Une approche tensorielle pour entrainer des réseaux de neurones flexibles*

Yassine ZNIYED[1], Konstantin USEVICH[2], Sebastian MIRON[2], David BRIE[2]

[1]Université de Toulon, Aix Marseille Université, CNRS, LIS, Toulon, France

[2]Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

yassine.zniyed@univ-tln.fr, {konstantin.usevich, sebastian.miron, david.brie}@univ-lorraine.fr

Résumé – Dans ce travail, nous nous concentrons sur la compression des réseaux de neurones (RNs), en considérant une approche basée sur les tenseurs. En particulier, nous nous intéressons à l'estimation de fonctions d'activation flexibles à l'aide de solutions basées sur des tenseurs. La méthode proposée est basée sur une factorisation couplée matrice-tenseur (FCMT), fusionnant les informations d'ordre 0 et d'ordre 1 des RNs. L'algorithme d'apprentissage proposé est basé sur une approche contrainte des moindres carrés alternés. Les fonctions d'activation (FAs) sont exprimées sous la forme d'une somme pondérée d'éléments de base prédéfinis. Notre méthode permet de compresser fortement des grandes couches des RNs, avec une très légère dégradation de la précision de la classification. Cela donne également un aperçu du fonctionnement des RNs, qui est un problème du plus haut intérêt dans ce domaine.

Abstract – In this work, we focus on the compression of Neural Networks (NNs), considering a tensor-based approach. In particular, we are interested in the estimation of flexible activation functions using tensor-based solutions. The proposed method is based on a coupled matrix-tensor factorization (CMTF), fusing the zeroth and first order information of the NN. The proposed learning algorithm is based on a constrained alternating least squares (ALS) approach. The activation functions (AFs) are expressed as a weighted sum of pre-defined basis elements. Our method allows to strongly compress large NN layers, with a very slight degradation of the classification accuracy. This also provides insights into the functioning of NNs, which is a problem of utmost interest in NN domain.

1 Introduction

Les réseaux de neurones (RNs) sont des outils puissants, qui ont la capacité d'apprendre et de modéliser des fonctions non linéaires complexes, tout en montrant de bonnes performances de prédiction. Avant d'entraîner un réseau de neurones, il faut décider (i) combien de neurones et de couches cachées utiliser ? et (ii) quelles fonctions d'activation (FAs) utiliser ? Ces choix contrôlent la capacité d'apprentissage du RN à partir des données. Les architectures des RNs modernes de pointe sont trop volumineuses et comportent de nombreux paramètres. Les exécuter dans des systèmes avec une capacité de calcul limitée est une tâche difficile. Pour cette raison, plusieurs travaux dans la littérature se sont intéressés à la compression de ces modèles.

Les approches tensorielles les plus populaires [1] pour la compression visent principalement à compresser les couches des réseaux convolutifs, qui peuvent être considérés comme des tenseurs. Par une décomposition polyadique canonique (CPD) [2], ils remplacent les convolutions multidimensionnelles par des convolutions unidimensionnelles. Une autre direction de recherche se concentre sur la relation entre les décompositions de tenseurs et les réseaux de neurones avec des unités de produit (au lieu d'unités de sommation) [3] ; ce type de représentations est cependant peu utilisé en pratique.

Dans ce travail, nous considérons une approche totalement différente. Tout en conservant la structure traditionnelle du réseau de neurones (poids linéaires + fonctions d'activation non linéaires), nous visons à ajouter de la flexibilité [4] aux FAs, par opposition aux FAs fixes utilisées classiquement. En particulier, les fonctions d'activation peuvent être différentes dans différents nœuds. Une telle architecture est particulièrement intéressante grâce à la théorie de l'identifiabilité (unicité) disponible dans le cas polynomial [5]. Les propriétés d'identifiabilité peuvent donner un aperçu du fonctionnement de ces RNs et aider à renforcer la stabilité de la représentation.

