ITERATIVE MEMORY NETWORK ET DEEP INTEREST NETWORK POUR LA MODÉLISATION DU COMPORTEMENT D'UTILISATEUR DE LONG TERME

YUAN Fangzheng, **HUANG Bozhang**, **LI Haoran**,

Master2 informatique DAC, Sorbonne Université

Introduction

Les systèmes de recommandation prennent en charge les décisions dans divers domaines tels que les services financiers, les équipements de télécommunication et les systèmes logiciels. Dans ce contexte, des recommandations sont déterminées, par exemple, par l'analyse des préférences des utilisateurs dans une séquence temporelle. Dans le cadre de notre projet, nous découvrirons certains états de l'art les plus récents dans ce domaine tel que:

- Deep Interest Network (DIN)
- Iterative Memory Network (IMN)

Formulation du problème:
Soit utilisateur $u \in U$ et objet $i \in I$, $c_{mn} = 1$ indique que l'utilisateur u_m a cliqué sur l'objet i_n. Le problème consiste à prédire la probabilité qu'un utilisateur u clique sur un objet i se basant sur la séquence historique $i_1, i_2, i_3, ..., i_L$ avec L la longueur de la séquence.

Deep Interest Network

Couche Embedding: Le comportement d'un utilisateur se modélise par une séquence d'objets qu'il interagit. Chaque objet i est un vecteur de caractéristique (e.g. l'identifiant de l'objet, la catégorie) $i = (f_1, f_2, ..., f_N)$ avec f_i le feature. N couches embeddings $\boldsymbol{e}$ seront appliqués sur chaque feature différent et les résultats sont agrégés par une concaténation :

$$\boldsymbol{e}(i) = Concat(\boldsymbol{e_i}(f_1), \boldsymbol{e_2}(f_2), ...\boldsymbol{e_n}(f_n))$$

Modèle de base: Le modèle de base obtient un vecteur d'embedding des intérêts de l'utilisateur en regroupant tous les vecteurs de comportements d'utilisateurs.

- Le vecteur de représentation de l'utilisateur avec une dimension limitée sera un goulot d'étranglement pour exprimer les divers intérêts de l'utilisateur.
- risque de sur-apprendre et consommer la mémoire pour stocker l'embedding de grande taille.

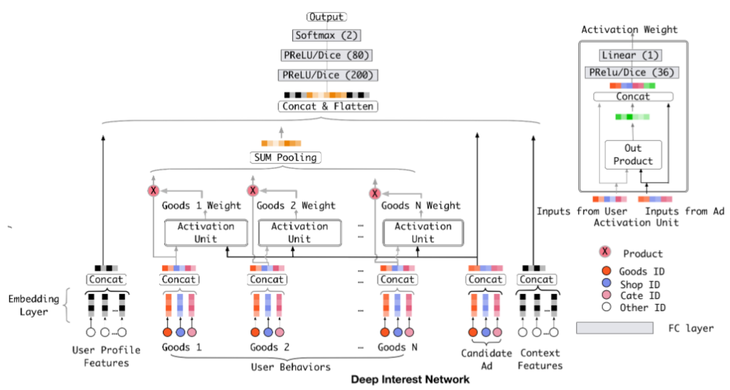

Fig. 1: Architecture du DIN

Modèle avec l'unité d'activation locale (DIN): L'unité d'activation sert à capturer l'intérêt de l'utilisateur. La méthode est inspirée par le mécanisme d'attention. Il permet de pondérer le comportement historique pour calculer l'embedding final.

$$\boldsymbol{v}_U(A) = f\left(\boldsymbol{v}_A, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N\right) = \sum_{j=1}^{N} a\left(\boldsymbol{e}_j, \boldsymbol{v}_A\right) \boldsymbol{e}_j = \sum_{j=1}^{N} w_j \boldsymbol{e}_j$$

étant $\boldsymbol{v}_A$ l'objet candidate et $\boldsymbol{v}_U(A)$ l'embedding de comportement d'utilisateur U par rapport à ce candidat.

Interative Memory Network

Modèle IMN: Un modèle qui combine le mécanisme d' *Attention* et celui de *GRU* pour retirer et mémoriser les similarités entre le target objet et la séquence d'utilisateur pour la prédiction des comportements d'utilisateur de longue séquence. L'objet de target est calculé plusieurs fois avec les objet dans la séquence pour retirer les informations communes. Il calcule non seulement la similarité entre la séquence et le target mais aussi celle entre les objets dans la séquence.

