

DOCUMENT DE SUIVI INDIVIDUEL (CSI) DU DOCTORANT DE L'ÉCOLE DOCTORALE MATHSTIC

Angers ☐ **Brest** ☐ **Le Mans** ☐ **Lorient-Vannes** ☐ **Nantes** ☐ **Rennes** ☐

PARTIE RÉSERVÉE AU DOCTORANT

Cette première partie est à remplir par le doctorant et le document de CSI est à transmettre aux membres de son comité - CSI

LE DOCTORANT

NOM et Prénoms du(de la) doctorant(e) : WANG Yichang		
Établissement d'inscription : Univ Rennes 1		
Date de début de la thèse : 16 Avril 2018		
Unité de recherche : irisa / inria - rba		
Nature du financement : bourse CSC (China)		Durée ou date de fin du financement : 3 ans
S'agit-il d'une thèse à temps partiel ?	Non	Indiquez la quotité en % (le cas échéant) :
S'agit-il d'une thèse CIFRE ?	Non	Nom de l'entreprise (le cas échéant) :
S'agit-il d'une thèse en cotutelle ?	Non	Établissement de cotutelle (le cas échéant) :
Titre de la thèse :		Interpretable shapelets for anomaly detection in time series

Cochez la case correspondant à votre domaine et spécialité d'inscription :

AST			EGE		INFO	MI
Automatique, Productique et Robotique	Signal, Image, Vision	Télécommunications	Électronique	Génie Électrique	Informatique	Mathématiques et leurs Interactions
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

LA DIRECTION ET L'ENCADREMENT DE LA THÈSE

	Nom et Prénoms	Grade/Fonction	Établissement/Laboratoire	E-mail
Directeur de thèse	Elisa Fromont	Professeur	Univ Rennes 1 / IRISA	elisa.fromont@irisa.fr
Co-directeur (*)				
Co-encadrant (*)	Romain Tavenard	MdC	Univ Rennes 2 / LETG	romain.tavenard@univ-rennes2.fr
Autre co-encadrant(*)	Simon Malinowski	MdC	Univ Rennes 1 / irisa	simon.malinowski@irisa.fr
	Rémi Emonet	MdC	Université de Saint-Etienne	remi.emonet@univ-st-etienne.fr

LE COMITÉ DE SUIVI INDIVIDUEL DU DOCTORANT

	Nom et Prénoms	Grade/Fonction	Établissement/Laboratoire	E-mail
Le référent CSI (**)	Guillaume Gravier	DR CNRS	IRISA	guillaume.gravier@irisa.fr
Les membres CSI	Vincent Lemaire	Senior researcher, HDR	Orange Labs	vincent.lemaire@orange.com

--	--	--	--	--

(*) Pour les règles de direction et d'encadrement de thèse dans l'école doctorale MathSTIC, voir l'article 11 du règlement intérieur <https://ed-mathstic.u-bretagne-loire.fr/sites/default/files/u50/mathstic-reglement-interieur-2018.01.29.pdf>

(**) Le référent-CSI est le membre du comité-CSI en charge de piloter les entretiens CSI, de rédiger le compte-rendu et de donner l'avis sur la réinscription, en concertation avec les autres membres du CSI.

RAPPORT ANNUEL D'AVANCEMENT DES TRAVAUX DEMANDE DE RÉINSCRIPTION EN THÈSE

Les 2 pages suivantes sont à remplir par le doctorant. Le document CSI est à transmettre ensuite aux membres de son comité CSI

<input type="checkbox"/> CSI 1 (Demande d'inscription en 2 ^e année)	<input checked="" type="checkbox"/> CSI 2 (Demande d'inscription en 3 ^e année)	<input type="checkbox"/> CSI 3 et + (Demande de dérogation pour une inscription en 4 ^{ème} année ou plus)
---	--	---

Récapitulatif des formations scientifiques/disciplinaires suivies depuis le début de la thèse

Dates de la formation	Titre de la formation	Nb d'heures équivalent	Validées (Oui, En cours)
19 et 20 avril 2018	AI days	9	Oui
22 novembre 2018	Python intermédiaire	3	Oui
25 et 26 mars 2019	TS days	9	Oui
Du 1 au 5 juillet 2019	PFIA2019 conférence	30	Oui
Du 15 au 19 juillet 2019	Deep learning summer school	30	Oui
Nombre total d'heures de formations scientifiques réalisées :		81	

