Actes de CORIA-TALN-RJCRI-RECITAL 2025
Marseille, France, 30 juin au 4 juillet 2025

Prédiction des préférences et génération de revue personnalisée basées sur les aspects et attention

Ben Kabongo[1] Vincent Guigue[1, 2] Pirmin Lemberger[3]
(1) Sorbonne Université, CNRS, ISIR, 4 Place Jussieu, 75005 Paris, France
(2) AgroParisTech, UMR MIA Paris-Saclay, 22 place de l'Agronomie, 91120 Palaiseau, France
(3) Onepoint, 29 Rue des Sablons, 75016 Paris, France
ben.kabongo@sorbonne-universite.fr, vincent.guigue@agroparistech.fr,
p.lemberger@groupeonepoint.com

RÉSUMÉ

Le filtrage collaboratif alimente de nombreux systèmes de recommandation performants, mais il peine à saisir les interactions fines entre utilisateurs et articles et à fournir des explications claires. Face à la demande croissante de transparence, la génération d'explications textuelles via des modèles de langage est devenue un axe de recherche majeur. Nous proposons AURA, un modèle multi-tâches combinant prédiction de notes et génération de revues personnalisées. AURA apprend simultanément des représentations globales et spécifiques aux aspects en optimisant les notes globales, les notes par aspect et la génération de revues, avec une attention personnalisée. Ces représentations produisent une invite personnalisée qui guide un modèle de langage pour générer la revue finale. Implémenté avec le modèle T5 pré-entraîné et une stratégie de réglage par invite, AURA a été testé sur TripAdvisor et RateBeer. Les résultats montrent qu'il surpasse nettement les modèles de référence, surtout en génération de revues, renforçant ainsi la transparence des recommandations et la satisfaction des utilisateurs.

ABSTRACT

Aspect-based Unified Ratings Prediction and Personalized Review Generation with Attention

Collaborative filtering drives many successful recommender systems but struggles with fine-grained user-item interactions and explainability. As users increasingly seek transparent recommendations, generating textual explanations through language models has become a critical research area. We propose AURA, a multi-task model combining rating prediction with personalized review generation. AURA jointly learns global and aspect-specific representations of users and items, optimizing overall ratings, aspect-level ratings, and review generation, with personalized attention to emphasize aspect importance. These learned representations produce a personalized continuous prompt, guiding a language model to generate the final review. We implement AURA with the pre-trained T5 language model, adopting a prompt tuning strategy, and conduct experiments on two real-world multi-aspect datasets : TripAdvisor and RateBeer. Experimental results demonstrate that AURA significantly outperforms strong baseline models, especially in review generation, underscoring its effectiveness in enhancing recommendation transparency and user satisfaction.

MOTS-CLÉS : Systèmes de Recommandation, Grands Modèles de Langue, Génération d'Explication, Attention Neuronale.

KEYWORDS: Recommender Systems, Large Language Models, Explaination Generation, Neural Attention.

Actes de la 20e Conférence en Recherche d'Information et Applications (CORIA), pages 151–170.

1 INTRODUCTION

De nombreux systèmes de recommandation performants reposent sur le filtrage collaboratif (CF), qui apprend les préférences des utilisateurs et les caractéristiques des articles à partir des données d'interaction (Koren *et al.*, 2009; He *et al.*, 2017). Parmi les méthodes de CF, les modèles à facteurs latents basés sur la factorisation matricielle (MF) (Koren *et al.*, 2009) ont démontré de fortes performances. Cependant, ces modèles s'appuient principalement sur des représentations latentes des utilisateurs et des articles, ce qui limite leur capacité à capturer des interactions fines et conduit à un manque d'interprétabilité et d'explicabilité.

Au-delà de la performance, l'explicabilité est devenue un axe central de la recherche sur les systèmes de recommandation. Les utilisateurs attendent de plus en plus non seulement des recommandations pertinentes, mais aussi des justifications transparentes pour ces recommandations (Zhang *et al.*, 2020). Cela a suscité un intérêt croissant pour l'utilisation de modèles de langage afin de générer des explications textuelles (Ni *et al.*, 2019; Li *et al.*, 2023), y compris sous forme de revues (Dong *et al.*, 2017). L'hypothèse sous-jacente est que les revues des utilisateurs encapsulent à la fois le sentiment global et des opinions fines sur des aspects spécifiques d'un article, ce qui en fait des candidats naturels pour des recommandations explicables. Plusieurs études ont exploré la génération d'explications en utilisant des réseaux de neurones récurrents (RNN) (Dong *et al.*, 2017; Li *et al.*, 2017; Ni & McAuley, 2018) et des Transformers (Li *et al.*, 2021, 2023; Cheng *et al.*, 2023; Shimizu *et al.*, 2024). Cependant, les approches basées sur les RNN ne parviennent pas à exploiter les avantages des modèles Transformer pré-entraînés (Brown *et al.*, 2020; Dubey *et al.*, 2024; Raffel *et al.*, 2020). Par ailleurs, les méthodes existantes basées sur les Transformers montrent souvent des gains de performance limités en raison de stratégies d'adaptation sous-optimales. De plus, la plupart de ces méthodes négligent la modélisation des aspects, essentielle pour capturer les préférences fines des utilisateurs et fournir des recommandations à la fois personnalisées et interprétables.

Les systèmes de recommandation basés sur les aspects (ABRS) répondent à cette limitation en modélisant les opinions des utilisateurs sur des aspects spécifiques d'un article. Ces méthodes extraient ou apprennent des représentations des aspects à partir des revues (Cheng *et al.*, 2018; Chin *et al.*, 2018). Une catégorie de méthodes ABRS exploite l'analyse de sentiments basée sur les aspects (ABSA) (Zhang *et al.*, 2022) pour extraire des opinions au niveau des aspects (Bauman *et al.*, 2017; Chen *et al.*, 2016). Bien que ces méthodes reposent sur des outils d'analyse de sentiments, elles garantissent que les préférences des utilisateurs inférées reflètent fidèlement leurs opinions réelles. De manière plus générale, la modélisation des aspects permet d'inférer à la fois la note globale d'un utilisateur pour un article et ses opinions détaillées sur des aspects spécifiques, améliorant ainsi l'explicabilité des recommandations et enrichissant les explications textuelles par des informations fines et spécifiques aux aspects (Ni *et al.*, 2019; Sun *et al.*, 2021).

Dans cet article, nous présentons **AURA** (Aspect-based Unified Rating Prediction and Personalized Review Generation with Attention), un modèle multi-tâches composé de deux modules : le module de prédiction de notes et le module de génération de revues personnalisées. Notre approche apprend à la fois des représentations globales et basées sur les aspects pour chaque utilisateur et chaque article à partir des données d'interaction, en optimisant trois objectifs prédictifs : la prédiction de la note globale, la prédiction des notes par aspect et la génération de revues utilisateurs. Nous utilisons une attention personnalisée pour estimer l'importance des différents aspects pour chaque utilisateur et chaque article. Le module de génération de revues déduit une invite continue personnalisée à partir des représentations de l'utilisateur et de l'article, qui est ensuite transmise à un modèle de langage

pour générer la revue. En utilisant un modèle de langage pré-entraîné, notre approche tire parti du pré-entraînement en fixant les paramètres du modèle durant l'apprentissage, conformément au prompt tuning (Lester *et al.*, 2021). Les informations liées aux aspects déduites par le modèle – notamment les notes par aspect et les poids d'importance déterminés grâce à l'attention personnalisée – ainsi que la revue générée offrent une explication complète de la recommandation de l'article à l'utilisateur.

Dans notre implémentation du modèle, nous utilisons le modèle de langage pré-entraîné T5 (Raffel *et al.*, 2020) en adoptant une stratégie de prompt tuning. Des expériences menées sur deux ensembles de données multi-aspects réelles, TripAdvisor et RateBeer, montrent que AURA surpasse de solides modèles de référence, y compris des architectures Transformer unifiées, en particulier dans la tâche de génération de revues. Des études d'ablation et des analyses empiriques mettent également en lumière les contributions de l'attention personnalisée et de la modélisation des aspects. Notre implémentation du modèle est disponible ici : https://github.com/BenKabongo25/aura.

2 TRAVAUX CONNEXES

Les systèmes de recommandation constituent l'un des domaines précurseurs de l'apprentissage de représentations, notamment grâce au filtrage collaboratif (CF) (Koren *et al.*, 2009). Les modèles de recommandation basés sur le CF capturent les préférences des utilisateurs en identifiant des similarités entre utilisateurs ou articles dérivées des données d'interaction. Différentes architectures basées sur la factorisation matricielle (Koren *et al.*, 2009) ainsi que sur des méthodes d'apprentissage profond (He *et al.*, 2017) ont été proposées. Cependant, ces méthodes font face à des défis tels que la sparcité de la matrice d'interactions et le problème du démarrage à froid, où un nombre limité d'interactions entrave la recommandation pour les nouveaux utilisateurs ou articles. Les modèles de recommandation basés sur le contenu (CB) (Lops *et al.*, 2011) atténuent ce problème en intégrant des caractéristiques des articles, telles que les revues, bien qu'ils aient tendance à recommander uniquement des articles similaires. Les méthodes hybrides combinent différentes techniques de recommandation afin de surmonter les limitations propres à chaque approche. L'intégration d'informations supplémentaires sur les utilisateurs ou les articles, comme les attributs des articles (Cheng *et al.*, 2018; Chin *et al.*, 2018) ou les revues (Sun *et al.*, 2021), améliore considérablement les performances des modèles basés sur le CF.