Contrairement aux méthodes existantes pour les FAs flexibles (FAFs) qui utilisent des techniques d'entraînement conventionnelles [4], nous utilisons un cadre original développé dans la communauté d'identification de systèmes non linéaires. Les travaux de [6] ont montré qu'une architecture avec une couche flexible cachée peut être identifiée comme une CPD d'un tenseur jacobien. Cependant, il n'est pas directement applicable dans la configuration d'apprentissage ; en particulier, il n'y a pas de moyen simple d'estimer les fonctions d'activation. Dans ce travail, nous proposons une nouvelle méthode de compression de réseaux de neurones pré-entraînés basée sur la factorisation couplée matrice-tenseur (FCMT). Cette technique fusionne les informations d'ordre 0 et 1 du RN, où les informations de premier ordre sont contenues dans un tenseur jacobien,

*Travail soutenu par l'ANR LeaFleT (ANR-19- CE23-0021).

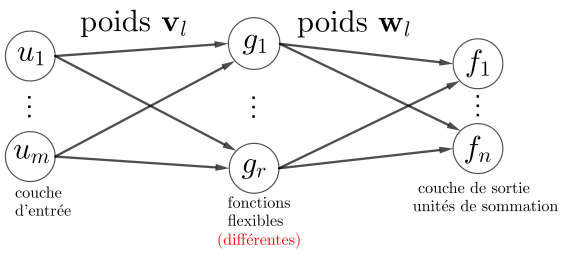

FIGURE 1 – Représentation graphique du bloc de base à une couche du RN flexible.

suivant une CPD contrainte.

Le nouveau RN flexible a des FAs différentes et flexibles dans chaque neurone. Les FAFs sont exprimées sous forme de somme pondérée de fonctions de base prédéfinies. Le but de ce modèle est d'augmenter l'expressivité du modèle pour pouvoir apprendre des fonctions avec un nombre de paramètres inférieur par rapport aux modèles classiques. Ce genre de résultat est susceptible d'accélérer l'exécution des réseaux de neurones et permettra leur utilisation sur des systèmes à faibles ressources de calcul.

2 Modèle de FAFs

Modèle flexible Le modèle proposé dans ce travail est un RN flexible, où chaque neurone a une FA différente et flexible. Les FAs sont exprimées sous la forme d'une combinaison linéaire de fonctions de base prédéfinies. Une illustration graphique d'un bloc de base à une couche de ce modèle est donnée dans la Figure 1. La fonction multivariée qui représente la relation entrée-sortie est notée $\boldsymbol{f}$. Cette fonction $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$ est donnée par :

$$\boldsymbol{f}(\boldsymbol{u}) = [f_1(\boldsymbol{u}) \quad f_n(\boldsymbol{u})]^\top, \quad \text{with } \boldsymbol{u} = [u_1 \quad u_m]^\top. \quad (1)$$

On dit que $\boldsymbol{f}$ admet un modèle RN flexible s'il peut s'exprimer sous la forme :

$$\boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{W}\boldsymbol{g}(\boldsymbol{V}^\top \boldsymbol{u}), \quad (2)$$

où $\boldsymbol{V} \in \mathbb{R}^{m \times r}$, $\boldsymbol{W} \in \mathbb{R}^{n \times r}$ sont des matrices de transformation, $\boldsymbol{v}_\ell$ et $\boldsymbol{w}_\ell$, pour $1 \leq \ell \leq r$, sont respectivement les colonnes de $\boldsymbol{V}$ et $\boldsymbol{W}$, et $\boldsymbol{g} : \mathbb{R}^r \to \mathbb{R}^r$ est exprimée comme

$$\boldsymbol{g}(t_1, \ldots, t_r) = [g_1(t_1) \quad g_r(t_r)]^\top, \quad (3)$$

avec $g_\ell : \mathbb{R} \to \mathbb{R}$ est une fonction non linéaire univariée.

