imagia

Applications: Séries temporelles

Francis Dutil M.Sc. (francis.dutil@imagia.com)

1er décembre 2018

Plan

1. Introduction aux réseaux de neurones récurrents
2. Données médicales
3. Diverses applications

Plan

1. **Introduction aux réseaux de neurones récurrents**
2. Données médicales
3. Diverses applications

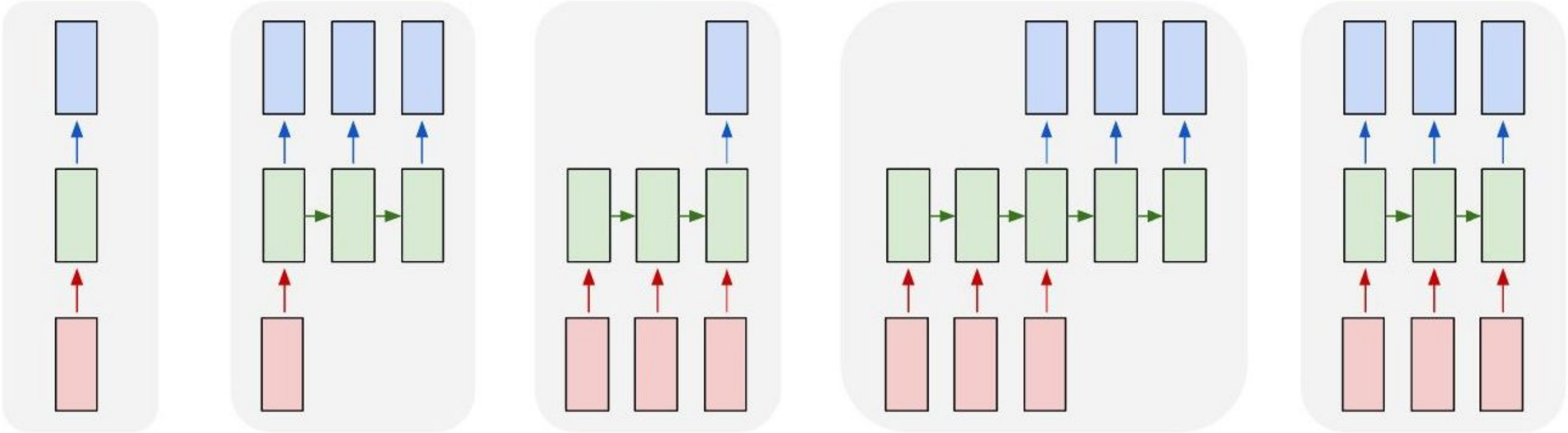
Rappel: type de tâches avec les séries temporelles

un pour un | un pour plusieurs | plusieurs pour un | plusieurs pour plusieurs | plusieurs pour plusieurs

Sortie

État

Entrée



Chat

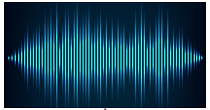
“Ceci est un chat”

“C’est poilu et j’y suis allergique”

Chat

“Ceci est un chat”

“This is a cat”

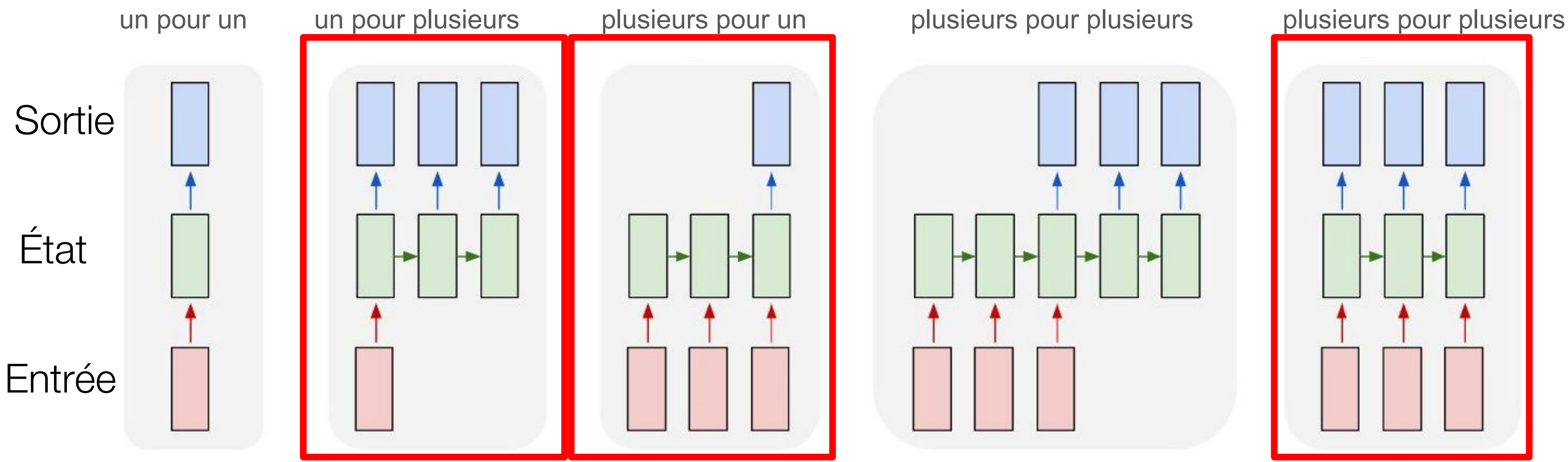


“Miaou Miaou”

