produit et on apelle ce triple un sample-feature space. Par exemple, dans le cas discret, on peut voir l'intéraction xi comme la coordonée de la matrice en entrée. Et Xs, Xf sont simplement les indices des lignes et des colonnes.

Maintenant, étant donné 2 sample-feature space, on peut définir le UCOOT entre eux comme le minimum de ce probleme d'optim, où on cherche simultanément 2 plans de transport qui minimisent cette fonction objective
Son terme d'intégrale représent le cout de transport qui est la différence élement-wise entre 2 intéractions. La 2è somme se ressemblent bien à celle dans FUGW, qui dit en gros, on doit pas forcément transporter toutes les masses du source vers le cible.
On peut montrer que ce problème admet toujours une solution. La propriété la plus importante de UCOOT, c'est qu'il est très robust aux bruits et aux anomalies.

Pour résumer, ma contribution, j'ai formalisé le framework et montré tous les résultats théoriques. J'ai aussi réalisé l'expériment sur l'adaptation de domaine hétérogène. On a aussi colaboré avec l'équipe biologistes à l'université de Brown et réussi à appliquer UCOOT sur les données multi-omics. A la fin, notre papier a été accepté au AAAI.

Donc, ce sont 2 projets réalisés l'année dernière. Maintenant je vous parle des projets les plus récents. Le 1er, c'est sur une variation de la distance de GW. On a 2 motivations principales. On sait que la distance de GW est invariante par transformation isometrique, car la matrice de distance n'est toujour pas impactée dans ce cas. Mais ça veut aussi dire qu'il y a beaucoup des isométries et qu'elles n'ont pas forcément les même sens, par exemple le chiffre 6 et 9 sont géometriquement équivalent mais apparament ils représentent des chiffres différents. En plus, dans GW, tous les features sont aggrégés dans la distance, donc elles ne sont pas bien exploités. Et là où on a besoin de COOT.

Donc, on propose de combiner COOT et GW et qu'on appelle cet combinaison Augmented GW. Comme COOT, on cherche 2 plans transport qui minimisent une combinaison convex de GW et COOT, un peu près. Intuitivement, si alpha tend vers 0, alors on approche COOT et quand alpha tend vers 1, on approche GW. On peut aussi montrer que AGW s'intéresse à vachement moins des isométries que GW et souvent dans nos expériments, ce sont des bonnes.

Pour résumer ma contribution, j'ai formalisé la framework et réalisé toute l'analyse théorique de cette divergence. En colaboration toujours avec l'équipe à l'université de Brown, on a obtenu des résultats très favorables avec AGW et on a soumis ce papier à Neurips la semaine derniere.

Il nous reste à bien comprendre les isométries concernées par AGW. Je dirais qu'en ce moment, on comprend 70% comment elle marche. Il nous faut aussi comparer avec les autres méthodes récemment proposées.

Le 2è projet que j'ai commencé la semaine dernière, c'est sur le joint distribibution COOT. Notre motivation est la suivante. Supposons on a des données de source labelisés et de cible non-labelisés, et que les données de source et de cible se trouvent dans le meme espace. Une tache classique en machine learning, c'est d'entrainer un classifier qui marche bien sur les données de cible, en utilisant les données de source labelisées. Il y a un travaux qui résoud ce problème à l'aide du transport optimal. Cet approche est appelée Joint distribution OT, où on cherche à apprendre un classifier qui minimise la distance de Wasserstein entre les données jointe de source et de cible estimés.

Notre intérêt pricipal dans ce projet, c'est le cas où le source et cible ne sont plus comparables, par exemple les points dans R3 versus les points dans R5. Grace au succès de COOT, on propose de remplacer Wasserstein par COOT et on introduit le Joint distribution COOT. On peut aussi tenter qqc de plus puissante. On l'appelle Deep-JDCOOT, où on veut en plus projeter les données dans les espaces latents en espérant que le classifier marche encore mieux sur ces espaces. Typiquement, les fonctions g ici sont les réseaux neurones. C'est pour ça on l'appelle Deep-JDCOOT.

Le projet a déjà commencé il y a quelques temps, donc le papier est déjà pas mal dirigé. Le code et les expériments sont implémentés par un ancien doctorant mais les résultats des expé ne sont pas encore communiqués dans le papier. Donc, ce qu'il me reste à faire, c'est de finaliser la rédaction du papier, relancer les experiments et revérifier le résultat théorique.

Donc pour résumer, depuis le début de thèse, j'ai 3 papiers accepté et 1 papier récemment soumis. Voici la liste des formations suivie. Pas bcp de nouveau en comparant avec l'année dernière. Merci pour votre attention