

Project SCATE - Evaluation 220987

QUALITY AND SCIENTIFIC AIM

- a - Clarity of research objectives and hypotheses
 - b - Scientific ambition of the project and position in relation to the state-of-the-art
- [Added value of the project in terms of scientific contribution - scope, problem and methodological approach - and in terms of knowledge production]
- c - Adequacy and relevance of the methods implemented
- [« Relevance » is also understood in terms of ethics, scientific integrity and social responsibility of the sciences - and as such, taking into account the sex and/or gender aspect -, of disciplinary coverage (mono-trans-inter-disciplinarity) and of scientific risk management. “Methods” also includes Open Science practices, namely: data management, reuse of existing data sets, development or contribution to open source software, standards, and adopting permanent identifiers for all research products]

COMMENT

(a)

Une description plus structurée, complète et précise des objectifs scientifiques dans le paragraphe 1.6 ("Objectives") aurait bénéficié l'identification et la compréhension de ceux-ci. En particulier, le descriptif détaillé des WP de la tâche B évoque la conception de nouveaux algorithmes à temps de calcul optimal pour les systèmes hétérogènes, au lieu d'une simple extension des algorithmes proposés dans la tâche A (qui traite des systèmes homogènes), comme décrit dans le paragraphe 1.6. Il est donc nécessaire de parcourir tout le document avant de bien comprendre la portée des objectifs fixés.

En ce qui concerne les hypothèses, le projet semble supposer qu'une implémentation parallèle (efficace) de l'approche classique de troncature de la HOSVD (THOSVD) doit forcément être plus performante que l'approche séquentiellement tronquée (STHOSVD). Bien que celle-ci se prête moins bien à la parallélisation, il est bien connu qu'elle exige un nombre total d'opérations pouvant être bien inférieur par rapport à la THOSVD. A priori, les arguments données ne suffisent donc pas pour établir la pertinence de cette hypothèse. Le document évoque aussi le fait que les coûts sont dominés par le calcul de l'étape Multi-TTM dans les deux approches. Or, cette étape n'est pas présente sous la même forme dans la STHOSVD, vu que le cœur de la décomposition résulte simplement des projections réalisées pour réduire la dimension tout au long du calcul séquentiel.

Enfin, il est à noter que pour calculer une CPD de rang faible d'un tenseur de grandes dimensions, typiquement on réalise d'abord une projection à l'aide d'une THOSVD (ou STHOSVD) afin de réduire les dimensions du tenseur (ce qui réduit la taille des facteurs de la CPD calculée). Cela affaiblit l'intérêt pratique d'un algorithme parallèle pour la réalisation d'un MTTKRP avec des matrices de très grande taille.

(b)

Le projet énonce un certain nombre d'objectifs qui peuvent certainement produire de nouvelles approches algorithmiques utiles pour les problèmes ciblés. Toutefois, il n'est pas tout à fait clair à quel point le programme proposé représente une avancée méthodologique importante en rupture avec l'état de l'art, et non pas simplement une application de l'approche déjà proposée par le porteur dans [4], surtout pour la tâche A. Plus d'accent aurait pu être donné à cet aspect dans la proposition.

Par ailleurs, le programme proposé manque d'un peu d'ouverture vers d'autres problèmes similaires où les techniques étudiées pourraient être également utiles. En particulier, l'opération MTTKRP n'est pas employée que dans le calcul d'un modèle de Tucker, étant un élément central du calcul numérique tensoriel.

Quant au positionnement par rapport à l'état de l'art, on constate l'absence de comparaison entre l'approche proposée dans le projet et les approches comme [R1] et [R2], basées sur l'utilisation de techniques de sketching pour réaliser la CPD d'un tenseur de grandes dimensions de manière efficace. En particulier, [R1] permet de distribuer le calcul en réalisant plusieurs CPD de tenseurs plus petits, et puis en combinant les résultats. Il conviendrait de préciser comment l'approche proposée se situe par rapport à ce travaux (sont-elles concurrentes, complémentarites ?).

[R1] Yang, Bo, Ahmed Zamzam, and Nicholas D. Sidiropoulos. "Parasketch: Parallel tensor factorization via sketching." Proceedings of the 2018 SIAM International Conference on Data Mining, 2018.