Plus particulièrement, les systèmes de recommandation basés sur les aspects (ABRS) améliorent la personnalisation en modélisant les préférences des utilisateurs et les caractéristiques des articles à un niveau de granularité plus fin grâce à des attributs ou aspects spécifiques. Une première catégorie de méthodes ABRS extrait les aspects des revues de manière non supervisée. Par exemple, ALFM (Cheng *et al.*, 2018) utilise un modèle d'extraction de thèmes par aspects (ATM) pour apprendre une distribution multivariée sur les thèmes à partir des revues, tandis qu'ANR (Chin *et al.*, 2018) apprend des représentations d'aspects pour les utilisateurs et les articles en employant des mécanismes d'attention pour se concentrer sur les parties les plus pertinentes des revues. En revanche, une seconde catégorie de méthodes ABRS se concentre sur l'analyse d'opinion (Bauman *et al.*, 2017; Chen *et al.*, 2016), en tirant parti des techniques d'analyse de sentiment basée sur les aspects (ABSA) (Zhang *et al.*, 2022) pour extraire les opinions sur divers aspects à partir des revues. L'avantage clé de ces méthodes réside dans leur meilleure interprétabilité et explicabilité par rapport aux approches de recommandation classiques.

L'explicabilité est devenue un axe central dans la recherche sur les systèmes de recommandation,

avec un intérêt croissant pour l'exploitation des modèles de langage afin de générer des explications textuelles (Li *et al.*, 2017; Dong *et al.*, 2017; Li *et al.*, 2021, 2023). Les premières méthodes reposaient sur des réseaux de neurones récurrents (RNN), Att2Seq (Dong *et al.*, 2017) et NRT (Li *et al.*, 2017) en étant des exemples notables. Ce dernier, en particulier, est un modèle multi-tâches qui génère des explications tout en prédisant simultanément la note globale. D'autres méthodes basées sur les RNN intègrent la modélisation des aspects pour mieux guider la génération d'explications (Ni & McAuley, 2018; Ni *et al.*, 2019; Sun *et al.*, 2021). Par la suite, plusieurs modèles de génération d'explications tirent parti des Transformers. L'un des premiers modèles multitâches basés sur les Transformers est PETER (Li *et al.*, 2021), qui intègre des données hétérogènes, y compris les embeddings de mots, d'utilisateurs et d'articles, pour la prédiction de la note globale et la génération d'explications. PEPLER (Li *et al.*, 2023) est dérivé de PETER et exploite en outre un Transformer pré-entraîné, GPT-2 (Radford *et al.*, 2019). Cette architecture a fait l'objet d'extensions (Cheng *et al.*, 2023; Raczyński *et al.*, 2023; Shimizu *et al.*, 2024). Malgré ces avancées, les gains de performance par rapport aux méthodes basées sur les RNN restent modestes, en particulier pour la génération de longues explications. Cela reflète les limites des adaptations antérieures des Transformers pour la génération d'explications personnalisées, la plupart de ces méthodes négligeant la modélisation des aspects. L'intégration d'informations spécifiques aux aspects permet d'améliorer la qualité des explications générées. Pour relever ce défi, nous proposons une approche qui combine le prompt tuning (Lester *et al.*, 2021) avec l'intégration d'informations basées sur les aspects, afin de mieux guider le processus de génération et d'améliorer la qualité des explications personnalisées.

3 AURA

Nous proposons **AURA** (Aspect-based Unified Rating Prediction and Personalized Review Generation with Attention), un modèle multi-tâches conçu pour prédire la note globale, les notes par aspect et la revue personnalisée pour une paire utilisateur-article donnée. AURA se compose de deux composants principaux : le **module de prédiction de notes** et le **module de génération de revues personnalisées**. La Figure 1 présente une vue d'ensemble complète de l'architecture d'AURA.

3.1 Notations et Formulation du Problème

Soit $\mathcal{U}$ l'ensemble des utilisateurs et $\mathcal{I}$ l'ensemble des articles. Nous notons $|\mathcal{U}|$ et $|\mathcal{I}|$ le nombre d'utilisateurs et d'articles, respectivement. Soit $\mathcal{R}$ l'ensemble des interactions entre utilisateurs et articles. Pour une interaction entre un utilisateur u et un article i, notons r_{ui} la note globale et t_{ui} la revue fournie par l'utilisateur u pour l'article i. La revue t_{ui} est représentée comme une séquence de tokens, $t_{ui} = (y_1, y_2, \ldots, y_{|t_{ui}|})$, où chaque token y_k appartient à l'ensemble du vocabulaire $\mathcal{V}$.

Pour un domaine d'application donné, nous notons $\mathcal{A}$ l'ensemble des aspects d'intérêt, et $|\mathcal{A}|$ le nombre d'aspects. Les notes d'aspect pour un article i fournies par un utilisateur u peuvent être soit explicitement disponibles, soit extraites de la revue t_{ui} à l'aide de méthodes d'analyse de sentiments basées sur les aspects (ABSA) (Zhang *et al.*, 2022). Nous notons r^a_{ui} la note attribuée par l'utilisateur u à l'aspect a de l'article i. L'ensemble des interactions est défini par :

$$\mathcal{R} = \left\{ \left(u, i, r_{ui}, t_{ui}, \{r^a_{ui}\}_{a \in \mathcal{A}} \right) \right\}. \tag{1}$$

Notre objectif est d'apprendre des représentations d'utilisateurs et d'articles à partir des données

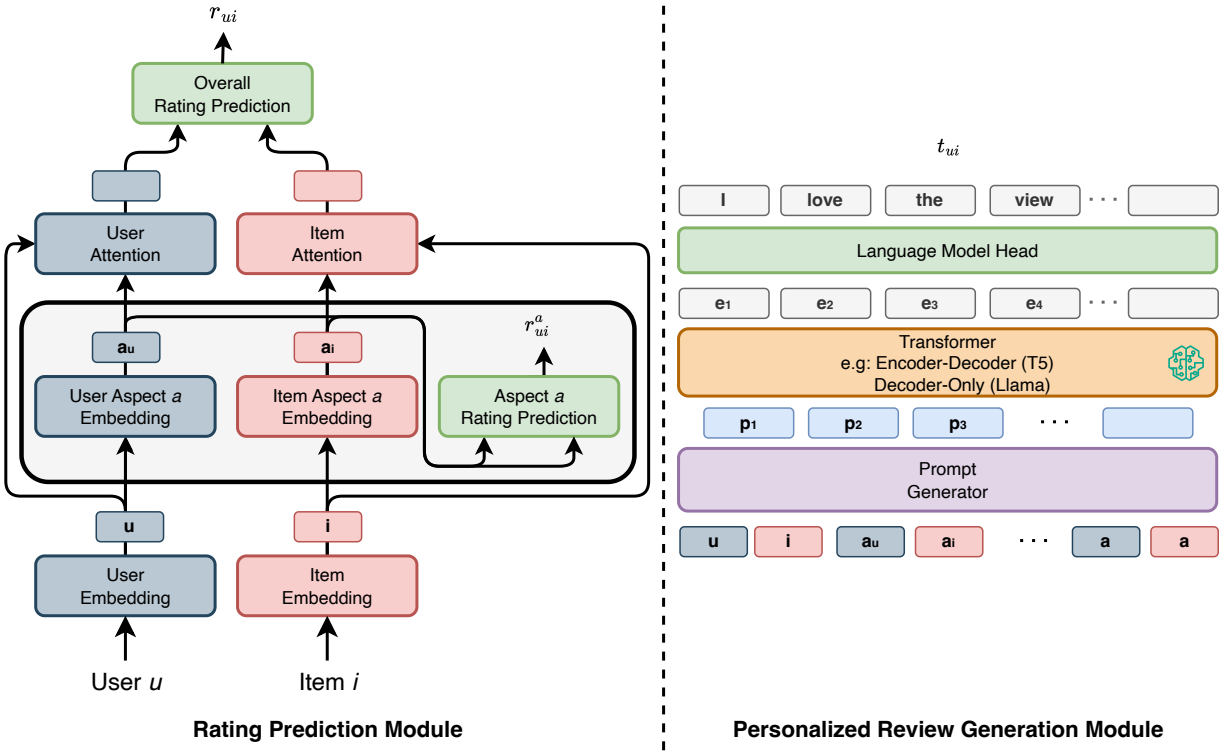


FIGURE 1 – Vue d'ensemble de l'architecture du modèle AURA, composé du module de prédiction de notes (à gauche) et du module de génération de revue personnalisée (à droite).

d'interaction, en capturant à la fois les interactions globales et au niveau des aspects afin de mieux modéliser les préférences globales et les intérêts fins. Notre approche représente les préférences des utilisateurs et les caractéristiques des articles sur ces deux niveaux, en utilisant une attention personnalisée pour évaluer l'importance de chaque aspect pour chaque utilisateur et chaque article. Nous y parvenons grâce à trois tâches unifiées : la prédiction de la note globale, la prédiction des notes spécifiques aux aspects et la génération de revues personnalisées, garantissant à la fois une haute performance et des recommandations explicables.