Tiré de http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf

Rappel: type de tâches avec les séries temporelles

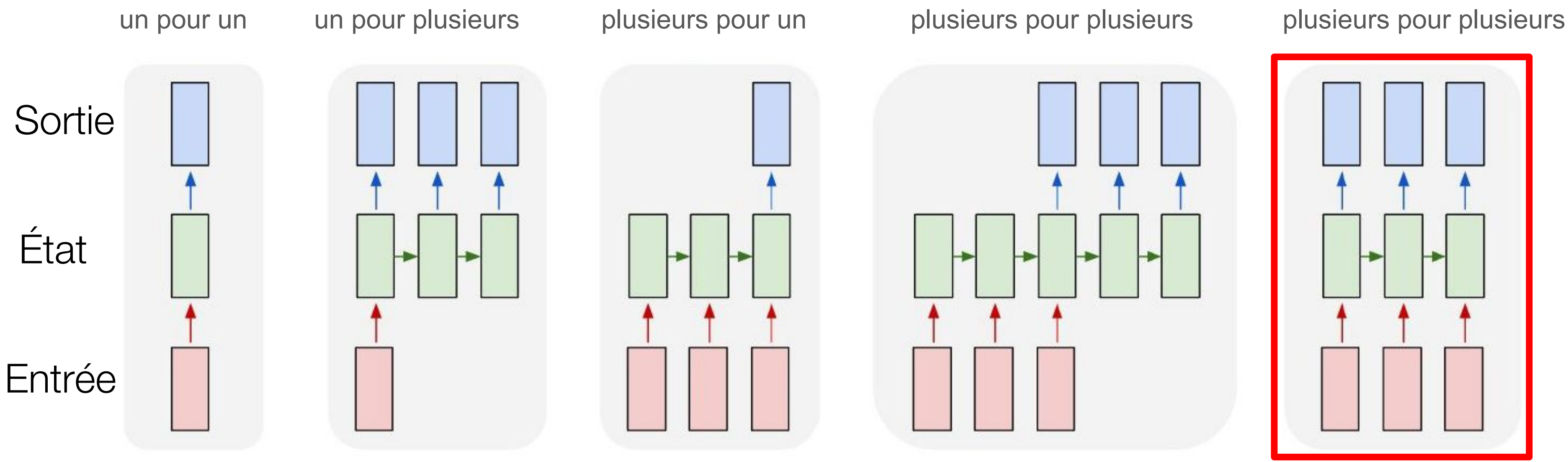
Dans cette présentation



Tiré de http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf

Rappel: type de tâches avec les séries temporelles

Dans cette présentation

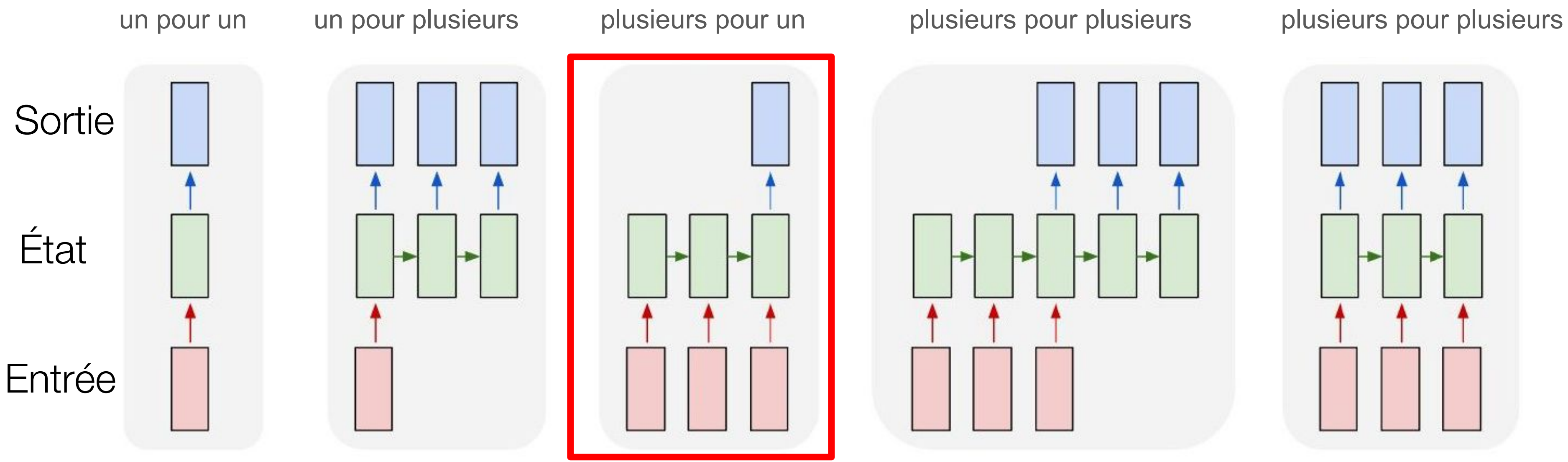


Suivi de patient

Tiré de http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf

Rappel: type de tâches avec les séries temporelles

Dans cette présentation

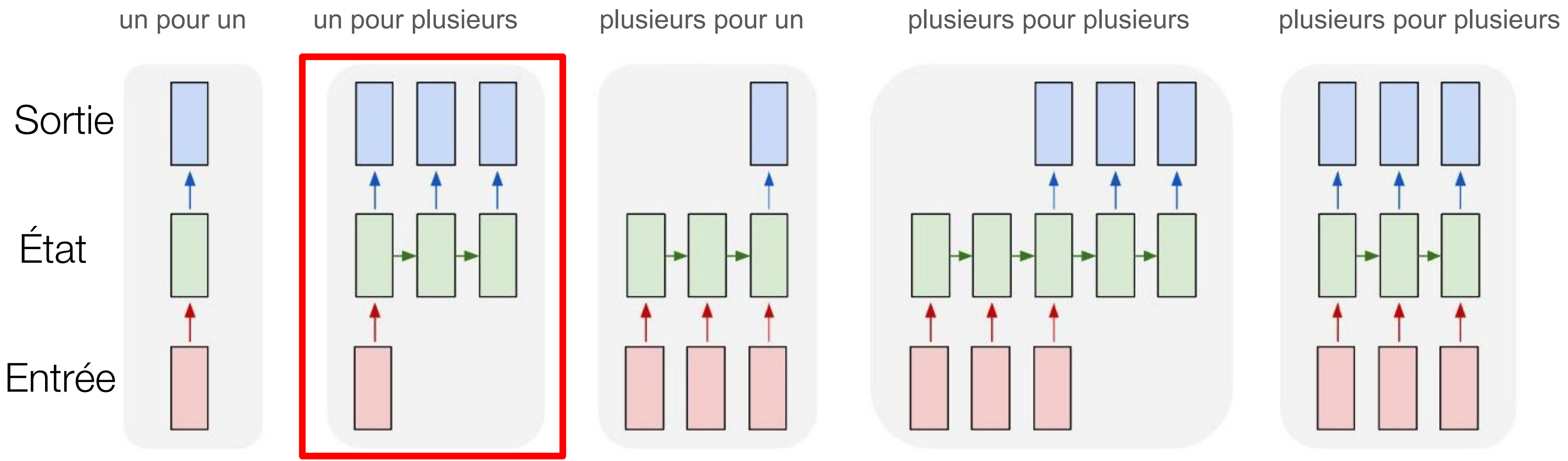


Diagnostic d'un épisode

Tiré de http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf

Rappel: type de tâches avec les séries temporelles

Dans cette présentation



Génération de note clinique
à partir d'image

Tiré de http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf

Réseaux de neurones récurrents

Un RNN applique une fonction à une **séquence d'entrée** $[x_1, x_2, \ldots, x_T]$, pour produire une **séquence de sortie** $[y_1, y_2, \ldots, y_T]$, en maintenant un **état interne** $[h_1, h_2, \ldots, h_T]$.