Fonctions de base Les fonctions non linéaires univariées g_ℓ ($1 \leq \ell \leq r$) dans (2) sont une combinaison linéaire de fonctions de base prédéfinies $\{\phi_1, \ldots, \phi_d\}$:

$$g_\ell(t) = c_{0,\ell} + c_{1,\ell}\phi_1(t) + \quad + c_{d,\ell}\phi_d(t). \quad (4)$$

Les fonctions ϕ_k, pour $1 \leq k \leq d$, dans (4) étant *a priori* choisies, les paramètres $c_{k,\ell}$, pour $0 \leq k \leq d$ et $1 \leq \ell \leq r$, sont les paramètres apprenables.

3 Décomposition d'une fonction multivariée basée sur un tenseur

Représenter une fonction non-linéaire $\boldsymbol{f}$ par une représentation plus simple, telle que (2) a été abordée dans plusieurs travaux, puisqu'elle permet d'approximer la fonction avec un nombre réduit de paramètres et peut donner un aperçu de la compréhension des mécanismes derrière la fonction non linéaire. Ces méthodes supposent que la fonction $\boldsymbol{f}$ est connue à travers une représentation paramétrique polynomiale, et ont été développées dans le contexte du problème de découplage pour les polynômes réels multivariés. L'une de ces méthodes est la méthode jacobienne [6] qui vise à décomposer $\boldsymbol{f}(\boldsymbol{u})$ selon (2) à partir de son information de premier ordre. Cette approche repose sur l'observation que la matrice jacobienne de $\boldsymbol{f}$, évaluée en un point donné $\boldsymbol{u}$ s'exprime par

$$\boldsymbol{J}_f(\boldsymbol{u}) := \begin{bmatrix} \frac{\partial f_1}{\partial u_1}(\boldsymbol{u}) & \frac{\partial f_1}{\partial u_m}(\boldsymbol{u}) \\ \vdots & \vdots \\ \frac{\partial f_n}{\partial u_1}(\boldsymbol{u}) & \frac{\partial f_n}{\partial u_m}(\boldsymbol{u}) \end{bmatrix} \quad (5)$$

$$= \boldsymbol{W} \; diag\left(g_1'(\boldsymbol{v}_1^\top \boldsymbol{u}) \quad g_r'(\boldsymbol{v}_r^\top \boldsymbol{u})\right) \; \boldsymbol{V}^\top. \quad (6)$$

En empilant les matrices jacobiennes évaluées en N points d'échantillonnage différents $\boldsymbol{u}^{(j)} \in \mathbb{R}^m$, pour $1 \leq j \leq N$, tel que $\boldsymbol{\mathcal{J}}_{:,:,j} = \boldsymbol{J}_f(\boldsymbol{u}^{(j)})$, nous construisons un tenseur jacobien d'ordre 3 $\boldsymbol{\mathcal{J}}$ de taille $n \times m \times N$ suivant une CPD exprimée comme

$$\boldsymbol{\mathcal{J}} = \sum_{\ell=1}^{r} \boldsymbol{w}_\ell \circ \boldsymbol{v}_\ell \circ \boldsymbol{h}_\ell, \quad (7)$$

où $\circ$ dénote le produit tensoriel. De manière équivalente, on note $\boldsymbol{\mathcal{J}} = [\![\boldsymbol{W}, \boldsymbol{V}, \boldsymbol{H}]\!]$. Comme montré dans [6, 7], les colonnes de $\boldsymbol{H} \in \mathbb{R}^{N \times r}$ ont la structure suivante

$$\boldsymbol{h}_\ell = [g_\ell'(\boldsymbol{v}_\ell^\top \boldsymbol{u}^{(1)}), \quad , g_\ell'(\boldsymbol{v}_\ell^\top \boldsymbol{u}^{(N)})]^\top, \quad (8)$$

ce qui montre que le facteur matriciel $\boldsymbol{H}$ dépend de $\boldsymbol{V}$. Il convient de noter que l'approche originale de [6] est assez sensible à la non-unicité de la CPD et au bruit [8, 9], puisqu'une solution ALS relaxée est utilisée pour estimer les facteurs.