3.2 Modèle

AURA apprend des représentations globales ($\mathbf{u}$ et $\mathbf{i}$) pour chaque utilisateur u et article i, et dérive un ensemble de représentations spécifiques aux aspects ($\{\mathbf{a}_u\}_{a\in\mathcal{A}}$ et $\{\mathbf{a}_i\}_{a\in\mathcal{A}}$). Le modèle utilise une attention personnalisée pour estimer l'importance relative de chaque aspect pour les utilisateurs et les articles. AURA se compose de deux modules clés : le module de prédiction de notes et le module de génération de revues personnalisées. Le module de prédiction de notes exploite les représentations d'aspects de l'utilisateur u et de l'article i pour prédire à la fois la note globale r_{ui} et les notes par aspect $\{r^a_{ui}\}_{a\in\mathcal{A}}$. Les représentations globales et spécifiques aux aspects de l'utilisateur et de l'article sont ensuite combinées pour générer une invite continue personnalisée $\mathbf{p}_{ui}$. Le module de génération de revues utilise alors un modèle de langage θ_{LM} pour générer la revue t_{ui} pour l'utilisateur u et l'article i, à partir de l'invite personnalisée $\mathbf{p}_{ui}$. Nous notons θ l'ensemble des paramètres du modèle.

3.2.1 Représentations aux niveaux global et par aspects

Nous apprenons des représentations globales, $\mathbf{u} \in \mathbb{R}^d$ pour chaque utilisateur $u \in \mathcal{U}$ et $\mathbf{i} \in \mathbb{R}^d$ pour chaque article $i \in \mathcal{I}$, qui encodent diverses informations, notamment les préférences des utilisateurs, les caractéristiques des articles et d'autres facteurs contextuels pertinents. À partir des représentations globales, nous cherchons à extraire les préférences de l'utilisateur u et les caractéristiques de l'article i concernant l'aspect a. Pour ce faire, nous définissons deux fonctions, $\phi^a_{\mathcal{U}}, \phi^a_{\mathcal{I}} : \mathbb{R}^d \to \mathbb{R}^d$, qui transforment les représentations globales $\mathbf{u}$ et $\mathbf{i}$ en représentations pour l'aspect a, notées respectivement $\mathbf{a}_u$ et $\mathbf{a}_i$. Ces représentations sont obtenues comme suit :

$$\mathbf{a}_u = \phi^a_{\mathcal{U}}(\mathbf{u}), \quad \mathbf{a}_i = \phi^a_{\mathcal{I}}(\mathbf{i}), \quad \mathbf{a}_u, \mathbf{a}_i \in \mathbb{R}^d. \tag{2}$$

Nous notons $(\theta_{\mathcal{U}}, \theta_{\mathcal{I}}, \theta_{\mathcal{A}})$ l'ensemble des paramètres du modèle correspondant aux utilisateurs, aux articles et aux aspects, respectivement.

3.2.2 Attention Personnalisée

Dans les systèmes de recommandation personnalisés, l'importance de chaque aspect varie considérablement en fonction des préférences de l'utilisateur et des caractéristiques de l'article (Cheng *et al.*, 2018; Chin *et al.*, 2018). Nous modélisons ces variations d'importance des aspects à l'aide d'une attention personnalisée, qui estime les poids d'importance de chaque aspect pour chaque utilisateur et chaque article.

Pour un utilisateur u, nous estimons les poids d'importance relatifs des aspects, notés $\{\alpha^a_u\}_{a\in\mathcal{A}}$, en utilisant trois fonctions de projection : $q_{\mathcal{U}}, k_{\mathcal{U}}, v_{\mathcal{U}} : \mathbb{R}^d \to \mathbb{R}^d$, inspirées du mécanisme d'attention de (Vaswani *et al.*, 2017). De même, pour un article i, nous appliquons trois fonctions de projection, $q_{\mathcal{I}}, k_{\mathcal{I}}, v_{\mathcal{I}} : \mathbb{R}^d \to \mathbb{R}^d$, afin de calculer les poids d'attention des aspects $\{\alpha^a_i\}_{a\in\mathcal{A}}$. Les poids d'importance de l'aspect a pour l'utilisateur u et l'article i, respectivement notés α^a_u et α^a_i, sont calculés comme suit :

$$\alpha^a_u = \exp\left(\frac{q_{\mathcal{U}}(\mathbf{u})^{\mathrm{T}} k_{\mathcal{U}}(\mathbf{a}_u)}{Z}\right), \quad \alpha^a_i = \exp\left(\frac{q_{\mathcal{I}}(\mathbf{i})^{\mathrm{T}} k_{\mathcal{I}}(\mathbf{a}_i)}{Z}\right), \tag{3}$$

où Z est un terme de normalisation. Pour un utilisateur, les poids d'attention des aspects reflètent l'importance relative qu'il accorde à chaque aspect. Pour un article, ces poids représentent l'importance moyenne attribuée à chaque aspect par les utilisateurs.

Nous calculons ensuite des représentations agrégées pour l'utilisateur u et l'article i, notées $\tilde{\mathbf{u}}$ et $\tilde{\mathbf{i}}$ respectivement. Ces représentations capturent de manière dynamique les préférences de l'utilisateur et les caractéristiques de l'article à travers divers aspects. Elles sont obtenues comme suit :

$$\tilde{\mathbf{u}} = \sum_{a\in\mathcal{A}} \alpha^a_u \, v_{\mathcal{U}}(\mathbf{a}_u), \quad \tilde{\mathbf{i}} = \sum_{a\in\mathcal{A}} \alpha^a_i \, v_{\mathcal{I}}(\mathbf{a}_i), \quad \tilde{\mathbf{u}}, \tilde{\mathbf{i}} \in \mathbb{R}^d. \tag{4}$$

3.2.3 Module de Prédiction de Notes

L'objectif du module de prédiction de notes est d'estimer à la fois la note globale r_{ui} et les notes spécifiques aux aspects $\{r^a_{ui}\}_{a\in\mathcal{A}}$ pour un utilisateur u et un article i. Nous notons θ_R l'ensemble des paramètres spécifiques à ce module.

Prédiction de la Note Globale. Pour prédire la note globale r_{ui} de l'utilisateur u pour l'article i, nous définissons une fonction $f : \mathbb{R}^{2\times d} \to \mathbb{R}$. Les entrées de cette fonction sont les représentations agrégées de l'utilisateur et de l'article, notées $\tilde{\mathbf{u}}$ et $\tilde{\mathbf{i}}$, respectivement. Ces représentations encapsulent les préférences essentielles de l'utilisateur et les caractéristiques de l'article, tout en intégrant de manière dynamique l'importance relative des différents aspects. La note globale est prédite comme suit :

$$\hat{r}_{ui} = f(\tilde{\mathbf{u}}, \tilde{\mathbf{i}}). \tag{5}$$

Prédiction des Notes par Aspect. Pour prédire la note r^a_{ui}, qui représente l'évaluation de l'utilisateur u de l'aspect a pour l'article i, nous définissons une fonction $g_a : \mathbb{R}^{2\times d} \to \mathbb{R}$. Les entrées de cette fonction sont les représentations de l'aspect a pour l'utilisateur et l'article, notées respectivement $\mathbf{a}_u$ et $\mathbf{a}_i$. Ces représentations capturent les préférences fines de l'utilisateur et les caractéristiques spécifiques de l'article pour l'aspect a. La prédiction de la note de l'aspect a est donnée par :

$$\hat{r}^a_{ui} = g_a(\mathbf{a}_u, \mathbf{a}_i). \tag{6}$$

3.2.4 Génération de Revue Personnalisée.

Le but du module de génération de revue personnalisée est de générer la revue t_{ui} pour l'utilisateur u et l'article i. Pour ce faire, le module génère une invite continue personnalisée, notée $\mathbf{p}_{ui}$. Nous notons θ_P l'ensemble des paramètres du modèle pour la génération d'invites. Ensuite, nous employons un modèle de langage (pré-entraîné) θ_{LM} de dimension d_w pour générer la revue t_{ui}, conditionnée par l'invite personnalisée.