Récapitulatif des formations transversales suivies depuis le début de la thèse

Dates de la formation	Titre de la formation	Nb d'heures équivalent	Validées (Oui, En cours)
15 novembre 2018	Sensibilisation à l'entrepreneuriat	4	oui
13 et 29 avril 2019	Ethique de la recherche et intégrité scientifique	6	oui
2ème semestre de l'année scolaire 2019-2020	Enseignant TP : Introduction à OOP en Python	42	Oui
Nombre total d'heures de formations transversales réalisées :		52	

Activités (autres que formations) réalisées

Description courte de l'activité (mobilité, etc.)	Dates de l'activité
Séminaire à Equipe Obelix, Vannes	18 mars 2019
AALTD conference	18 septembre 2020

Rapport sur l'état d'avancement des travaux de thèse et projet pour l'année suivante

(Mettre ici en **1 page maximum** une synthèse des travaux réalisés au cours de l'année, les résultats obtenus, les articles soumis/publicés, etc. ainsi que les projets pour l'année suivante et le planning prévisionnel des travaux)

In the second year of my PhD study, I focused on the interpretation of the shapelet-based time series classification model I developed in the first year. My work has been completely interrupted from April 1st until August 31st due to the COVID-19 situation (I was in China with my family).

To summarize the work of my first year: it was inspired by Learning Shapelet (LS) method [1], where the authors obtained good performance by jointly learning the shapelet representation and training the time series classifier. This work ([1]) is, in fact, quite similar to learning with a simple (few layers) Convolutional Neural Network (CNN), because both the shapelets and the convolutional kernels can be considered as feature extractors. However, we noticed that neither the shapelets produced by [1] nor the convolutional filters of CNN look like real pieces of time series which make them quite different from the original "shapelet" definition but also not an interpretable vocabulary to explain the classifier's decisions. To benefit from the interpretable aspect of "real" shapelets, we proposed a new adversarial regularization on the convolutional filters of our base CNN classifier.

Our method is as accurate as LS [1] but we learn more interpretable shapelets on the 85 datasets of the UCR/UEA archive [2]. This work was submitted to KDD 2019 (with quite good reviews), CIKM 2019, DAMI 2020, and ECMLPKDD 2020. In 2019, the idea was deemed very good but we were missing good illustrations about the "explainable" aspect of our method and we spent most of our time this year, trying to improve this point (see below). In 2020, a serious concurrent paper was published (at AAAI2020 [4]) which made publishing this work, almost impossible.

To improve the "explainable" aspect of our method, we adapted Grad-CAM [3], a famous method to explain the predictions of neural networks, to our time series context. In addition to a *post-hoc* explanation on a single test instance which would highlight which regions in the instance is useful for a particular prediction (as [3] does), we also 1) consider the negative shapelets for one prediction result, and 2) provide a global level shapelet-based explanations for entire classes.

Another similar work [4] was carried out in parallel as our work. Their network architecture is quite similar to ours but their vision of what is actually a shapelet in the network is different: in their case, the shapelets are the feature maps resulting from the convolution layers so their (also adversarial) regularization is made on the activation values of the feature maps and not on the convolutional filters. They obtain more accurate classification results with interpretable "shapelets" but they do not focus much on how to use these shapelets for providing explanations. Our work was finally accepted in AALTD 2020 (workshop of ECMLPKDD 2020) and ICTAI 2020.

I also started to work on the review of state-of-the-art methods for explaining machine learning methods and time series classifiers in particular. I wrote a chapter with the following sections: why interpretable AI matters, black-box vs white-box models. This chapter is not finished yet and I now have to focus on detailing the existing methods. In addition to writing my thesis, the next year will be dedicated to improving the usability of my method: for example, I will try to add new regularization terms (e.g. group-lasso) to automatically find the minimum number of (interpretable) shapelets needed for classification. I would also like to test our method on different use-cases (for example multivariate time series or images instead of univariate time series) to evaluate its generalizability. I would also like to work on different types of neural network classifiers, e.g. attention-based ones (such as transformers) and study how to explain the prediction results in this context.