Séquence de sortie

Etat interne

Séquence d'entrée

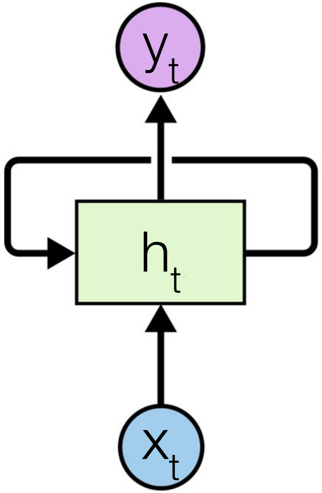


Image du blog de Christopher Olah, diapo de l'École d'automne 2018, César Laurent

Réseaux de neurones récurrents

- La version la plus simple:

$$h_t = \tanh(Ux_t + Wh_{t-1})$$
$$y_t = f(Vh_t)$$

- W, U et V sont les paramètres du réseau.

  ➢ Ils sont **partagés** à travers le temps.

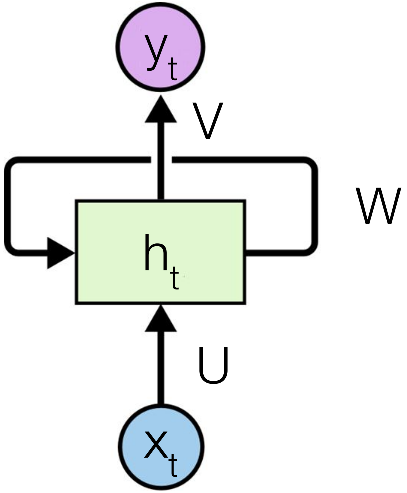
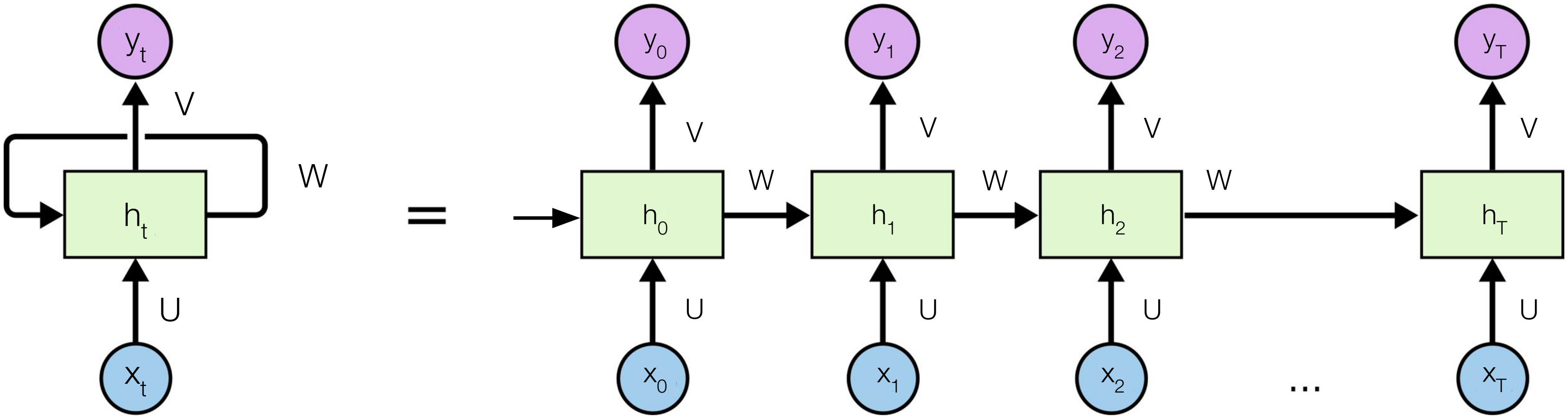


Image du blog de Christopher Olah, diapo de l'École d'automne 2018, César Laurent

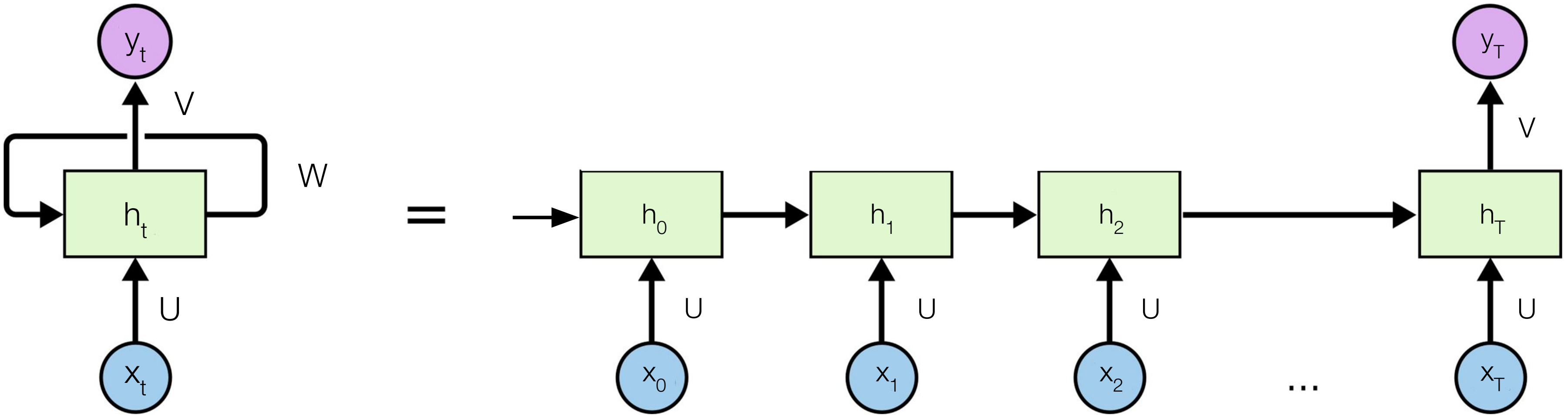
Déroulement dans le temps



Les paramètres sont **partagés** à travers le temps!

Image du blogue de Christopher Olah, diapo de l'École d'automne 2018, César Laurent

