Rapport Recherche d'Informations Visuelles Structurées

Loïc HERBELOT et Sébastien PEREIRA

29 décembre 2017

Résumé

Dans ces TME nous allons nous intéresser au paradigme d'apprentissage structuré appliqué à différentes problématiques de recherche d'informations visuelles. On instanciera dans un premier temps ce problème pour résoudre des tâches de classification d'images puis dans un second temps pour résoudre un problème de classement renvoyant une liste d'images qui correspondent à la recherche de l'utilisateur.

1 Introduction

Contrairement à la régression où notre espace de sortie est $Y = \mathbb{R}$ et à la classification où $Y = \{1, ..., C\}$ (où C est le nombre de classes), l'apprentissage structuré définit un espace de sortie Y où les $y \in Y$ sont des structures comme des arbres, des graphes, des ordonnancements etc.

1.1 Formalisme

Soit X l'espace d'entrée et Y l'espace de sortie structuré. On définit une *joint feature map* $\Psi(x, y)$ qui décrit le lien entre l'entrée $x \in X$ et la sortie $y \in Y$ et une fonction de coût $\Delta(y_i, \hat{y})$ qui mesure la dissimilarité entre notre prédiction $\hat{y}$ et la vérité terrain y_i. Le problème d'apprentissage revient à trouver un modèle linéaire qui donne un score à chaque paire (x, y), le score est défini par $< w, \Psi(x, y) >$, notre paramètre est $w \in \mathbb{R}^d$ La prédiction du modèle est de la forme :

$$\hat{y}(x, w) = \arg\max_{y \in Y} < w, \Psi(x, y) > \tag{1}$$

De plus on souhaite avoir une régularisation sur le paramètre w, le problème d'apprentissage s'écrit :

$$\min_w \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n} \Delta(y_i, \hat{y}) \tag{2}$$

Ce problème d'optimisation étant NP difficile, on optimise une borne supérieure convexe et sous différentiable de $\Delta(y_i, \hat{y})$ définie par :

$$\max_{y \in Y}[\Delta(y_i, \hat{y}) + < \Psi(x_i, y), w >] - < \Psi(x_i, y_i), w > \tag{3}$$

Le problème d'optimisation devient donc :

$$\min_w \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n}[\max_{y\in Y}[\Delta(y_i,\hat{y}) + < \Psi(x_i,y), w >] - < \Psi(x_i,y_i), w >] \quad (4)$$

1.2 Algorithme d'apprentissage

- Utilisation d'une descente de gradient stochastique pour trouver w qui minimise le pb

2 Apprentissage structuré multi-classes et hiérarchique

Dans la première partie on s'intéresse à la classification d'images à partir d'un apprentissage supervisé, on observera l'influence du choix de la fonction de coût sur les résultats obtenus.

2.1 Classification multi-classe

Dans le cas d'un problème d'apprentissage structuré instancié pour les problèmes de classification multi-classe on définit la *joint feature map* et la fonction de coût de la façon suivante :

— $\Psi(x,y) = \begin{bmatrix} 0^d \\ 0^d \\ ... \\ \phi(x) \\ ... \\ 0^d \\ 0^d \end{bmatrix}$

— $\Delta(y,y') = \begin{cases} 1 & \text{si } y \neq y' \\ 0 & \text{sinon.} \end{cases}$

Où $\phi(x)$ est la représentation de l'image sous forme d'un vecteur *Bag of Words*. $\Psi(x,y)$ est donc un vecteur de taille $C * d$ rempli de 0, sauf aux indices $i(y) * d, ..., (i(y)+1) * d$, avec $i(y)$ le numéro associé à la classe y.

2.2 Interprétation des résultats

— Les images 'taxi' (colonne 0) sont classées comme des 'taxi' ce qui est rassurant, mais aussi comme des 'minivan' (ligne 2) ou des 'ambulance' (ligne 1)
— Les 'ambulance' (colonne 1) sont classées comme des 'ambulance', mais aussi comme des 'minivan' ou des 'taxi'.
— la classe minivan (2) est confondue avec la classe ambulance (1)
— la classe acoustic guitar (3) est très confondue avec la classe electric guitar (4) et énormément confondue avec la classe harp (5)
— la classe electric guitar (4) est très confondue avec la classe harp (5)
— la classe harp (5) est très bien distinguée de toutes les autres classes
— la classe wood grog (6) est beaucoup confondue avec la classe tree grog (7) et énormément confondue avec la classe european fire salamander (8)
— la classe tree grog (7) est un peu confondue avec la classe european fire salamander (8)