Suivant [8, 9], cela peut être attribué au fait que l'algorithme ne tient pas compte de la structure particulière du tenseur jacobien (8), *ie* la relation entre les colonnes de $\boldsymbol{H}$ et $\boldsymbol{V}$. Cela devient problématique, en particulier lorsque l'unicité de la CPD est perdue, auquel cas l'ALS relâché ne parviendra pas à récupérer des FAs significatives. Cela a motivé le développement d'algorithmes structurés pour le découplage de polynômes multivariés [8, 9]. De plus, il a été montré dans [7] que la prise en compte de la structure (8) conduit à des conditions d'unicité moins restrictives par rapport au cas non structuré, élargissant ainsi le domaine d'application de la méthode.

4 Problème d'apprentissage

Pour surmonter les inconvénients susmentionnés des méthodes existantes, nous proposons dans cette section un algorithme

pour la CPD structurée du tenseur jacobien. Il consiste à formuler la CPD structurée comme une factorisation couplée matrice-tenseur contrainte, utilisant conjointement les informations d'ordre 0 et 1.

Information d'ordre 0 Soit F une matrice de taille $n \quad N$ dont les lignes sont les évaluations des n sorties du sous-réseau d'origine à N points d'échantillonnage différents $u^{(j)} \; 2 \; \mathsf{R}^m$:

$$F = \begin{bmatrix} f_1(u^{(1)}) & & f_1(u^{(N)}) \\ \vdots & & \vdots \\ f_n(u^{(1)}) & & f_n(u^{(N)}) \end{bmatrix}. \tag{9}$$

En se basant sur (9) et (2), nous pouvons voir que F est sujet à une factorisation donnée par

$$F = \begin{bmatrix} f_1(u^{(1)}) & & f_1(u^{(N)}) \\ \vdots & & \vdots \\ f_n(u^{(1)}) & & f_n(u^{(N)}) \end{bmatrix} \tag{10}$$

$$= W \underbrace{\begin{bmatrix} g_1(v_1^\top u^{(1)}) & & g_1(v_r^\top u^{(N)}) \\ \vdots & & \vdots \\ g_r(v_1^\top u^{(1)}) & & g_r(v_r^\top u^{(N)}) \end{bmatrix}}_{Z^\top}, \tag{11}$$

où Z est de taille $N \quad r$. On peut voir à partir de (11) que la matrice F partage le même facteur W avec le tenseur J. Sur la base de cette observation, nous construisons un problème FCMT contraint, impliquant la matrice F et le tenseur J.

Algorithme d'apprentissage L'idée de l'approche basée sur la FCMT contrainte est de formuler l'apprentissage des paramètres, $\mathbf{W}$, $\mathbf{V}$ et $c_{k,\cdot}$, pour $0 \quad k \quad d$ et $1 \quad \ell \quad r$, comme suit.

$$\min_{\mathbf{W},\mathbf{V},\mathbf{H},\mathbf{Z}} \quad J \quad [j\mathbf{W}, \mathbf{V}, \mathbf{H}j] \quad ^2 + \lambda \quad \mathbf{F} \quad \mathbf{W}\mathbf{Z}^T \quad ^2 \tag{12}$$

$$\text{t.q.} \quad h_{j,\cdot} = c_{1,\cdot}\phi_1^0(\mathbf{v}_\cdot^T \mathbf{u}^{(j)}) + \quad + c_{d,\cdot}\phi_d^0(\mathbf{v}_\cdot^T \mathbf{u}^{(j)}),$$
$$z_{j,l} = c_{0,l} + c_{1,\cdot}\phi_1(\mathbf{v}_\cdot^T \mathbf{u}^{(j)}) + \quad + c_{d,\cdot}\phi_d(\mathbf{v}_\cdot^T \mathbf{u}^{(j)}).$$