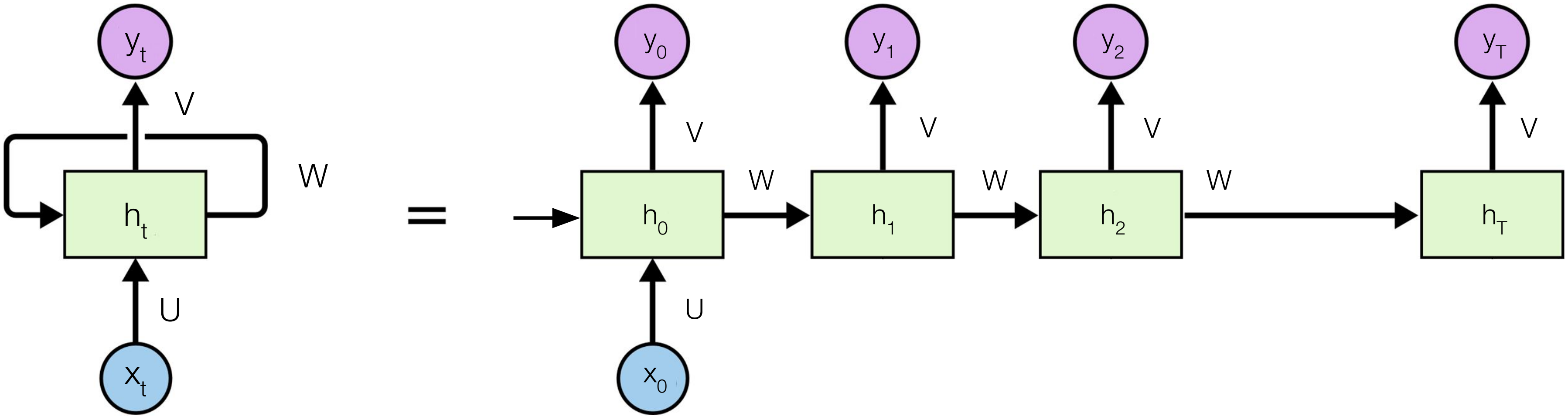
Déroulement dans le temps



Les paramètres sont **partagés** à travers le temps!

Image du blogue de Christopher Olah

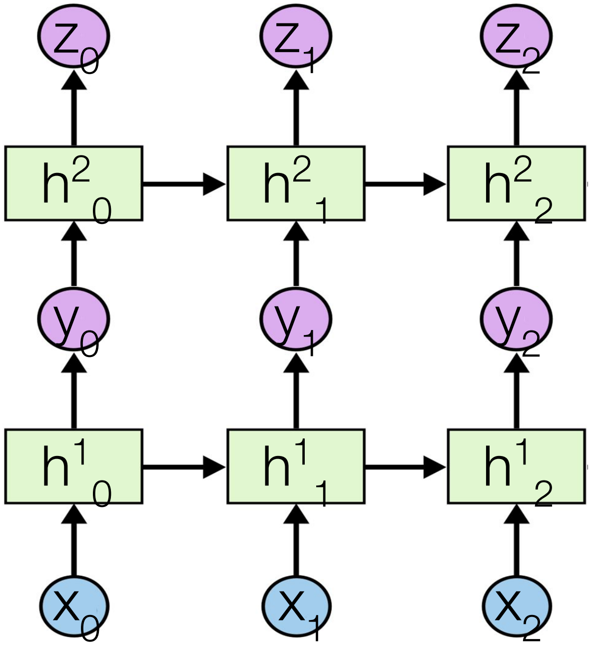
Déroulement dans le temps



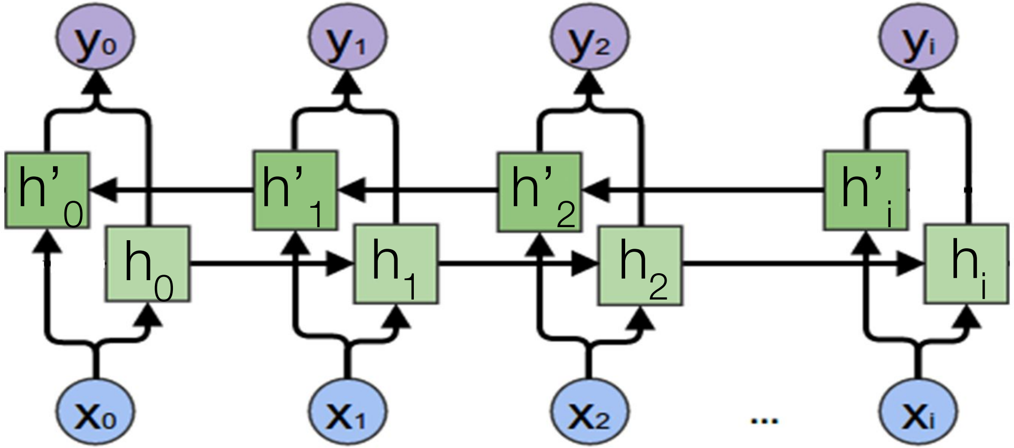
Les paramètres sont **partagés** à travers le temps!

Image du blogue de Christopher Olah

On peut être créatif !