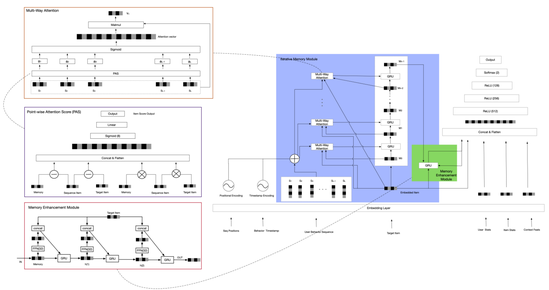

Fig. 2: Architecture de IMN

Le réseau est composé de 4 composants :

- **Encoder Layer** : encoder la séquence des utilisateurs en sommant les embeddings d'objet dans la séquence, de l'intervalle de temps de l'objet et de sa position dans la séquence:

$$\boldsymbol{e}_b^j = \boldsymbol{e}_i \oplus \boldsymbol{e}_t \oplus \boldsymbol{e}_p$$

- **Multi-Way Attention** : calculer le poids d'attention de chaque objet dans la séquence:

$$\alpha(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{m}) = [\boldsymbol{e} - \boldsymbol{m}, \boldsymbol{e} - \boldsymbol{v}, \boldsymbol{e} \circ \boldsymbol{m}, \boldsymbol{e} \circ \boldsymbol{v}]$$

$$a(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{m}) = \sigma\left(W^{(2)}\sigma\left(W^{(1)}\alpha(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{m}) + b^{(1)}\right) + b^{(2)}\right)$$

- **Iterative Memory Update Module**: retirer les similarités entre le target objet et les objets de séquence:

$$\boldsymbol{m}^0 = \boldsymbol{v}_T$$

,

$$\boldsymbol{v}_u^t = f\left(\boldsymbol{v}_T, \boldsymbol{e}_b^1, \boldsymbol{e}_b^2, \ldots, \boldsymbol{e}_b^L\right) = \sum_{j=1}^{L} a\left(\boldsymbol{e}_b^j, \boldsymbol{v}_T, \boldsymbol{m}^{t-1}\right) \boldsymbol{e}_b^j = \sum_{j=1}^{L} w_j \boldsymbol{e}_b^j$$

$$\boldsymbol{m}^t = GRU\left(\boldsymbol{v}_u^t, \boldsymbol{m}^{t-1}\right)$$

- **Memory Enhancement Module**: retirer plus d'informations entre le target et la séquence et réduire l'erreur:

$$\boldsymbol{u}_0 = \boldsymbol{m}^N$$

,

$$\boldsymbol{u}^t = GRU\left(\left[W^u \boldsymbol{u}^{t-1}, \boldsymbol{v}_T\right], \boldsymbol{u}^{t-1}\right)$$

Protocole d'Évaluation

Amazon Dataset. Cet ensemble de données contient des avis sur les produits et les métadonnées d'Amazon. Il contient des avis (notes, texte), des métadonnées de produits (descriptions, catégories de produits, prix, marque et caractéristiques de l'image). Sans perte de généralité, les datasets **Electronics**, **Kindle_Store**, **CDs_and_Vinyl** sont choisis pour l'évaluation.

Nom	nb. user	nb. item	nb. category	nb. samples	max sequence length
Elec	192403	63001	801	1689188	136
Kindle	39387	23033	484	278677	431
CD	68223	61934	1386	982619	3583

Métrique d'évaluation. AUC (Area Under the Curve) pour mesurer la performance des modèles.

Résultat & Comparaison

tableau de performance:

Modèle	Elec AUC	Elec Impr	Kindle AUC	Kindle Impr	CD AUC	CD Impr
BASE	0.587	0.000	0.506	0.000	0.500	0.000
DIN	0.623	0.036	0.547	0.041	/	/
IMN 2P	0.755	0.168	0.631	**0.125**	0.638	0.138
IMN 3P	0.756	**0.169**	0.631	**0.125**	**0.639**	**0.139**

étant Impr (improvement) l'augmentation de la performance par rapport au baseline.

tableau du temps d'entraînement (Unité : s/epoch) :

Modèle	Elec	Kindle	CD
BASE	53s	41s	102s
DIN	120s	228s	∞
IMN 2P	78s	93s	280s
IMN 3P	88s	112s	364s

Conclusion

Dans ce projet, nous avons implémenté et testé les 2 états de l'art les plus récents dans le domaine du système de recommandation. L'utilisation de la représentation de longueur fixe est un goulot d'étranglement pour capturer le diversité des intérêts des utilisateurs. Pour améliorer la capacité expressive de modèle, DIN a proposé une solution inspirée par le mécanisme d'attention. Alors que le modèle IMN combine l'attention et GRU pour retirer les intérêts des utilisateurs. D'un point de vue expérimentale, les défis sont la vitesse d'entraînement. La contribution de IMN sont surtout l'accélération sur l'entraînement et la capacité de traiter les séquences longues..

References

[1] Guorui Zhou, Chengru Song (2018) Deep Interest Network for Click-Through Rate Prediction, Alibaba Group.

[2] Anonymous authors (2021) Iterative Memory Network for Long Sequential User Behavior Modeling in Recommender Systems