Invite Personnalisée. À partir des représentations globales $\mathbf{u}$ et $\mathbf{i}$ de l'utilisateur u et de l'article i, ainsi que de leurs représentations par aspects $\{\mathbf{a}_u, \mathbf{a}_i\}_{a\in\mathcal{A}}$, nous utilisons une fonction $\psi : \mathbb{R}^{2(1+|\mathcal{A}|)\times d} \to \mathbb{R}^{\eta\times d_w}$ pour obtenir l'invite continue personnalisée destinée à u et i. Ici, η est un hyperparamètre du modèle qui dépend du modèle de langage θ_{LM}. Notre hypothèse est qu'en plus des représentations globales, les préférences de l'utilisateur et les caractéristiques de l'article encapsulées dans les représentations par aspects contribuent à guider plus efficacement le processus de génération de la revue. L'invite personnalisée, notée $\mathbf{p}_{ui} \in \mathbb{R}^{\eta\times d_w}$, se compose de η tokens dans l'espace latent du modèle de langage θ_{LM} de dimension d_w. Elle capture les mêmes informations que les représentations globales et par aspects de l'utilisateur et de l'article, et est calculée comme suit :

$$\mathbf{p}_{ui} = \psi\big(\mathbf{u}, \mathbf{i}, \{\mathbf{a}_u, \mathbf{a}_i\}_{a\in\mathcal{A}}\big). \tag{7}$$

Génération de la Revue. Après avoir obtenu l'invite continue personnalisée $\mathbf{p}_{ui}$ pour l'utilisateur u et l'article i, nous employons le modèle de langage θ_{LM} pour générer la revue t_{ui}, conditionnée par l'invite. La génération de la revue t_{ui} est formulée comme suit :

$$\begin{aligned} P_{\theta_P,\theta_{LM}}\big(t_{ui}|\mathbf{p}_{ui}\big) &= P_{\theta_P,\theta_{LM}}\big((y_1, y_2, \ldots, y_{|t_{ui}|})|\mathbf{p}_{ui}\big) \\ &= \prod_{k=1}^{|t_{ui}|} P_{\theta_P,\theta_{LM}}\big(y_k|\mathbf{p}_{ui}, y_{<k}\big). \end{aligned} \tag{8}$$

3.3 Optimisation

Dans cette section, nous abordons l'optimisation du modèle, en présentant les fonctions de perte ainsi que la procédure d'entraînement.

3.3.1 Fonctions de Perte

L'ensemble des paramètres du modèle, noté $\theta = (\theta_{\mathcal{U}}, \theta_{\mathcal{I}}, \theta_{\mathcal{A}}, \theta_R, \theta_P, \theta_{LM})$, est optimisé en minimisant trois fonctions de perte correspondant aux trois objectifs : la prédiction de la note globale, la prédiction des notes par aspect et la génération de revues. Pour la prédiction de la note globale, nous utilisons la fonction de perte par erreur quadratique moyenne (MSE), définie comme suit :

$$\mathcal{L}_R = \frac{1}{|\mathcal{R}|} \sum_{(u,i)\in\mathcal{R}} \left(r_{ui} - \hat{r}_{ui}\right)^2. \tag{9}$$

Pour la prédiction des notes par aspect, nous employons la moyenne des pertes MSE calculées pour chaque aspect :

$$\mathcal{L}_A = \frac{1}{|\mathcal{A}|} \sum_{a\in\mathcal{A}} \frac{1}{|\mathcal{R}|} \sum_{(u,i)\in\mathcal{R}} \left(r^a_{ui} - \hat{r}^a_{ui}\right)^2. \tag{10}$$

La perte totale du module de prédiction de notes est une combinaison pondérée de ces deux pertes :

$$\mathcal{L}_{rating} = \alpha\mathcal{L}_R + (1-\alpha)\mathcal{L}_A, \tag{11}$$

où α est un hyperparamètre qui contrôle l'importance relative de la perte de la note globale $\mathcal{L}_R$ et de la perte des notes par aspect $\mathcal{L}_A$. Pour la génération de revues personnalisées, nous utilisons la perte de vraisemblance négative (NLL), donnée par :

$$\mathcal{L}_{review} = -\frac{1}{|\mathcal{R}|} \sum_{(u,i)\in\mathcal{R}} \frac{1}{|t_{ui}|} \sum_{k=1}^{|t_{ui}|} \log P_{\theta_P,\theta_{LM}}\left(y_k \mid \mathbf{p}_{ui}, y_{<k}\right). \tag{12}$$

3.3.2 Entraînement

Notre approche offre une flexibilité quant au choix du modèle de langage, qui peut être basé soit sur des réseaux de neurones récurrents (RNN) soit sur des architectures basées sur des Transformers. Lorsqu'on utilise un modèle de langage pré-entraîné (Raffel *et al.*, 2020; Dubey *et al.*, 2024), notre méthode s'aligne naturellement avec le prompt tuning (Lester *et al.*, 2021), ce qui nous permet de tirer efficacement parti des connaissances acquises. Dans cette configuration, nous fixons les paramètres du modèle de langage θ_{LM} pendant l'entraînement et n'optimisons que les paramètres de génération d'invite θ_P. Cette stratégie réduit significativement le nombre de paramètres entraînables, tout en préservant la richesse des connaissances encodées dans le modèle pré-entraîné.

Pour maximiser l'efficacité dans ce contexte, nous adoptons une approche d'entraînement séquentielle. Étant donné que l'invite personnalisée dépend directement des représentations globales et spécifiques aux aspects des utilisateurs et des articles, nous optimisons d'abord ces représentations en entraînant

le module de prédiction de notes et en minimisant la perte $\mathcal{L}_{rating}$. Une fois que les représentations apprises sont suffisamment informatives, nous entraînons uniquement les paramètres de génération d'invite dans le module de génération de revues personnalisées, en minimisant la perte de génération $\mathcal{L}_{review}$, tout en maintenant les paramètres du modèle de langage figés.

3.4 Ablations

Nous présentons diverses ablations du modèle AURA afin d'évaluer la contribution de chaque composant, tels que l'attention personnalisée et la modélisation des aspects.

3.4.1 Attention Personnalisée

Nous désignons *AURA -Attention* comme l'ablation du modèle AURA dans laquelle l'attention personnalisée, utilisée pour estimer l'importance des aspects pour chaque utilisateur et article et pour agréger les représentations par aspects, est remplacée par une opération de max pooling. Les représentations agrégées par aspects de l'utilisateur u et de l'article i sont calculées en appliquant un max pooling sur l'ensemble de leurs représentations par aspects, comme suit :

$$(\tilde{\mathbf{u}})_j = \max_{a \in \mathcal{A}} (\mathbf{a}_u)_j, \quad (\tilde{\mathbf{i}})_j = \max_{a \in \mathcal{A}} (\mathbf{a}_i)_j, \quad \forall j \in \{1, \ldots, d\}. \tag{13}$$

3.4.2 Modélisation des Aspects

Nous désignons *AURA -Aspects* comme l'ablation du modèle AURA dans laquelle la modélisation des aspects est supprimée. Dans cette variante, le modèle n'apprend pas de représentations spécifiques aux aspects pour les utilisateurs et les articles, et nous omettons également la prédiction des notes par aspect ainsi que l'attention personnalisée. Pour la prédiction de la note globale, nous redéfinissons la fonction $f : \mathbb{R}^{2\times d} \to \mathbb{R}$ et, pour la génération d'invite personnalisée, nous redéfinissons la fonction $\psi : \mathbb{R}^{2\times d} \to \mathbb{R}^{\eta \times d_w}$. Puisque la modélisation des aspects est omise, ces fonctions prennent en entrée uniquement les représentations globales de l'utilisateur et de l'article, comme suit :

$$\hat{r}_{ui} = f(\mathbf{u}, \mathbf{i}), \quad \mathbf{p}_{ui} = \psi(\mathbf{u}, \mathbf{i}). \tag{14}$$

3.4.3 Représentations Globales

Nous désignons *AURA -Global* comme une ablation du modèle AURA dans laquelle la prédiction de la note globale, la prédiction des notes par aspect et la génération d'invite continue personnalisée reposent uniquement sur les représentations globales des utilisateurs et des articles. Cette ablation nous permet d'évaluer l'importance combinée de la modélisation des aspects et de l'attention personnalisée pour l'ensemble des tâches du modèle. Dans cette variante, les fonctions $f : \mathbb{R}^{2\times d} \to \mathbb{R}$, $g_a : \mathbb{R}^{2\times d} \to \mathbb{R}$ et $\psi : \mathbb{R}^{2\times d} \to \mathbb{R}^{\eta \times d_w}$, respectivement pour la prédiction de la note globale, la prédiction de la note pour l'aspect a et la génération de l'invite, sont définies comme suit :

$$\hat{r}_{ui} = f(\mathbf{u}, \mathbf{i}), \quad \hat{r}^a_{ui} = g_a(\mathbf{u}, \mathbf{i}), \quad \mathbf{p}_{ui} = \psi(\mathbf{u}, \mathbf{i}). \tag{15}$$

Jeux de Données	TripAdvisor	RateBeer
Aspects	6	4
Utilisateurs	716 870 (8 830)	40 213 (8 384)
Articles	10 008 (2 903)	110 419 (5 093)
Interactions	1,4M (62,7K)	2,8M (201,7K)

TABLE 1 – Description des jeux de données. Les statistiques après filtrage sont indiquées entre parenthèses.