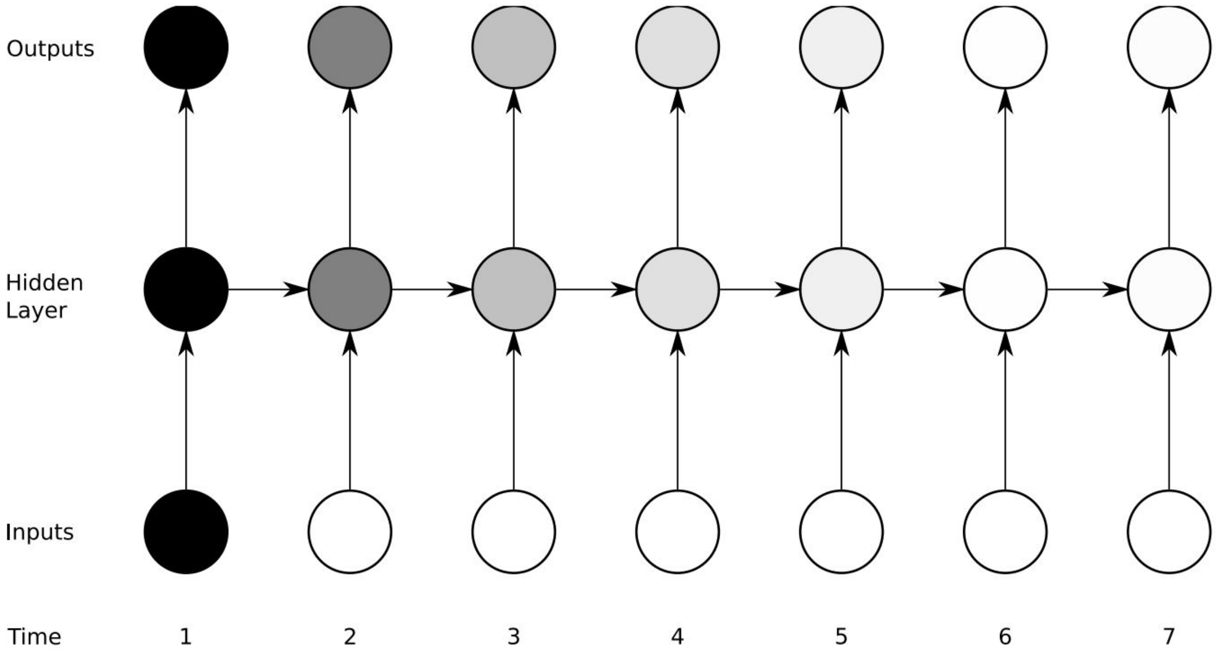
Pile de RNNs



RNN Bidirectionnel

diapo de l'École d'automne 2018, César Laurent

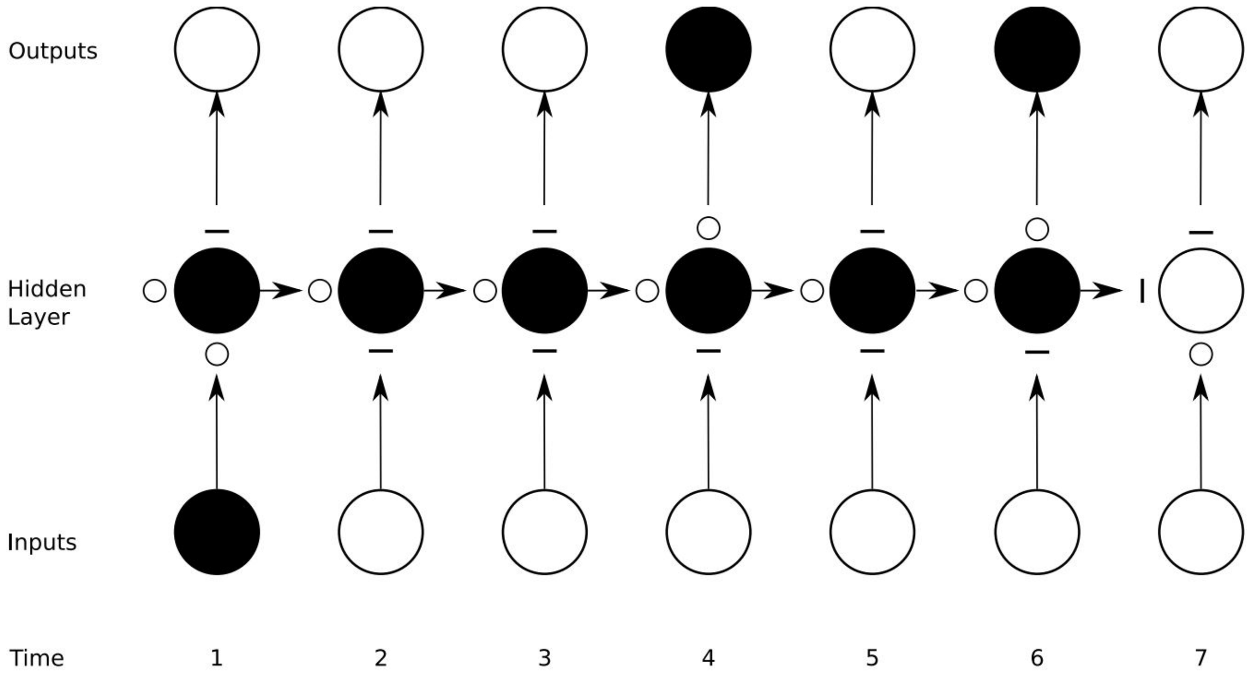
Problème avec le RNN basic



La nuance de gris montre l'influence de l'entrée du RNN au temps 1. Elle décroît au cours du temps, comme le RNN oublie peu à peu sa première entrée.

diapo de l'École d'automne 2018, César Laurent

Problème avec le RNN basic

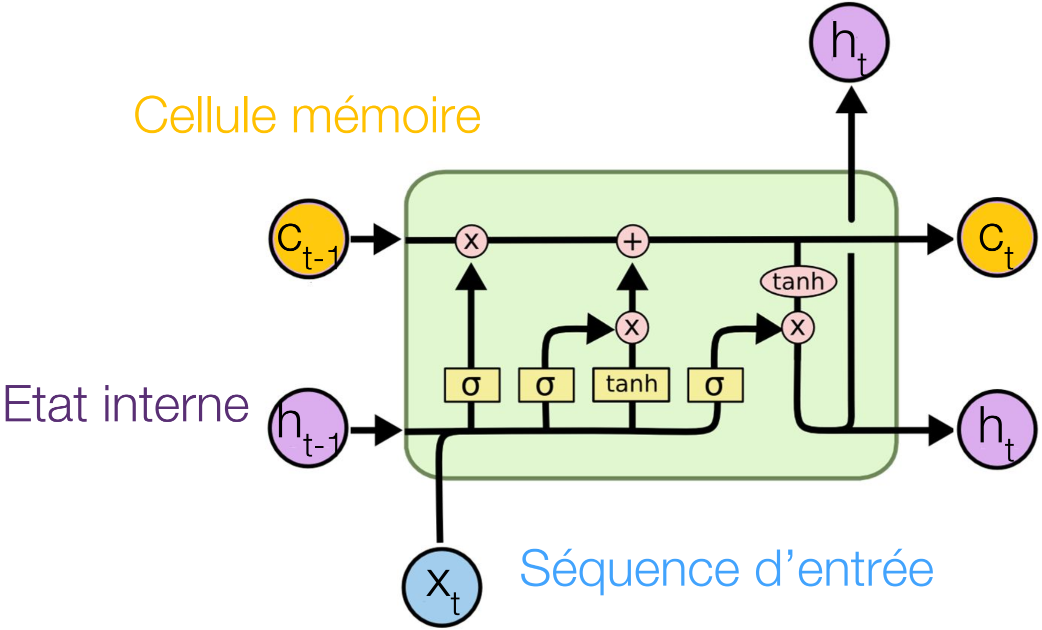


En ajoutant 3 gates (o ouvert; - fermé), qui contrôlent l'entrée, la sortie et l'effacement de l'état, on peut retenir et propager l'information dans le temps.

diapo de l'École d'automne 2018, César Laurent

Long Short-Term Memory (LSTM)

Réduction du problème de dissipation avec **un mécanisme de gates** et une **cellule mémoire**.

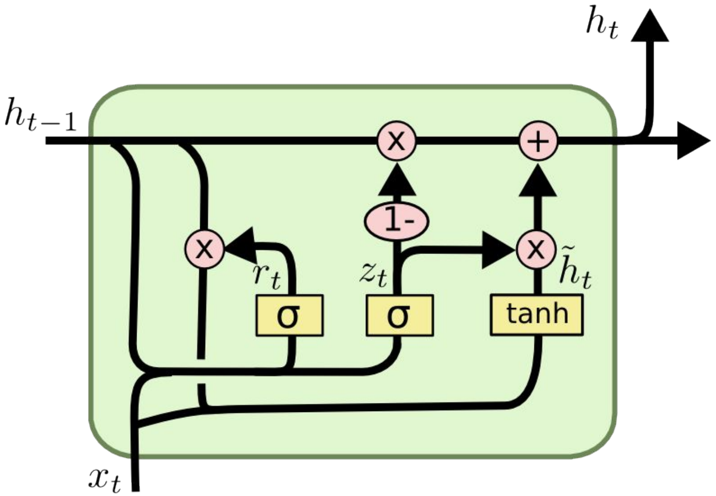


$$
\begin{aligned}
i_t &= \sigma(U_i x_t + W_i h_{t-1} + b_i) \\
f_t &= \sigma(U_f x_t + W_f h_{t-1} + b_f) \\
o_t &= \sigma(U_o x_t + W_o h_{t-1} + b_o) \\
g_t &= \tanh(U_g x_t + W_g h_{t-1} + b_g) \\
c_t &= i_t \odot g_t + f_t \odot c_{t-1} \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