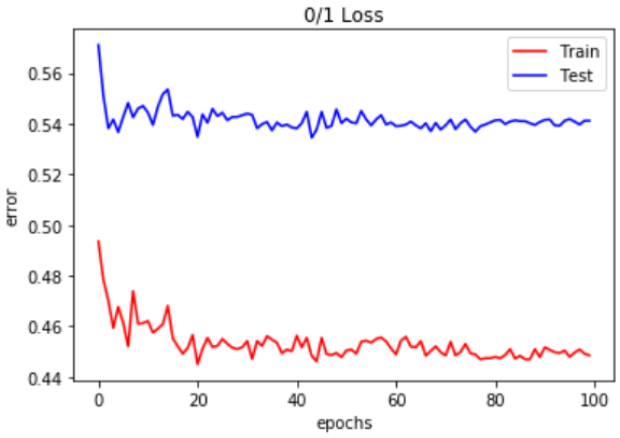


FIGURE 1 – Évaluation de la 0/1 loss en train et en test sur 100 itérations.

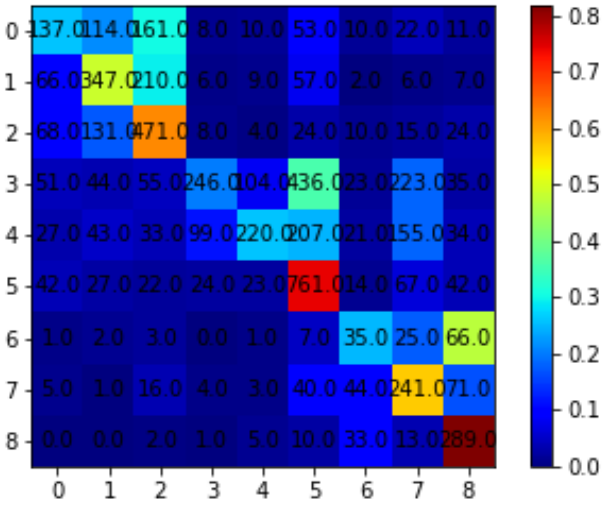


FIGURE 2 – Matrice de confusion pour un modèle entrainé avec une 0/1 loss sur 100 epochs. 0 = taxi, 1 = ambulance, 2 = minivan, 3 = acoustic guitar, 4 = electric guitar, 5 = harp, 6 = wood frog, 7 = tree frog, 8 = european fire salamander

— la classe european fire salamander (8) est très bien distinguée de toutes les autres classes

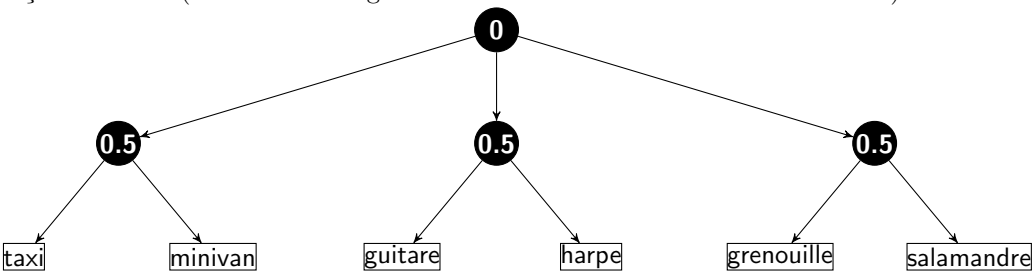
Conclusion : on remarque qu'il existe 3 catégories sémantiques dans nos 9 classes, la première catégorie sémantique rassemble des véhicules (taxi, ambulance et minivan) la deuxième catégorie sémantique rassemble des instruments à cordes (guitare acoustique, guitare electrique et harpe) enfin la troisième catégorie sémantique rassemble des amphibiens (grenouille des bois, grenouille arboricole et salamandre). De plus on voit d'après la matrice de confusion que la grande majorité des erreurs de notre modèle provient de confusions au sein d'une même catégorie sémantique mises à part deux exceptions (harpe et salamandre), on peut donc en conclure que notre modèle n'arrivent pas à capter les différences entre les données provenant d'une même catégorie sémantique.