[1] Grabocka et al., Learning time-series shapelets, KDD 2014: 392-401

[2] Bagnall et al., The UEA & UCR Time Series Classification Repository, www.timeseriesclassification.com

[3] Selvaraju et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, ICCV 2017: 618-626

[4] Ma et al., Adversarial Dynamic Shapelet Network, AAAI 2020 : 5069-5076

Autres informations à destination des membres du comité de suivi individuel

Mettre ci-dessous (ou sur papier libre) les observations/remarques, informations complémentaires et souhaits éventuels de tout ordre que vous souhaitez communiquer aux membres de votre comité de suivi individuel. Ceci peut également être fait lors d'un entretien avec votre comité CSI que vous pouvez solliciter à tout moment.

RAS

Date et signature du doctorant :

15 octobre 2020

王逸暢

COMPTE-RENDU DU COMITÉ DE SUIVI INDIVIDUEL DU DOCTORANT

Cette partie incluant l'avis du CSI pour la réinscription est à remplir par le référent-CSI (en concertation avec les membres du CSI)
Le document CSI est à transmettre ensuite au doctorant et au directeur de thèse.

Il s'agit ci-dessous d'une trame dont le comité CSI peut se servir en cochant les cases appropriées. Le compte-rendu peut toutefois se faire sous format libre, où devront cependant apparaître les points ci-dessous.		
L'entretien avec le doctorant sans les encadrants est <u>obligatoire pour une réinscription en 3^{ème} année ou plus</u> . Dans les autres cas, le conseil de l'ED MatSTIC recommande un entretien avec le doctorant pour discuter du déroulement de la thèse. <u>Cet entretien a-t-il eu lieu ?</u>	X OUI Date de l'entretien : 19/10/2020	<input type="checkbox"/> NON Autre formule :
Le doctorant a-t-il fait une présentation orale de ses activités au comité ?	X OUI	<input type="checkbox"/> NON Autre formule :
Le doctorant s'est-il approprié le sujet de thèse et fait-il preuve d'autonomie ?	X OUI	<input type="checkbox"/> NON
Le doctorant a-t-il présenté ses résultats dans une conférence/revue dans l'année écoulée ?	X OUI	<input type="checkbox"/> NON
Le doctorant est-il dans une dynamique de publication de ses résultats ? *	X OUI	<input type="checkbox"/> NON
Le doctorant a-t-il présenté un planning prévisionnel de ses activités pour l'année suivante ?	X OUI	<input type="checkbox"/> NON
Ce planning prévisionnel vous paraît-il crédible/réalisable ?	X OUI	<input type="checkbox"/> NON
Le doctorant a-t-il suivi des formations scientifiques/disciplinaires pendant l'année écoulée ?	<input type="checkbox"/> OUI	X NON
Si non, dites pourquoi (dispense, etc.) : 80h de formation atteinte à la date de juillet 2019		
Le doctorant a-t-il suivi des formations transversales/professionnelles pendant l'année écoulée ?	X OUI	<input type="checkbox"/> NON
Si non, dites pourquoi (Salarié, cotutelle, CIFRE, etc.) :		
Le doctorant a-t-il effectué des enseignements durant l'année écoulée ? ^{1*}	X OUI nb d'heures = 42	<input type="checkbox"/> NON

Rapport de synthèse du comité CSI

État d'avancement de la thèse : En raison de la crise sanitaire et de la difficulté de publication dans des conférences de bons niveaux dans le domaine de l'IA, les travaux de thèse n'ont pas progressé autant qu'espéré. La thèse a cependant avancé convenablement, avec notamment la publication d'un article à ICTAI 2020 qui pourra servir de cœur à la rédaction d'un manuscrit. Cette première contribution notable pose des bases pour de nouvelles contributions dans les prochains mois afin d'étoffer le futur manuscrit de thèse et de renforcer l'impact des travaux menés jusqu'à présent.