3.5 Implémentation

Nous implémentons les fonctions clés d'AURA à l'aide de perceptrons multicouches (MLP). Cela inclut f pour la prédiction de la note globale, $\{g_a\}_{a\in\mathcal{A}}$ pour la prédiction des notes par aspect, $\{\phi^a_{\mathcal{U}}, \phi^a_{\mathcal{I}}\}_{a\in\mathcal{A}}$ pour les représentations par aspects, et ψ pour la génération d'invite personnalisée. Pour l'attention personnalisée de l'utilisateur $(q_{\mathcal{U}}, k_{\mathcal{U}}, v_{\mathcal{U}})$ et de l'article $(q_{\mathcal{I}}, k_{\mathcal{I}}, v_{\mathcal{I}})$, nous apprenons des matrices de projection distinctes. Par exemple, pour la fonction $q_{\mathcal{U}}$, nous utilisons une matrice de projection $\mathbf{W}^q_{\mathcal{U}} \in \mathbb{R}^{d\times d}$. La génération de la revue est donnée par $P_{\theta_P,\theta_{LM}}\left(t_{ui} \mid \mathbf{p}_{ui}\right)$. Dans notre implémentation d'AURA, nous utilisons le modèle de langage T5, basé sur l'architecture Transformer pré-entraînée (Raffel *et al.*, 2020), et nos expériences s'inscrivent dans une approche de prompt tuning au sein d'un entraînement séquentiel. L'hyperparamètre η, qui spécifie le nombre de tokens d'invite, dépend du modèle de langage, comme démontré dans (Lester *et al.*, 2021).

4 EXPERIMENTATIONS

4.1 Configuration Expérimentale

4.1.1 Jeux de Données

Nos expériences sont menées sur deux jeux de données multi-aspects :

— **TripAdvisor** [1] (Wang *et al.*, 2010) : comprend des revues couvrant six aspects des hôtels : propreté (*cleanliness*), localisation (*location*), chambre (*rooms*), service (*service*), qualité du sommeil (*sleep quality*) et rapport qualité-prix (*value*).

— **RateBeer** [2] (McAuley *et al.*, 2012) : comprend des revues axées sur quatre aspects des bières : apparence (*appearance*), arôme (*aroma*), palais (*palate*) et goût (*taste*).

Pour chaque jeu de données, nous excluons les interactions contenant des informations manquantes et ne conservons que les utilisateurs et articles ayant au moins 5 revues. La Table 1 fournit un résumé des jeux de données avant et après application de ces étapes de filtrage.

4.1.2 Métriques d'Évaluation

Nous utilisons la racine de l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE) pour évaluer les modèles sur la prédiction des notes. Pour la génération de revues, nous évaluons les

1. https://www.cs.virginia.edu/~hw5x/Data/LARA/TripAdvisor/
2. https://cseweb.ucsd.edu/~jmcauley/datasets.html#multi_aspect

modèles à l'aide de métriques de qualité textuelle, notamment METEOR (Banerjee & Lavie, 2005), BLEU (Papineni *et al.*, 2002), ROUGE (Lin, 2004) et BERTScore (Zhang *et al.*, 2019).

4.1.3 Baselines

Nous comparons AURA à un ensemble diversifié de méthodes de référence, incluant des méthodes traditionnelles de prédiction de notes, des modèles de recommandation basés sur les aspects, des approches de génération d'explications, ainsi que des modèles multi-tâches. Pour évaluer la contribution des différents composants d'AURA, nous considérons également ses diverses ablations, notamment AURA -Attention, AURA -Aspects et AURA -Global.

Pour la prédiction de notes, nous utilisons les baselines suivantes : Average, MF (Koren *et al.*, 2009), MLP (He *et al.*, 2017) et NeuMF (He *et al.*, 2017). La méthode Average prédit la note moyenne sur l'ensemble des paires utilisateur-article. Pour la recommandation basée sur les aspects, nous évaluons deux méthodes de référence : ALFM (Cheng *et al.*, 2018) et ANR (Chin *et al.*, 2018). Ces méthodes apprennent des représentations d'aspects à partir des revues de manière non supervisée, rendant difficile l'alignement des aspects appris avec ceux présents dans le jeu de données. Nous considérons également les ablations AURA -Global et AURA -Attention comme baselines supplémentaires spécifiquement pour la prédiction des notes par aspect.

Pour la génération de revues, nous distinguons deux catégories de modèles de génération d'explications. La première catégorie regroupe les modèles basés sur des RNN, notamment Att2Seq (Dong *et al.*, 2017) et l'architecture multi-tâche NRT (Li *et al.*, 2017). La seconde catégorie inclut les modèles multi-tâches basés sur des Transformers, tels que PETER (Li *et al.*, 2021) et PEPLER (Li *et al.*, 2023). PETER utilise un Transformer non pré-entraîné, tandis que PEPLER fait appel au GPT-2 pré-entraîné (Radford *et al.*, 2019) dans une approche de fine-tuning. La plupart des méthodes de génération d'explications textuelles considérées reposent sur des représentations latentes des utilisateurs et des articles — typiquement encodées sous la forme de seulement deux tokens — pour conditionner l'ensemble de l'explication. Nous excluons des modèles tels que ceux proposés dans (Cheng *et al.*, 2023; Raczyński *et al.*, 2023; Shimizu *et al.*, 2024), qui sont des extensions de PETER et PEPLER, car leurs performances se rapprochent étroitement de celles des modèles originaux.

4.1.4 Configuration

Chaque jeu de données est divisé en ensembles d'entraînement, de validation et de test selon un ratio de 80 :10 :10. Toutes les notes, y compris les notes globales et par aspect, sont standardisées sur une échelle de 1 à 5. La longueur des revues est limitée à 128 mots pour tous les modèles. Les modèles sont entraînés sur l'ensemble d'entraînement, et les hyperparamètres sont ajustés à l'aide de l'ensemble de validation. Les résultats rapportés correspondent aux performances des modèles sur l'ensemble de test. Pour la plupart des modèles, nous utilisons les hyperparamètres spécifiés dans les articles originaux. Plus précisément, pour PEPLER (Li *et al.*, 2023), nous utilisons GPT-2 (124M) (Radford *et al.*, 2019) comme détaillé dans l'article d'origine. Pour AURA, nous utilisons le modèle T5-Small (60M) (Raffel *et al.*, 2020), qui est deux fois plus petit que GPT-2.

Après des expériences préliminaires, nous fixons η, le nombre de tokens dans l'invite personnalisée, à 50. La dimension du modèle, d, est fixée à 256, tandis que la dimension du modèle T5-Small est de 512 (Raffel *et al.*, 2020). Le nombre de couches dans les MLP est fixé à 2, et nous utilisons la

	TripAdvisor		RateBeer	
Model	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓
Average	0.9325	0.6458	0.5711	0.4249
MF	0.8409	0.6463	**0.4114**	**0.3008**
MLP	0.8332	0.5656	0.4648	0.3244
NeuMF	0.8408	0.5702	0.4731	0.3295
ALFM	0.8967	0.6912	0.4335	0.3142
ANR	0.8473	0.6075	0.4231	0.3084
NRT	0.8592	0.5481	0.4208	0.3066
PETER	0.8078	0.5327	0.4156	**0.3008**
PEPLER	0.7792	0.4782	0.4305	0.3059
AURA	**0.7482**	**0.4477**	0.4166	0.3050
-Attention	0.7716	0.5132	0.4217	0.3111
-Global	0.8651	0.6320	0.4439	0.3326

TABLE 2 – Performance sur la prédiction de la note globale.

fonction d'activation Rectified Linear Unit (ReLU) (Nair & Hinton, 2010). La probabilité de dropout est réglée à 0,1 afin de prévenir le surapprentissage. Nous constatons que fixer $\alpha = \frac{1}{|\mathcal{A}|+1}$, où $|\mathcal{A}|$ représente le nombre d'aspects, permet d'obtenir le meilleur compromis entre la performance sur les notes globales et les notes par aspect. Tous les modèles sont entraînés pendant un maximum de 100 époques. Pour AURA, nous utilisons un entraînement séquentiel avec prompt tuning. Le module de prédiction de notes est entraîné pendant 50 époques, suivi de 50 époques pour le module de génération de revues. L'entraînement est effectué à l'aide de l'optimiseur Adam (Kingma, 2014) avec un taux d'apprentissage de 10^{-3}.

