

Rapport sur le mémoire de thèse de Monsieur Mohamed BOUAZIZ

Mémoire intitulé : *Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles*

Le mémoire de thèse présenté par Mohamed BOUAZIZ s'intitule « *Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles* ». La version relue comporte 154 pages. Le mémoire s'organise en deux parties, pour un total de sept chapitres incluant une introduction (**chapitre 1**). Une annexe est également fournie. De manière classique, les deux parties du mémoire se divisent en une première partie dédiée à l'état de l'art, alors que la seconde partie présente les contributions de l'auteur. L'état de l'art traite principalement de la structuration automatique de contenus télévisuels et des techniques d'apprentissage supervisé utilisées pour le traitement de données séquentielles. Ce mémoire est bien rédigé et agréable à lire.

1. Contexte et objectifs

La thèse de Mohamed BOUAZIZ, dirigée par Georges LINARES (Laboratoire Informatique d'Avignon, Université d'Avignon et des Pays de Vaucluse) et co-encadrée par Richard DUFOUR et Mohamed MORCHID (Laboratoire d'Informatique d'Avignon), s'inscrit dans le cadre d'une collaboration de type CIFRE avec l'entreprise EDD qui propose des services de veille médiatique en lien avec la presse, la télévision et les réseaux sociaux. Dans ce cadre, les travaux de Mohamed BOUAZIZ visent à prédire le genre de la prochaine émission qui sera observée dans le flux d'une chaîne de télévision. Cette prédiction est motivée par l'application en temps réel de traitements automatiques comme la segmentation automatique ou l'extraction de moments forts d'une émission, traitements qui sont facilités par la connaissance *a priori* du genre de l'émission visée.

Pour répondre à cet objectif applicatif, Mohamed BOUAZIZ a proposé plusieurs méthodes permettant cette prédiction, tout particulièrement en utilisant la connaissance de plusieurs flux d'émissions parallèles asynchrones et hétérogènes.

2. Structuration automatique de contenus télévisuels et traitement de données séquentielles par apprentissage automatique supervisé

Le **chapitre 2** du mémoire s'intitule « *Traitement automatique de contenu télévisuel* » mais n'est pas aussi général que son titre semble l'indiquer. Il consiste principalement en un survol de l'état de l'art dans le domaine de la structuration automatique de contenus télévisuels et de la classification en genre d'une émission. Cet aperçu montre les bonnes connaissances du candidat sur ces domaines, mais finalement aucune des techniques présentées ou des traits caractéristiques (indicateurs) détaillés ne seront exploités plus tard dans le manuscrit. L'auteur fait le lien entre ce chapitre et ses contributions en soulignant l'aspect séquentiel des contenus audiovisuels. Avant cela, ce chapitre présente les différentes taxonomies existantes pour déterminer le genre d'une émission.

Le **chapitre 3** présente un état de l'art sur l'apprentissage supervisé pour le traitement de données séquentielles. À ce titre, il est beaucoup plus en phase avec les contributions du candidat, en abordant des notions comme les modèles n-grammes, les réseaux de neurones récurrents ou encore des méthodes de classification non directement liées à la modélisation de séquences comme les *Support Vector Machines (SVM)*.

3. Contributions méthodologiques et expérimentations

La première contribution de l'auteur apparaît au **chapitre 4**, dans lequel une nouvelle taxonomie des genres d'émissions est proposée, justifiée par la nécessité de regroupements en genre moins ambigus. Cette taxonomie est composée de 15 genres, dont seuls 11 seront utilisés dans les expérimentations à venir. La tâche visée est également décrite, ainsi que le corpus de données. Nous y apprenons que les données sont extraites à partir de guides de programmes TV des années 2013, 2014 et 2015, avant d'être projetées automatiquement dans la taxonomie proposée par l'auteur. Il existe ici un véritablement questionnement sur l'intérêt applicatif réel de la tâche de prédiction de genre qui se trouve au cœur de ce travail de thèse. En effet, l'information recherchée est déjà disponible dans les méta-données que constituent les guides de programme : pourquoi ne pas simplement la projeter sur les genres de la taxonomie proposée *via* la même approche que celle présentée en annexe A, même si elle rencontre parfois quelques difficultés ? Et surtout, pourquoi par la suite ne jamais utiliser les genres proposés par les guides de programmes comme indicateurs en plus de l'horaire exact de diffusion de l'émission traitée ?