2.3 Classification Hierarchique

Dans le cas d'un problème d'apprentissage structuré instancié pour les problèmes de classification hierarchique on garde la *joint feature map* précédente et on définit une fonction de coût de la façon suivante :

— $\Delta(y, y') = \begin{cases} 1 - \textit{Similarité}(y, y') & \text{si } y \neq y' \\ 0 & \text{sinon.} \end{cases}$

Où la similarité est définit par une distance dans un arbre par exemple de la façon suivante (les noeuds désignent les similarités et les feuilles les classes) :



2.4 Inteprétation des résultats

3 Apprentissage structuré appliqué au problème de ranking

Dans cette seconde partie on s'intéresse à la résolution d'un problème d'ordonnancement grâce à l'apprentissage structuré.

3.1 Ranking

On définit cette fois notre espace d'entrée X comme étant l'ensemble des représentations des images de notre ensemble de données sous forme de vecteurs de descriptions (*Bag of Words*). On définit l'espace de sortie structuré Y comme étant une liste d'ordonnancement des images par rapport à une requête. Pour ce problème de ranking, on considère un étiquetage binaire des données. Dans ce cas on définit la *joint feature map* et la fonction de coût de la façon suivante :

$$\Psi(x, y) = \sum_{i \in \oplus} \sum_{j \in \ominus} y_{ij}(\phi(x_i) - \phi(x_j)) \tag{5}$$

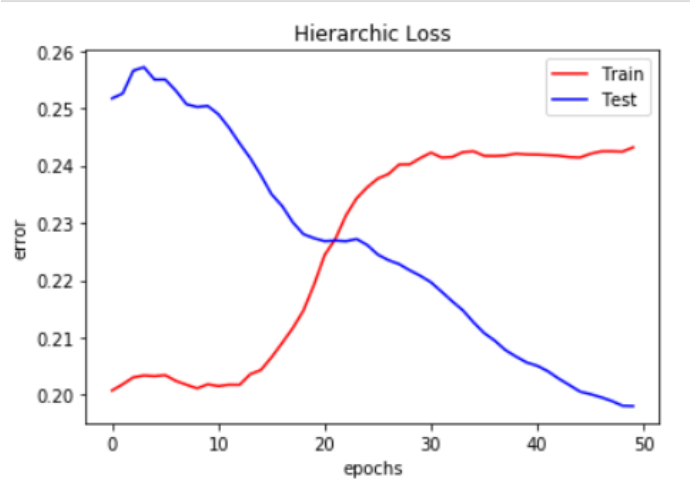


FIGURE 3 – Évaluation de la Hierarchic loss en train et en test sur 100 epochs.

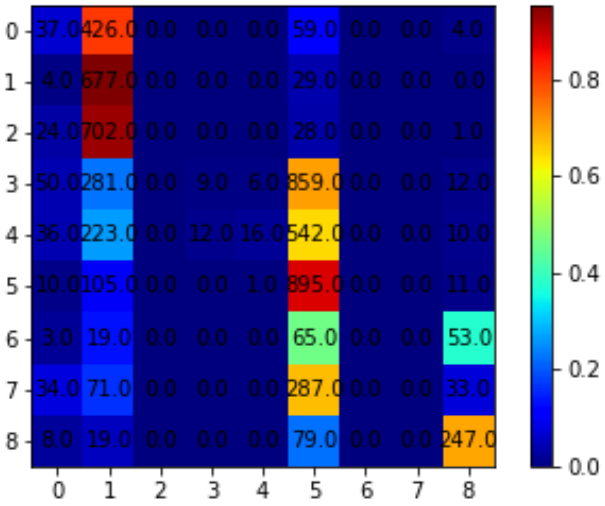


FIGURE 4 – Matrice de confusion pour un modèle entrainé avec une Hierarchic loss sur 100 epochs. 0 = taxi, 1 = ambulance, 2 = minivan, 3 = acoustic guitar, 4 = electric guitar, 5 = harp, 6 = wood frog, 7 = tree frog, 8 = european fire salamander

$$\Delta(y_i, y) = 1 - AP(y) \tag{6}$$

Où

— $\oplus$ représente l'ensemble des images placées avant l'ensemble $\ominus$ dans l'ordonnancement y

— AP représente l'aire sous la courbe de précision/rappel (Précision Moyenne)

3.2 Évaluation et inteprétation des résultats

4 Conclusion