4.2 Prédiction des Notes

Note Globale. Nous comparons AURA à divers modèles de référence pour la prédiction de la note globale, les résultats étant présentés dans la Table 2. AURA surpasse systématiquement tous les baselines sur l'ensemble des métriques, notamment sur le jeu de données TripAdvisor. Sur le jeu de données RateBeer, il se positionne parmi les modèles les plus performants, avec des performances proches de celles de MF et PETER. La supériorité d'AURA par rapport aux approches traditionnelles et multi-tâches peut être attribuée à son intégration de la modélisation des aspects. De même, son avantage par rapport aux baselines basés sur les aspects souligne les bénéfices d'une modélisation supervisée, notamment grâce à l'extraction directe des opinions sur les aspects à partir des revues. Les ablations AURA -Attention et AURA -Global affichent des performances inférieures à la version complète d'AURA, confirmant l'importance de l'attention personnalisée et de l'utilisation conjointe des représentations globales et spécifiques aux aspects pour atteindre une précision optimale.

Notes par Aspect. Nous comparons AURA avec la méthode Average et les ablations AURA -Global et AURA -Attention pour la prédiction des notes par aspect. Pour chaque modèle, nous calculons le RMSE et le MAE pour l'ensemble des aspects, puis nous rapportons la moyenne et l'écart-type de ces métriques sur l'ensemble des aspects. Les résultats sont présentés dans la Table 3. AURA surpasse de manière constante et significative toutes les autres méthodes sur l'ensemble des métriques et des

	TripAdvisor		RateBeer	
Model	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓
Average	1.014 (0.0879)	0.8014 (0.0572)	0.6054 (0.0117)	0.4893 (0.0231)
AURA	**0.7532 (0.0811)**	**0.4514 (0.0538)**	**0.4657 (0.0347)**	**0.3540 (0.0307)**
-Attention	0.7851 (0.0736)	0.5541 (0.0514)	0.4866 (0.0316)	0.3731 (0.0303)
-Global	0.8607 (0.0760)	0.6313 (0.0561)	0.4950 (0.0359)	0.3832 (0.0318)

TABLE 3 – Performance sur la prédiction des notes par aspect. Nous indiquons la moyenne et l'écart-type de tous les aspects.

TripAdvisor	METEOR ↑	BLEU ↑	ROUGE-1 ↑	ROUGE-2 ↑	ROUGE-L ↑	BERT-P ↑	BERT-R ↑	BERT-F1 ↑
Att2Seq	18.6113	04.6900	28.7839	06.4736	18.5239	85.3487	83.6769	84.4902
NRT	17.2198	03.4053	25.8336	05.1943	17.5390	82.8282	81.5335	82.1613
PETER	17.9550	03.9435	27.9742	05.9062	18.2520	85.0379	83.8235	84.4064
$\text{PEPLER}_{\text{GPT-2}}$	24.3400	11.4000	33.8312	11.6797	22.4529	82.6355	84.9450	83.7264
$\text{AURA}_{\text{T5-Small}}$	**42.7527**	**33.5446**	**53.2856**	**37.8780**	**44.0538**	**90.6867**	**88.4785**	**89.5549**
-Aspects	27.6423	10.0294	39.0768	21.9701	29.5941	88.0013	85.1939	86.5400
RateBeer	METEOR ↑	BLEU ↑	ROUGE-1 ↑	ROUGE-2 ↑	ROUGE-L ↑	BERT-P ↑	BERT-R ↑	BERT-F1 ↑
Att2Seq	18.6113	04.6900	28.7839	06.4736	18.5239	85.3487	83.6769	84.4902
NRT	24.9634	08.7375	32.5892	11.4721	26.6292	85.0467	82.9921	83.9859
PETER	28.8189	11.5183	35.5043	13.6200	29.6688	87.3401	85.6216	86.4486
$\text{PEPLER}_{\text{GPT-2}}$	28.2665	10.1432	32.4444	11.1827	26.2481	84.0207	86.0634	84.9906
$\text{AURA}_{\text{T5-Small}}$	**40.7637**	**24.1609**	**46.3715**	**25.8183**	**39.4616**	**90.4830**	**89.1356**	**89.7921**
-Aspects	32.6755	13.6520	39.0688	17.1069	32.4644	89.3652	87.3239	88.3102

TABLE 4 – Performance sur la génération de revue en %.

jeux de données. Les deux ablations dépassent la baseline Average. Nous observons que l'ablation AURA -Global performe moins bien que AURA -Attention. Cet écart de performance s'explique par le fait que AURA -Global repose uniquement sur les représentations globales, tandis que AURA -Attention conserve les représentations spécifiques aux aspects, démontrant ainsi la valeur d'apprendre explicitement des représentations basées sur les aspects pour une prédiction précise des notes par aspect.

De plus, le modèle complet AURA surpasse systématiquement l'ablation AURA -Attention, qui remplace l'attention personnalisée par une opération de max pooling. Cela confirme que l'attention personnalisée offre une stratégie d'agrégation plus efficace des représentations par aspects des utilisateurs et des articles que le max pooling. Enfin, un avantage supplémentaire d'AURA par rapport aux méthodes traditionnelles de prédiction de notes ou aux approches multi-tâches réside dans sa capacité à prédire explicitement les notes par aspect en parallèle de la note globale et de la revue personnalisée, offrant ainsi des recommandations plus riches et informatives.

4.3 Génération de Revue

Nous évaluons AURA par rapport à divers baselines et à l'ablation AURA -Aspects pour la génération de revues en utilisant des métriques de qualité textuelle, les résultats étant présentés dans la Table 4. AURA et AURA -Aspects surpassent de manière constante tous les autres modèles sur l'ensemble des métriques et des jeux de données, avec une performance supérieure pour AURA par rapport à son ablation AURA -Aspects. Ces résultats mettent en évidence les limites des approches basées sur les RNN pour capturer efficacement le contexte à long terme, et montrent que les baselines basées

η	METEOR ↑	BLEU ↑	ROUGE-2 ↑
PEPLER	24.3400	11.4000	11.6797
2	12.1594	01.2821	04.4362
5	16.7304	03.5462	06.0468
10	21.1600	07.6928	10.3301
20	<u>29.3798</u>	<u>17.0189</u>	<u>20.2004</u>
50	**42.7527**	**33.5446**	**37.8780**

TABLE 5 – Impact du nombre de tokens du prompt personnalisé (η) sur la génération de revue pour le jeu de données TripAdvisor.

sur les Transformers (PETER et PEPLER) ne surpassent pas toujours de manière significative les modèles RNN, en particulier sur le jeu de données TripAdvisor. La sous-performance de l'ablation AURA -Aspects confirme notre hypothèse selon laquelle l'incorporation d'informations basées sur les aspects, en complément des représentations globales, est cruciale pour améliorer la qualité des revues générées.

De plus, nous analysons l'impact de η, le nombre de tokens dans l'invite personnalisée, sur la qualité de génération de revues pour le jeu de données TripAdvisor. Comme le montre la Table 5, à partir de 20 tokens, AURA surpasse la baseline PEPLER. La meilleure performance est obtenue avec 50 tokens. Toutes les méthodes de référence conditionnent leur génération uniquement sur des représentations apprises des utilisateurs et des articles, généralement représentées par seulement deux tokens. En revanche, notre approche utilise une invite continue personnalisée plus riche, dont la longueur est optimisée en tant qu'hyperparamètre en fonction du modèle de langage utilisé. Notamment, tant AURA -Aspects que le modèle complet AURA surpassent le modèle PEPLER malgré l'utilisation de T5-Small, qui compte environ la moitié des paramètres du GPT-2 utilisé par PEPLER. Cela souligne encore l'efficacité et la performance de notre approche de prompt tuning personnalisé.

4.4 Modélisation des Aspects

La modélisation des aspects, en apprenant des représentations spécifiques aux aspects pour chaque utilisateur et chaque article, améliore les performances du modèle sur l'ensemble des tâches. Pour la prédiction des notes globales et spécifiques aux aspects, comme le montrent les Tables 2 et 3, le modèle AURA ainsi que son ablation AURA -Attention — qui remplace l'attention par une opération de max pooling tout en conservant les représentations par aspects — surpassent significativement l'ablation AURA -Global, qui repose uniquement sur des représentations globales. Dans la génération de revues, l'ablation AURA -Aspects, qui n'apprend ni les représentations par aspects ni ne prédit les notes par aspect, est également largement dépassée par AURA. Ces résultats soulignent le rôle crucial des représentations basées sur les aspects pour encoder les préférences des utilisateurs et les caractéristiques des articles au niveau des aspects, conduisant à des prédictions plus précises et à une génération de revues plus informative.