Image du blogue de Christopher Olah
diapo de l'École d'automne 2018, César Laurent
Hochreiter et al., *Long short-term memory*, Neural Computation 1997

Gated Recurrent Unit (GRU)

- Une variante populaire de la LSTM.
  - ➢ Pas de cellule mémoire explicite.
  - ➢ Input et forget gates combinées.
- En pratique, performances égales à la LSTM.
  - ➢ Plus rapide à calculer.



$$z_t = \sigma(U_z x_t + W_z h_{t-1} + b_z)$$
$$r_t = \sigma(U_r x_t + W_r h_{t-1} + b_r)$$
$$g_t = \tanh(U_g x_t + W_g(r_t \odot h_{t-1}) + b_g)$$
$$h_t = z_t \odot g_t + (1 - z_t) \odot h_{t-1}$$

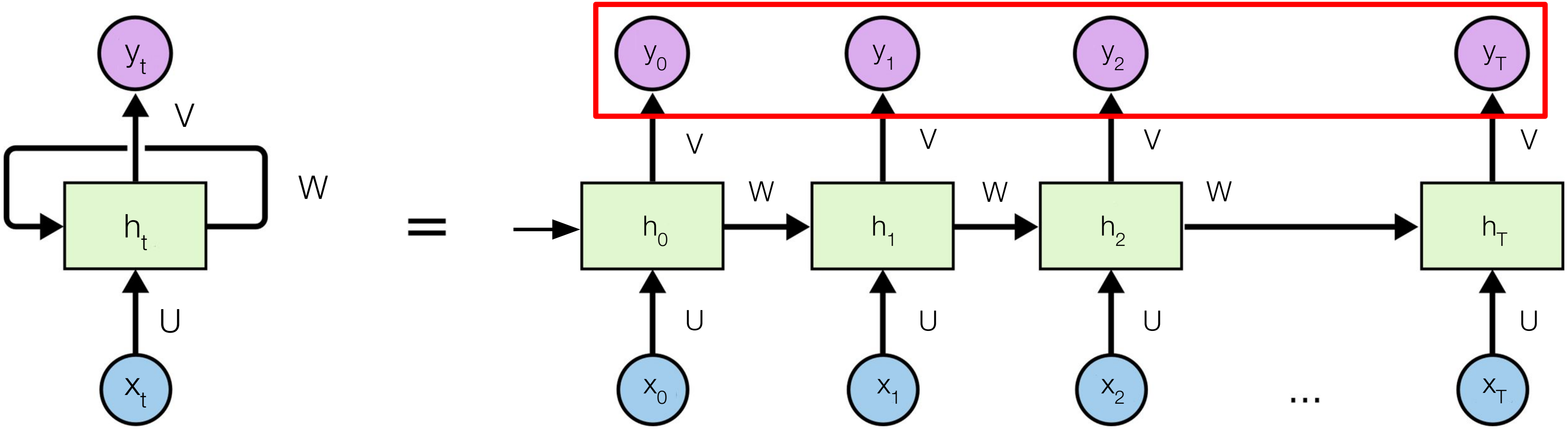
Image du blogue de Christopher Olah
diapo de l'École d'automne 2018, César Laurent
Chung et al.: *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*

Plan

1. Introduction au réseaux de neurones récurrents
2. **Données médicales**
3. Diverses applications

Séquence de sortie

- Prédire la probabilité d'un ***événement*** futur
- On doit s'entendre sur une définition!



CIM (Classification Internationale des Maladies)

World Health Organization

La Classification internationale des maladies (CIM) est une classification médicale codifiée classifiant les maladies et une très vaste variété de signes, symptômes, lésions traumatiques, empoisonnements, circonstances sociales et causes externes de blessures ou de maladies. Elle est publiée par l'Organisation mondiale de la santé (OMS) et est mondialement utilisée pour l'enregistrement des causes de morbidité et de mortalité touchant le domaine de la médecine.

1975 - CIM-9 - (OMS)

1990 - CIM-10 - (OMS)

2019 - CIM-11 - (OMS)

“Bientôt!”

CIM-9 (Classification Internationale des Maladies)

World Health Organization

Exemples de codes CIM-9:

786	Symptoms involving respiratory system and other chest symptoms
786.0	Dyspnea and respiratory abnormalities
786.1	Stridor
786.2	Cough
786.3	Hemoptysis
786.4	Abnormal sputum
786.5	**Chest pain**
786.6	Swelling, mass or lump in chest
786.7	Abnormal chest sounds
786.8	Hiccough
786.9	Other

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD9-CM/2011

CIM-9 (Classification Internationale des Maladies)

World Health Organization

Exemples de code CIM-9 (786.5, ***Chest pain***)

Cardialgia (see also Pain, precordial) 786.51
Diaphragmalgia 786.52
chest 786.59
anginoid (see also Pain, precordial) 786.51
chest (central) 786.50
atypical 786.59
midsternal 786.51
musculoskeletal 786.59
noncardiac 786.59
substernal 786.51
wall (anterior) 786.52
costochondral 786.52
diaphragm 786.52
heart (see also Pain, precordial) 786.51
intercostal 786.59
over heart (see also Pain, precordial) 786.51
pericardial (see also Pain, precordial) 786.51
pleura, pleural, pleuritic 786.52
precordial (region) 786.51
respiration 786.52
retrosternal 786.51
rib 786.50
substernal 786.51
respiration 786.52
Pleuralgia 786.52
Pleurodynia 786.52
Precordial pain 786.51
chest 786.59
Prinzmetal-Massumi syndrome (anterior chest wall) 786.52
painful 786.52

Beaucoup de regroupement!