Alors que les données concernent plusieurs chaînes de télévision (TF1, M6, France 5, TV Monde), seuls les genres des émissions de M6 seront à prédire dans les expérimentations de cette étude : pourquoi ne pas étendre les expérimentations de prédiction de genres des émissions à celles des autres chaînes disponibles dans les données ? L'interprétation des résultats expérimentaux, et surtout leur fiabilité statistique, n'en aurait été que bonifiée sans surcoût significatif. Une autre interrogation est soulevée dans ce chapitre dans la description de la tâche. Il est question de ne considérer dans l'historique que les genres réels, et non pas les genres précédemment prédits : comment cela se justifie-t-il au regard de la tâche applicative qui ne semble pas mettre l'humain dans la boucle pour corriger d'éventuelles erreurs ?

Avec le **chapitre 5**, nous entrons dans le cœur du travail de thèse. Dans ce chapitre deux approches de classification classiques (SVM et perceptron multicouche) sont expérimentées pour la tâche de prédiction décrite dans le chapitre précédent, ainsi que des modèles adaptés aux séquences (n-grammes et réseau de neurones récurrent de type *Long Short Term Memory (LSTM)*). Ces expériences montrent, comme attendu, l'intérêt accru des modèles de séquence sur les approches de classifications classiques pour traiter des séquences. Elles montrent également la supériorité des modèles LSTM sur les modèles n-grammes. Puisque la tâche à réaliser s'apparente à la modélisation d'un langage d'un vocabulaire de 11 mots (les 15 genres de la taxonomie ramenés à 11 sur les données utilisées), ces résultats sont conformes à la littérature. Cependant, une analyse pertinente des erreurs montre les limites de l'approche LSTM qui gère plus mal que les modèles n-gramme les événements peu fréquents. Dans ce même chapitre, une combinaison entre LSTM et SVM (nommé SE-SVM) est expérimentée avec succès. Je pense qu'il aurait été opportun de citer des travaux ayant déjà montré l'intérêt de combiner réseaux de neurones et SVM, par exemple parmi les plus récents : « *Deep learning using linear support vector machines* », Yichuan Tang, ICML 2013. Enfin, le chapitre se termine par l'expérimentation de l'utilisation exclusive de l'historique d'autres chaînes télévisuelles pour conclure de l'intérêt potentiel d'exploiter ces historiques.

Le **chapitre 6** présente les contributions les plus originales de ce mémoire. Elles autorisent l'exploitation de séquences parallèles d'événements asynchrones (les historiques d'émissions de chaînes de télévision différentes). L'une d'elles, nommée PLSTM consiste à considérer chaque séquence indépendamment et à pondérer les sorties de chaque couche récurrente, de manière analogue à celle utilisée dans un réseau bi-LSTM qui utilise deux séquences. La seconde se nomme MSE-SVM et consiste à appliquer un SVM sur un super-vecteur construit à partir de la concaténation des représentations vectorielles de séquences construites à partir des différents LSTM. Les résultats obtenus sont tout à fait remarquables. Je m'interroge simplement sur le fait qu'aucune comparaison avec une interpolation linéaire de modèles n-grammes ne soit proposée. En reprenant l'analogie évoquée plus haut avec la modélisation du langage, il est en effet très courant de construire un modèle de langage composite à partir de modèles construits chacun sur une source différente (ici les flux de chaînes télévisuelles différentes). Quoi qu'il en soit, les résultats obtenus à partir des contributions du candidat sont convaincants.

Le **chapitre 7** clôt le mémoire sur une conclusion, et fournit un ensemble de perspectives au travail présenté.

4. Appréciation

Tant sur la forme que sur le fond, le mémoire de thèse de M. Mohamed BOUAZIZ est très intéressant et agréable à lire. Les différentes remarques et questions que j'ai pu écrire dans ce rapport sont liées à l'intérêt que je porte à ce travail que j'estime de qualité. Je suis convaincu que les contributions du candidat ont une portée qui dépasse la tâche de prédiction de genres d'émissions telle que définie dans ce travail, réalisé dans le cadre d'une thèse sous convention CIFRE.

L'ensemble de l'étude présentée dans ce mémoire a été valorisé par des publications scientifiques internationales qui valident la qualité et le sérieux de ses travaux de recherche : **pour toutes ces raisons, j'émet un avis favorable à la soutenance de thèse de Mohamed BOUAZIZ en vue de l'obtention du titre de docteur.**

Le 22 novembre 2017
Yannick Estève
Professeur en informatique
Le Mans Université