Nous menons une étude empirique pour examiner les représentations des utilisateurs basées sur les aspects apprises par le modèle sur le jeu de données TripAdvisor. Ces représentations sont projetées dans un espace bidimensionnel, et regroupées par aspect. Les résultats sont présentés dans la Figure 2. Les représentations basées sur les aspects apprises par AURA permettent une séparation cohérente des utilisateurs selon les aspects, démontrant ainsi que le modèle capte efficacement les préférences

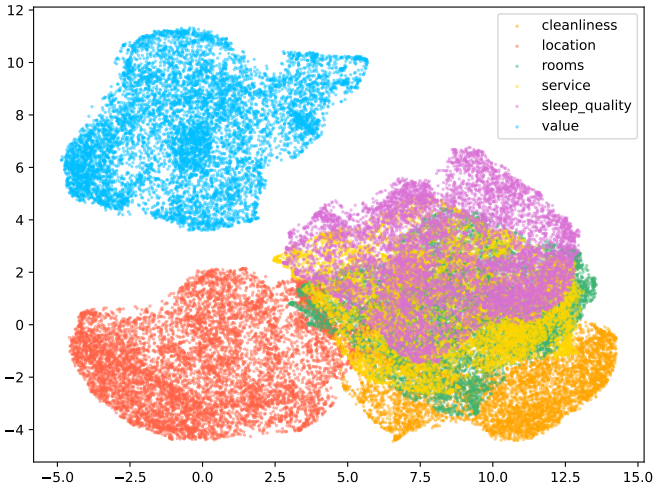
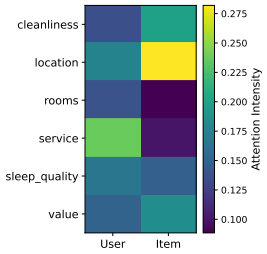


FIGURE 2 – Projection et clustering des représentations par aspect des utilisateurs pour le jeu de données TripAdvisor.



Aspect	Rating	Ground truth review
Cleanliness	4.9 (5.0)	if we go back to paris, we are staying here again. the place is so charming and
Location	5.0 (5.0)	overlooks the beautiful luxembourg gardens. the staff were sooo hospitiable.
Rooms	5.0 (5.0)	always asking what they could do to help us. they arranged two tours for us,
Service	5.0 (5.0)	recommended places to eat and then made the reservations for us, arranged
Sleep	5.0 (5.0)	transportation from and to the airport, etc. royce and xavier, i can't thank you
Value	5.0 (5.0)	enough ! also, so many places are in walking distance, like notre dame and the
Overall	4.9 (5.0)	louvre. you can't help but fall in love with this place !

TABLE 6 – Visualisation de l'attention sur un exemple tiré du jeu de données TripAdvisor. Pour chaque aspect, nous reportons la note prédite par AURA, ainsi que la note réelle entre paranthèses. Un alignement est observé entre les notes des aspects et leurs importances et le contenu de la revue.

des utilisateurs selon des critères spécifiques. Notamment, nous révèlevons un chevauchement entre les aspects *rooms, service, sleep quality* et *cleanliness*, ce qui suggère que ces aspects partagent une plus grande similarité comparativement à *location* et *value*. Ceci met en avant l'avantage de la modélisation des aspects, qui permet d'apprendre les préférences des utilisateurs et les caractéristiques des articles à un niveau fin, conduisant à des représentations mieux structurées et plus interprétables.

4.5 Attention Personnalisée

Le mécanisme d'attention personnalisée d'AURA déduit l'importance relative des aspects pour chaque utilisateur et chaque article. AURA obtient des gains de performance significatifs par rapport à ses ablations AURA -Attention et AURA -Global, notamment dans la prédiction de la note globale et des notes par aspect (voir Tables 2 et 3). Il est à noter que AURA -Attention, qui remplace l'attention personnalisée par une opération de max pooling, entraîne une baisse des performances, soulignant ainsi que l'attention personnalisée agrège l'information des aspects de manière plus efficace que le max pooling.

Les poids d'attention appris peuvent être interprétés comme des indicateurs de l'importance de chaque aspect pour un utilisateur ou un article donné. Pour valider cette hypothèse, nous avons analysé les poids d'attention pour différentes paires utilisateur-article et examiné leur cohérence avec le contenu des revues réelles. Un exemple tiré du jeu de données TripAdvisor est présenté dans la Table 6. Dans cet exemple, l'aspect le plus important pour l'utilisateur, selon l'attention, est *service*, tandis que pour l'article, c'est *location*. La comparaison de ces poids d'attention avec le contenu de la revue révèle une forte concordance avec les préférences de l'utilisateur et l'importance des aspects déduite par AURA. Notamment, *service* est l'aspect le plus fréquemment mentionné dans la revue avec un sentiment positif, tandis que *location*, l'aspect le plus important pour l'article, est également mis en avant. Ces analyses empiriques confirment que l'attention personnalisée d'AURA capte efficacement l'importance relative des aspects pour les utilisateurs et les articles. Au-delà de la prédiction des notes globales et par aspect et de la génération de revues, l'importance des aspects déduite peut servir de facteur explicatif supplémentaire, renforçant ainsi l'interprétabilité des recommandations.

5 CONCLUSION

Dans cet article, nous présentons AURA, un modèle multi-tâches composé d'un module de prédiction de notes — pour les notes globales et par aspect — et d'un module de génération de revue personnalisée. Notre approche apprend conjointement des représentations globales et spécifiques aux aspects pour les utilisateurs et les articles, et génère une invite continue personnalisée dans l'espace sémantique d'un modèle de langage. L'attention personnalisée nous permet de déduire l'importance relative des aspects pour les utilisateurs et les articles, confirmant notre hypothèse selon laquelle l'intégration d'informations basées sur les aspects améliore à la fois les recommandations et leurs explications. Nous implémentons AURA en utilisant le modèle de langage T5 pré-entraîné, en adoptant une stratégie de prompt tuning et un entraînement séquentiel pour exploiter efficacement le pré-entraînement. Nos expériences sur deux jeux de données multi-aspects réels, TripAdvisor et RateBeer, démontrent la supériorité d'AURA par rapport à des baselines performants pour l'ensemble des tâches, en particulier pour la génération de revues. Les analyses empiriques et les études d'ablation confirment les contributions significatives de chaque composant du modèle, en particulier la modélisation des aspects et l'attention personnalisée. Des travaux futurs viseront à adapter le modèle à des jeux de données ne disposant pas d'annotations explicites des aspects.

Références

BANERJEE S. & LAVIE A. (2005). Meteor : An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, p. 65–72.

BAUMAN K., LIU B. & TUZHILIN A. (2017). Aspect based recommendations : Recommending items with the most valuable aspects based on user reviews. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, p. 717–725.

BROWN T., MANN B., RYDER N., SUBBIAH M., KAPLAN J. D., DHARIWAL P., NEELAKANTAN A., SHYAM P., SASTRY G., ASKELL A. *et al.* (2020). Language models are few-shot learners. *Advances in neural information processing systems*, **33**, 1877–1901.

CHEN X., QIN Z., ZHANG Y. & XU T. (2016). Learning to rank features for recommendation over multiple categories. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, p. 305–314.

CHENG H., WANG S., LU W., ZHANG W., ZHOU M., LU K. & LIAO H. (2023). Explainable recommendation with personalized review retrieval and aspect learning. *arXiv preprint arXiv :2306.12657*.

CHENG Z., DING Y., ZHU L. & KANKANHALLI M. (2018). Aspect-aware latent factor model : Rating prediction with ratings and reviews. In *Proceedings of the 2018 world wide web conference*, p. 639–648.

CHIN J. Y., ZHAO K., JOTY S. & CONG G. (2018). Anr : Aspect-based neural recommender. In *Proceedings of the 27th ACM International conference on information and knowledge management*, p. 147–156.

DONG L., HUANG S., WEI F., LAPATA M., ZHOU M. & XU K. (2017). Learning to generate product reviews from attributes. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 1, Long Papers*, p. 623–632.

DUBEY A., JAUHRI A., PANDEY A., KADIAN A., AL-DAHLE A., LETMAN A., MATHUR A., SCHELTEN A., YANG A., FAN A. *et al.* (2024). The llama 3 herd of models. *arXiv preprint arXiv :2407.21783*.

HE X., LIAO L., ZHANG H., NIE L., HU X. & CHUA T.-S. (2017). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, p. 173–182.

KINGMA D. P. (2014). Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*.

KOREN Y., BELL R. & VOLINSKY C. (2009). Matrix factorization techniques for recommender systems. *Computer*, **42**(8), 30–37.

LESTER B., AL-RFOU R. & CONSTANT N. (2021). The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv :2104.08691*.

LI L., ZHANG Y. & CHEN L. (2021). Personalized transformer for explainable recommendation. *arXiv preprint arXiv :2105.11601*.

LI L., ZHANG Y. & CHEN L. (2023). Personalized prompt learning for explainable recommendation. *ACM Transactions on Information Systems*, **41**(4), 1–26.

LI P., WANG Z., REN Z., BING L. & LAM W. (2017). Neural rating regression with abstractive tips generation for recommendation. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, p. 345–354.

