

# Shallow representation - From Shallow to Deep representation for multimedia data - Lecture 7 : Deep and trick video representation

Johnny Nguyen<sup>1</sup>

**Abstract**—This electronic document permits me to synthesis the seventh course of analysis and indexation.

## I. INTRODUCTION

CNN et Deep Neural Network ont été vu avec le back-propagation pour minimiser l'erreur. En commençant par une architecture déjà existante, nous gagnons en rapidité (ImageNet). La couche softmax permet de transformer les scores en probabilité. La sigmoïd peut être utilisée pour la couche en sortie en ayant des valeurs comprises entre 0 et 1. La tangente est plus rapide. Nous utilisons PCA avant de modéliser le problème. Tous les algorithmes sont disponibles sur Keras. L'overfitting correspond à apprendre par coeur au lieu d'être générique. Le but est d'avoir le moins d'overfitting.

## II. LES PARAMÈTRES DU RÉSEAU

Nous devons utiliser les données d'entraînement pour déterminer les paramètres du réseau. Pour les problèmes de classifications, nous utilisons la validation croisée, ceux de régression, la moyenne et ceux de la récupération le triplet ou le similarité du cosinus. La fonction de coût permet d'évaluer si notre modèle est valable.

## III. OPTIMIZATION

La descente de gradient peut être utilisée pour converger vers un minimum. Les mini-batches récupèrent un échantillon du dataset pour en évaluer le coût pour savoir si le modèle est valable. Il faut minimiser puis normaliser. Lorsque le gradient est petit, l'entraînement se fait lentement. Le ReLU (Rectified Linear Unit) permet de répondre au Vanishing problem. Le momentum permet de minimiser correctement l'erreur. La combinaison de RMSProp et Adam sont très bien à utiliser. La régularisation permet de minimiser les poids et d'ajouter de la stabilité. Le dropout permet d'activer ou désactiver certaines neurones pendant un certain temps. La sélection des neurones se fait de manière efficace. En multipliant par la probabilité, nous normalisons.

## IV. VIDÉO

L'actualité d'une vidéo la plus rapide fera le buzz. Nous savons que CNN traite les images. Pour les vidéos, nous ajoutons une dimension correspondant au temps. Le filtre 3D permet de prendre en compte une séquence d'image pendant l'analyse.

<sup>1</sup>This work was not supported by any organization

## V. RMAC VERS RC3D

Après avoir fait la convolution, des propositions sont effectuées et elles sont filtrées. Nous obtenons les meilleurs propositions pour en extraire le début et la fin de la séquence. Nous arrivons à prédire la séquence qui nous intéresse.

## VI. CONCLUSION

Lab session 2 hour

## ACKNOWLEDGMENT

Thanks to Frederic Precioso for his work.

## REFERENCES

- [1] <https://moodle.polytech.unice.fr/course/view.php?id=31>