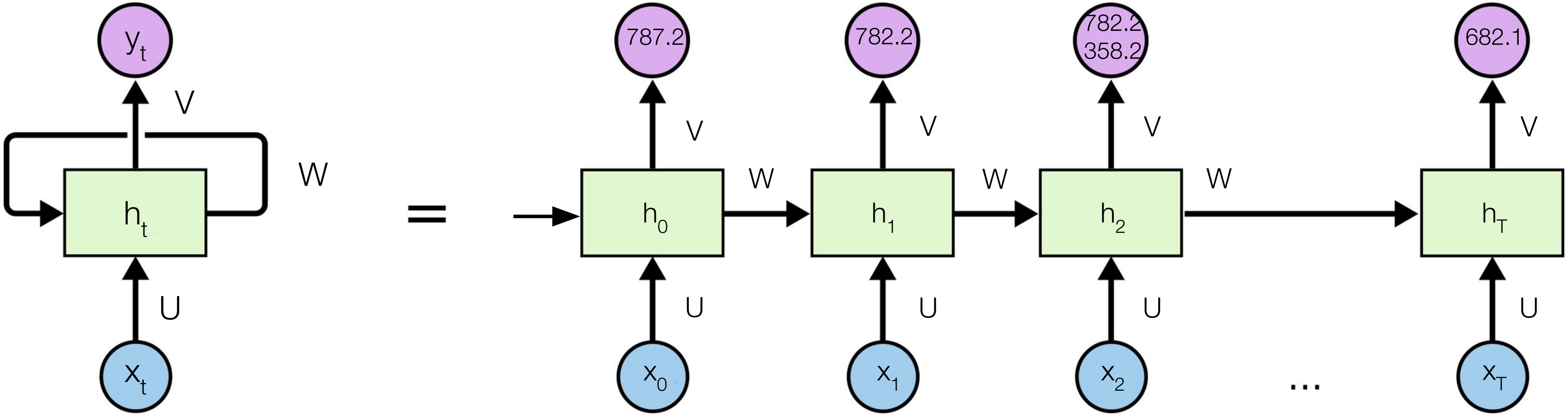
CIM-10 est très détaillé

V97.33XD: **Sucked into jet engine**, subsequent encounter

V00.15: **Heelies Accident**
Applicable To Rolling shoe, Wheeled shoe, Wheelies accident.

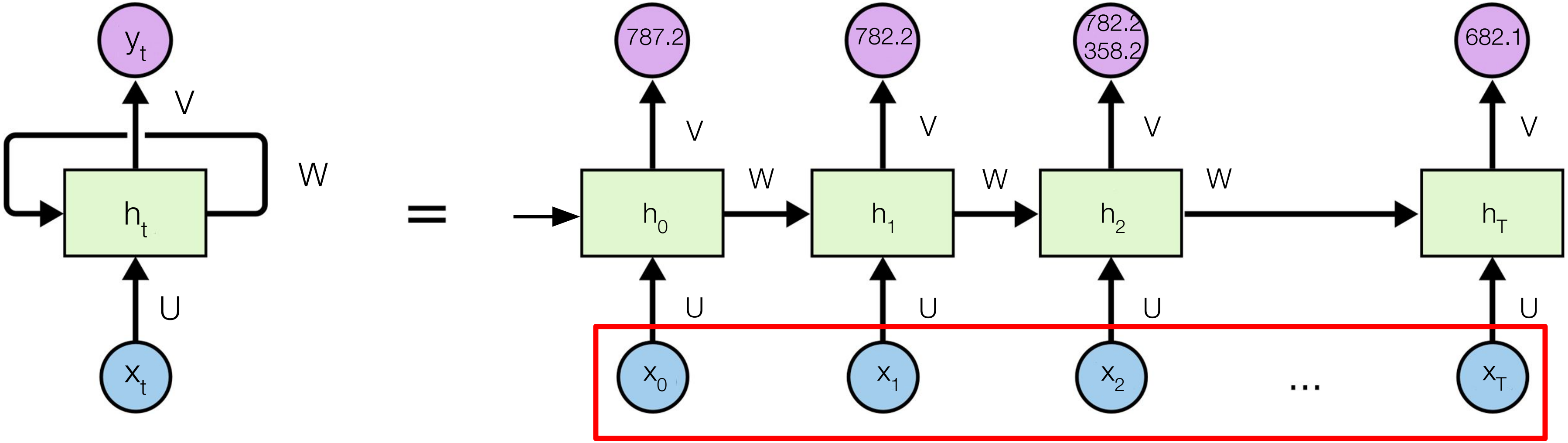
Supertypes:
V00-Y99 External causes of morbidity
V95-V97 Air and space transport accidents
V00 Pedestrian conveyance accident
V00.1 Rolling-type pedestrian conveyance accident

Séquences de sortie



Série de codes médicaux à travers le temps

Et les entrées?



Et les entrées?

- Information demographique
- Signes vitaux
  - Fréquence cardiaque, etc.
- Notes cliniques
- Médications
- etc.

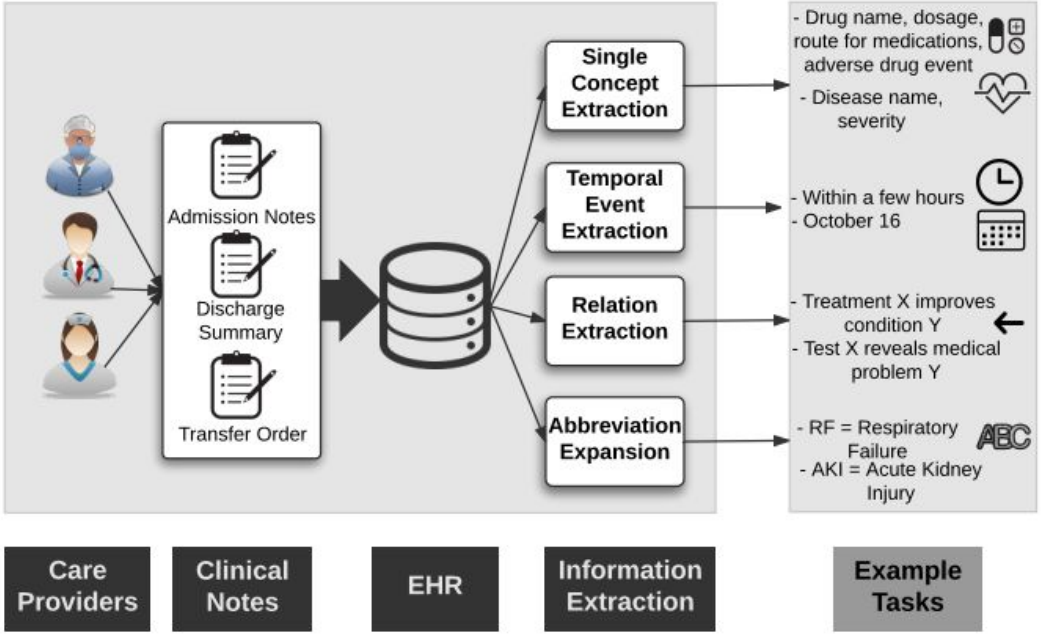
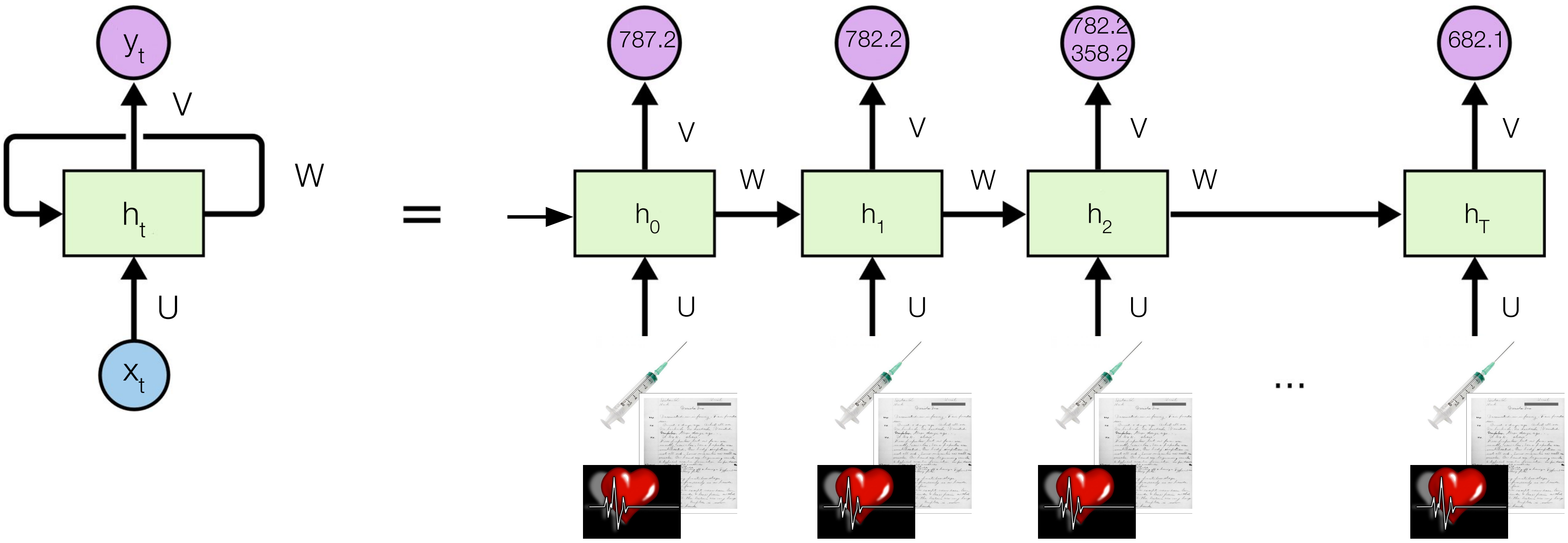