Interrogations et points de vigilance : Le CSID invite Yichang Wang et l'équipe d'encadrement à se concentrer sur les extensions du travail accompli et sur sa valorisation. Il convient de rester vigilant sur la rédaction du manuscrit de thèse dans un délai assez contraint (soutenance envisagée à l'été 2021).

Recommandations (incluant les jalons pour la fin de thèse et la date de soutenance pour une demande dérogatoire d'inscription en 4^{ème} année ou plus) : Le CSID recommande de se concentrer en premier lieu sur la rédaction des premiers chapitres du manuscrit. Il conviendra notamment de développer un chapitre de positionnement de la problématique général afin d'inscrire les contributions dans un cadre scientifique bien défini (définir interprétabilité, explicabilité, le lien entre les deux, mettre en avant et justifier les hypothèses sous-jacentes effectuées, etc.). Il est important que le manuscrit de thèse explicite la démarche scientifique suivie, ce qui n'était pas le cas du plan proposé lors de l'entretien. Parmi les nombreuses pistes de recherche proposées, nous recommandons de se concentrer sur des extensions du modèle proposé dans ICTAI 2020, en termes de technique (group lasso, réseau profond, combinaison signal/feature maps, etc.) et de publication : il serait peu judicieux de changer d'angle d'attaque à ce stade.

¹

* Le comité prendra soin d'attirer l'attention du doctorant sur le fait que ces critères sont importants pour la qualification CNU, si le doctorant souhaite présenter sa candidature à un poste de maître de conférences

--

<p>Le comité CSI a-t-il noté des difficultés, de quel ordre que ce soit (scientifique, personnel, relation avec ses encadrants, son équipe, etc.) préjudiciables à la progression de la thèse ?</p>	<p><input checked="" type="checkbox"/> OUI Merci d'en informer dès à présent le directeur-adjoint de site correspondant</p>	<p><input type="checkbox"/> NON</p>
--	--	-------------------------------------

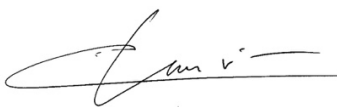

LES AVIS POUR LA RÉINSCRIPTION EN THÈSE

Avis du comité CSI pour la réinscription en thèse

☒ Favorable ☐ Réservé ☐ Défavorable
 (Justification obligatoire pour une demande dérogatoire d'inscription en 4^{ème} année ou plus)

.....

.....

	Référent-CSI	Membres du CSI	
NOM & Prénom	Guillaume Gravier	Vincent Lemaire	
Date et Signature	 20/10/2020	 20/10/2020	

PARTIE RÉSERVÉE AU DIRECTEUR DE THÈSE

A remplir par le directeur de thèse (en concertation avec les co-directeurs et co-encadrants). Le document CSI est à transmettre à la direction du laboratoire.

Avis sur le déroulement de la thèse :

Problèmes éventuels rencontrés :

Avis du directeur de thèse pour la réinscription en thèse

☐ Favorable ☐ Réservé ☐ Défavorable
 (Une justification circonstanciée est exigée pour une demande dérogatoire d'inscription en 4^{ème} année ou plus et/ou en cas d'avis autre que favorable)

.....

Nom et Prénoms du directeur de thèse :	Date et Signature :
--	---------------------

PARTIE RÉSERVÉE AU DIRECTEUR DE LABORATOIRE

A remplir par le directeur de laboratoire ou son représentant légal. Le document CSI est ensuite à transmettre à la direction de l'école doctorale.

Avis du directeur de laboratoire pour la réinscription en thèse

☐ Favorable ☐ Réservé ☐ Défavorable
 (Une justification ainsi que la signature du directeur de laboratoire sont exigées en cas d'avis autre que favorable)

.....

Nom et Prénoms :	Date et Signature :
------------------	---------------------

PARTIE RÉSERVÉE À LA DIRECTION DE L'ÉCOLE DOCTORALE MathSTIC

A remplir par le directeur-adjoint de site de l'école doctorale.

Avis du directeur-adjoint de site pour la réinscription en thèse

☐ Favorable

☐ Réservé

☐ Défavorable

(Une justification ainsi que la signature du directeur de l'école doctorale sont exigées en cas d'avis défavorable)

.....

Nom et Prénoms :	Date et Signature :
------------------	---------------------