Ces contraintes s'expriment de manière équivalente comme

$$h_\cdot = X_\cdot \quad c_\cdot, \quad z_\cdot = Y_\cdot \quad c_\cdot, \quad \text{for } 1 \quad \ell \quad r,$$

où les matrices $X_\cdot$ et $Y_\cdot$, de taille $N \quad (d+1)$, sont données par

$$X_\cdot = \begin{bmatrix} 0 & \phi_1^0(v_\cdot^\top u^{(1)}) & & \phi_d^0(v_\cdot^\top u^{(1)}) \\ \vdots & \vdots & & \vdots \\ 0 & \phi_1^0(v_\cdot^\top u^{(N)}) & & \phi_d^0(v_\cdot^\top u^{(N)}) \end{bmatrix}, \tag{13}$$

$$Y_\cdot = \begin{bmatrix} 1 & \phi_1(v_\cdot^\top u^{(1)}) & & \phi_d(v_\cdot^\top u^{(1)}) \\ \vdots & \vdots & & \vdots \\ 1 & \phi_1(v_\cdot^\top u^{(N)}) & & \phi_d(v_\cdot^\top u^{(N)}) \end{bmatrix}, \tag{14}$$

et le ℓ-ième vecteur de coefficient $c_\cdot$ est défini comme $c_\cdot = \begin{bmatrix} c_{0,\cdot} & c_{1,\cdot} & \dots & c_{d,\cdot} \end{bmatrix}^\top$.

Pour résoudre le problème (12), nous proposons un algorithme des moindres carrés alternés à facteurs H et Z contraints, dans lequel nous mettons à jour les coefficients des FAs $g_1, \quad , g_r$ selon (4). La procédure complète est donnée dans Algorithme 1, où dénote le produit de Khatri-Rao.

Algorithme 1 Algorithme d'apprentissage basé sur la FCMT

Entrée : Tenseur J de dimenson $n \quad m \quad N$, matrice F de taille $n \quad N$, les fonctions $f\phi_1, \dots, \phi_d g$, paramètre d'ajustement λ, rang r.
Sortie : Facteurs W, V, H et Z, et les coefficients $c_\cdot$.

1: Initialiser V, H, Z
2: **répéter**
3: Mettre à jour W avec $\min_W \; \text{unfold}_1 J \quad W \quad H \quad V \quad ^\top \; ^2 + \lambda \quad F \quad W Z^\top \; ^2$
4: Mettre à jour V avec $\min_V \; \text{unfold}_2 J \quad V \quad H \quad W \quad ^\top \; ^2$
5: Mettre à jour H avec $\min_H \; \text{unfold}_3 J \quad H \quad V \quad W \quad ^\top \; ^2$
6: Mettre à jour Z avec $\min_Z \; F \quad W Z^\top \; ^2$
7: **pour** $\ell = 1 \dots r$ **faire**
8: Construire $X_\cdot$ selon (13).
9: Construire $Y_\cdot$ selon (14).
10: **fin pour**
11: Calculer c comme $\min_c \; \text{vec}(H) \quad \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_r \end{bmatrix} \quad ^2 c + \lambda \quad \text{vec}(Z) \quad \begin{bmatrix} Y_1 & & 0 \\ & \ddots & \\ 0 & & Y_r \end{bmatrix} \quad ^2 c$
12: **pour** $\ell = 1 \dots r$ **faire**
13: Calculer $h_\cdot$ t.q : $h_\cdot = X_\cdot \quad c_\cdot$
14: Calculer $z_\cdot$ t.q : $z_\cdot = Y_\cdot \quad c_\cdot$
15: **fin pour**
16: **jusqu'à** un test de convergence est satisfait ou le nombre maximal d'itérations est atteint
17: **retourne** W, V, $c_\cdot$.

5 Compression des réseaux de neurones pré-entraînés

Pour illustrer l'efficacité de l'apprentissage de réseaux de neurones flexibles basé sur des tenseurs, la compression de réseaux de neurones pré-formés est considérée dans cette section. L'un des principaux objectifs de la compression est de réduire la complexité d'un réseau de neurones afin de permettre son implémentation sur des systèmes embarqués avec des capaci-