LIN C.-Y. (2004). Rouge : A package for automatic evaluation of summaries. In *Text summarization branches out*, p. 74–81.

LOPS P., DE GEMMIS M. & SEMERARO G. (2011). Content-based recommender systems : State of the art and trends. *Recommender systems handbook*, p. 73–105.

MCAULEY J., LESKOVEC J. & JURAFSKY D. (2012). Learning attitudes and attributes from multi-aspect reviews. In *2012 IEEE 12th International Conference on Data Mining*, p. 1020–1025 : IEEE.

NAIR V. & HINTON G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, p. 807–814.

NI J., LI J. & MCAULEY J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, p. 188–197.

NI J. & MCAULEY J. (2018). Personalized review generation by expanding phrases and attending on aspect-aware representations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, p. 706–711.

PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002). Bleu : a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, p. 311–318.

RACZYŃSKI J., LANGO M. & STEFANOWSKI J. (2023). The problem of coherence in natural language explanations of recommendations. In *ECAI 2023*, p. 1922–1929. IOS Press.

RADFORD A., WU J., CHILD R., LUAN D., AMODEI D., SUTSKEVER I. *et al.* (2019). Language models are unsupervised multitask learners. *OpenAI blog*, **1**(8), 9.

RAFFEL C., SHAZEER N., ROBERTS A., LEE K., NARANG S., MATENA M., ZHOU Y., LI W. & LIU P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, **21**(140), 1–67.

SHIMIZU R., WADA T., WANG Y., KRUSE J., O'BRIEN S., HTAUNGKHAM S., SONG L., YOSHIKAWA Y., SAITO Y., TSUNG F. *et al.* (2024). Disentangling likes and dislikes in personalized generative explainable recommendation. *arXiv preprint arXiv :2410.13248*.

SUN P., WU L., ZHANG K., SU Y. & WANG M. (2021). An unsupervised aspect-aware recommendation model with explanation text generation. *ACM Transactions on Information Systems (TOIS)*, **40**(3), 1–29.

VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L. & POLOSUKHIN I. (2017). Attention is all you need.

WANG H., LU Y. & ZHAI C. (2010). Latent aspect rating analysis on review text data : a rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 783–792.

ZHANG T., KISHORE V., WU F., WEINBERGER K. Q. & ARTZI Y. (2019). Bertscore : Evaluating text generation with bert. *arXiv preprint arXiv :1904.09675*.

ZHANG W., LI X., DENG Y., BING L. & LAM W. (2022). A survey on aspect-based sentiment analysis : Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, **35**(11), 11019–11038.

ZHANG Y., CHEN X. *et al.* (2020). Explainable recommendation : A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, **14**(1), 1–101.

Reviews

Review 1

Score : 1 (weak accept)
L'article intitulé Prédiction des préférences et génération de revue personnalisée basées sur les aspects et attention présente un modèle permettant d'avoir une prédiction des notes et la génération de reviews personnalisés. L'article est bien écrit et la problématique est bien posé. Le problème est bien formulé et le modèle bien décrit. De Nombreuses expérimentations sont proposés avec des tests d'ablation sur deux corpus de références et comparés à plusieurs baselines. Plusieurs questions et remarques restent toutefois à éclaircir. Premièrement, pourquoi avoir choisi T5 Comme base pour le développement du modèle ? De plus ample justifications permettrait d'être mieux convaincu de l'intérêt du modèle proposé. Au niveau des expérimentations, il aurait été intéressant de comparer le modèle présenté à des baseline un peu plus récente, où utilisant des LLM plus récent et ayant plus de paramètres. Les corpus choisis possèdent assez peu d'aspect, il aurait été intéressant d'avoir des corpus ayant un nombre d'aspects plus important. Du fait de l'utilisation de modèles peu récent, la longueur des reviews est limité à 128 mots ce qui est assez limité actuellement. De plus, les auteurs affirment que leur modèle est supérieur aux baselines or s y on regarde le tableau 2 ce n'est pas toujours vrai. Il aurait été intéressant d'avoir une discussion plus poussée pour expliquer ce fait.
Questions : Une indication qu'il peut aussi être ajoutée dans les expérimentations est comment a été sélectionnée la valeur de performance. Est-ce la meilleure ou bien est ce une moyenne de plusieurs runs ?
Réponse des auteurs : Pour chaque modèle, nous reportons la meilleure performance obtenue lors des différentes expérimentations.
Respect de la catégorie : yes

Review 2

Score : 3 (strong accept)
Cet article présente AURA, un modèle multitâche qui combine la prédiction des notes et la génération d'avis personnalisés pour les systèmes de recommandation. L'approche proposée apprend conjointement des représentations globales et spécifiques à chaque aspect pour les utilisateurs et les articles, tout en utilisant une attention personnalisée pour modéliser l'importance variable des différents aspects. AURA utilise une stratégie de prompt tuning avec le modèle linguistique T5 pré-entraîné pour générer des avis explicatifs. Des expériences menées sur les ensembles de données TripAdvisor et RateBeer démontrent les performances supérieures d'AURA par rapport à des baselines, en particulier dans la génération d'avis.
Points forts :

- l'article combine efficacement la modélisation basée sur les aspects avec des mécanismes d'attention personnalisés afin de saisir les préférences fines des utilisateurs et les caractéristiques des éléments ;
- l'architecture unifiée qui optimise conjointement la prédiction globale des notes, la prédiction des notes au niveau des aspects et la génération d'avis est bien conçue et théoriquement solide ;
- l'évaluation est approfondie et couvre plusieurs ensembles de données, métriques et références. Les études d'ablation démontrent clairement la contribution de chaque composant ;
- AURA surpasse largement les modèles de référence, en particulier en termes de métriques de

génération d'avis, avec des améliorations considérables par rapport aux approches de l'état de l'art.

Points faibles :

- l'article mentionne que les évaluations des aspects peuvent être explicitement disponibles ou extraites à l'aide d'une analyse des sentiments basée sur les aspects (ABSA), mais ne décrit pas clairement la méthode spécifique utilisée pour les ensembles de données expérimentaux. Plus de détails sur la manière dont les aspects ont été identifiés ou extraits renforceraient la compréhension et la reproductibilité ;
- ça serait intéressant d'avoir des informations sur le temps d'inférence et génération de revue, aussi bien que le temps d'entraînement ;
- l'article s'appuie exclusivement sur des mesures automatiques pour évaluer les avis générés, l'évaluation humaine pourrait renforcer les contributions ;
- une analyse qualitative examinant des aspects tels que la cohérence, l'exactitude factuelle et la spécificité des avis générés améliorerait l'évaluation

Questions : Comment le modèle pourrait être adapté à des domaines où des annotations explicites sur les aspects ne sont pas disponibles ?

Réponse des auteurs : Nous travaillons sur différentes extensions du modèle à des jeux de données non annotés. Une première adaptation consiste à extraire les aspects des revues en amont, avec des techniques d'ABSA ou en few-shot avec des modèles de langue et utiliser les données extraites en entrée du modèle. Une seconde extension consiste à apprendre conjointement à extraire les aspects de la revue et à reconstruire la revue dans une approche inspirée des auto-encodeurs.

Respect de la catégorie : yes

Review 3

Score :3 (strong accept)

L'article présente, un modèle multitâche (AURA) pour la recommandation personnalisée qui combine deux modules principaux : la prédiction des notes (à la fois globales et spécifiques à différents aspects comme la propreté, l'emplacement, etc.) et la génération personnalisée de revues (explications en langage naturel des recommandations). AURA apprend conjointement des représentations globales et locales (orientées aspects) pour les utilisateurs et les articles. Il utilise une attention personnalisée pour moduler l'importance relative de chaque aspect selon les préférences de l'utilisateur et les caractéristiques de l'article. Cette attention est utilisée à la fois pour améliorer la qualité des prédictions et la pertinence des explications générées. Le modèle repose sur une stratégie de prompt tuning et un entraînement séquentiel appliqués sur le modèle préentraîné T5. Des expérimentions sont réalisés sur deux jeux de données et multi-aspects, TripAdvisor (hôtels avec six aspects : propreté, localisation, chambre, service, sommeil, rapport qualité-prix) et RateBeer (bières avec quatre aspects : apparence, arôme, palais, goût). L'article est à la fois bien rédigé et bien construit (structuré). La contribution principale réside dans l'exploitation conjointe de l'importance relative des aspects pour chaque utilisateur et chaque article, grâce à une attention personnalisée, ainsi que dans l'intégration des opinions textuelles et des notes par aspect afin d'améliorer à la fois la qualité des recommandations et la pertinence des explications générées. Le travail expérimental est très bien construit. Les résultats montrent qu'AURA surpasse des modèles de référence. Les études d'ablation confirment l'apport de chaque composant, notamment la modélisation des aspects et l'attention personnalisée.

Respect de la catégorie : yes