[R2] Ma, Linjian, and Edgar Solomonik. "Fast and accurate randomized algorithms for low-rank tensor decompositions." Advances in neural information processing systems 34: 24299-24312, 2021.

(c)

L'objectif énoncé dans le paragraphe 1.6 est de concevoir des algorithmes dont le temps de calcul est optimal, ce qui comprend à la fois les temps dû à la communication et au calcul. Par exemple, les livrables du WPA3 sont des algorithmes censés être optimaux selon ce critère. Or, les problèmes d'optimisation évoqués (par exemple dans le WPA1 et le WPA2) ne portent que sur le coût de communication. Bien que les compromis entre les coûts de communication et de calcul seraient aussi étudiés, la façon dont les résultats obtenus permettraient de garantir l'optimalité des algorithmes conçus au sens souhaité (et non pas simplement optimalité par rapport au coût de communication) n'est pas claire.

ORGANISATION AND IMPLEMENTATION OF THE PROJECT

- a - Skills, expertise and involvement of the scientific coordinator
- b - Contribution to the coordinator’s level of responsibility and team development
- c - Adequacy of implemented and requested means to the project’s objectives

Warning:The French National Research Agency (ANR) has signed the San Francisco Declaration on Research Assessment (DORA). Consequently, all the results of research work must be considered (scientific publications, data sets, software, etc.). The use of bibliometric indicators such as the impact factor and the h-index must be banned in favor of qualitative indicators on the works, such

as their influence on policies and practices.

COMMENT

(a)

Les compétences et l'expertise du coordinateur sont bien adaptées au projet, notamment compte tenu de ses résultats préliminaires dans [4].

On constate un plus forte implication relative du coordinateur dans la tâche B menée par le(a) chercheur(se) postdoctorant(e), au sens d'un plus grand ratio entre son implication et celle du(de la) chercheur(se), par rapport au même ratio calculé pour la tâche A, réalisée par le(a) doctorant(e). Par exemple, ce ratio vaut 2/7 et 3/11 dans les WPA3 et WPA4, respectivement, et 2/2, 4/7, 2/3, 6/8 et 4/4 pour les WPAB1 à WPAB5. A priori cela représente un certain déséquilibre, compte tenu du fait qu'un niveau d'autonomie plus élevé est normalement attendu d'un(e) chercheur(se) postdoctorant(e), par rapport à un(e) doctorant(e).

(b)

Les bénéfices en terme de formation d'une équipe et d'un réseau de coopération semblent clairs et suffisants.

(c)

Les moyens demandés sont en bonne cohérence avec les objectifs fixés.

IMPACT AND BENEFITS OF THE PROJECT

- a - Scientific impact and potential economic, social or cultural impact
- b - Strategy for the dissemination and exploitation of the results; promotion of scientific, technical and industrial culture

COMMENT

(a)

Compte tenu des éléments évoqués dans les items (a) et (b) de la rubrique "Qualité et ambition scientifique" (notamment l'absence de rupture importante par rapport à l'état de l'art), l'impact scientifique pourrait être modéré.

(b)

La stratégie de diffusion, et en particulier l'initiative de mettre à disposition le documents et le code produit par le projet, est à mon avis très positive.

GENERAL OPINION including the strengths and weaknesses of the project

COMMENT

- Points forts :
- Les contributions du projet pourraient être utiles dans plusieurs domaines scientifiques où on traite des tenseurs de grande taille.
 - Bonne organisation et planification détaillée des tâches à mener et des livrables à produire.
 - Bonne adéquation du sujet du projet aux activités scientifiques du coordinateur.

- Points faibles :
- Pertinence questionnable de certaines hypothèses, comme détaillé dans l'item (a) de la rubrique "Qualité et ambition scientifique".
 - Ambition scientifique potentiellement limitée, comme discuté dans l'item (b) de la rubrique "Qualité et ambition scientifique".
 - Possible inadéquation des méthodes détaillées dans certains WP vis-à-vis des objectifs scientifiques fixés, cf. item (c) de la rubrique "Qualité et ambition scientifique".