Figure de Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis

Et les entrées?



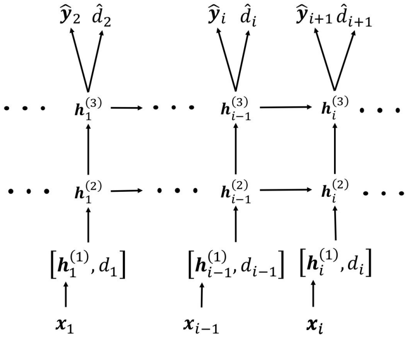
Plan

1. Introduction aux réseaux de neurones récurrents
2. Données médicales
3. **Diverses applications**

Prédiction de la prochaine visite (Doctor AI)

y = codes médicaux R^{1778}

d = durée écoulée depuis la dernière visite



Codes médicalaux R^{40000}

- Codes médicales: CIM-9, diagnostiques, procédures, médications
- Groupe les codes CIM-9

Table 1: Basic statistics of the the clinical records dataset.

# of patients	263,706	Total # of codes	38,594
Avg. # of visits	54.61	Total # of 3-digit Dx codes	1,183
Avg. # of codes per visit	3.22	# of top level Rx codes	595
Max # of codes per visit	62	Avg. duration between visits	76.12 days

Patients provenant de Sutter Health Palo Alto Medical Foundation

Algorithms	Dx,Rx,Time Recall @k $k = 10$	$k = 20$	$k = 30$	R^2
Last visit	26.25			—
Most freq.	48.11	60.23	66.00	—
Logistic	36.04	46.32	52.53	0.0726
→ MLP	38.82	49.09	55.74	0.1221
RNN-1	53.86	65.10	71.24	0.2519
RNN-2	53.61	64.93	71.14	0.2528
RNN-1-IR	54.37	65.68	71.85	0.2492
→ RNN-2-IR	**54.96**	**66.31**	**72.48**	**0.2534**

Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori MP2893,BAHADORI@GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Andy Schuetz, Walter F. Stewart SCHUETA1,STEWARWF@SUTTERHEALTH.ORG
Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA

Jimeng Sun JSUN@CC.GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

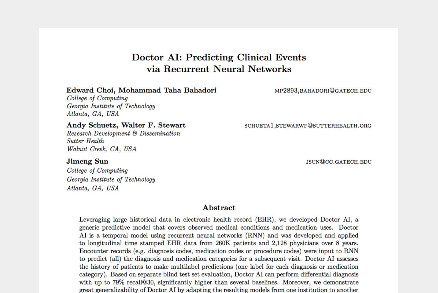
Abstract

Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another

Choi, Doctor AI: Predicting Clinical Event via Recurrent Neural Networks, 2016

Aller au delà des scores

- Difficile à interpréter (black box)
- Mais! On peut étudier le comportement du modèle



Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori MP2893,BAHADORI@GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
Andy Schuetz, Walter F. Stewart SCHUETA1,STEWARWF@SUTTERHEALTH.ORG
Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA
Jimeng Sun JSUN@CC.GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Abstract

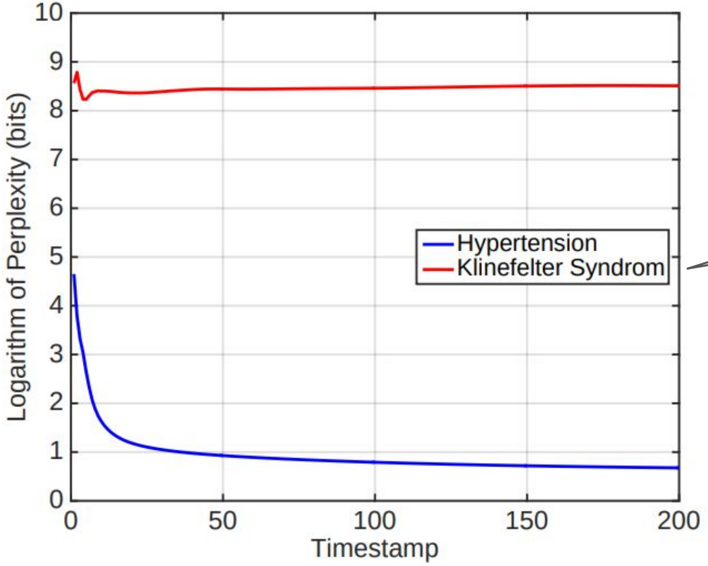
Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another

Algorithms	Dx,Rx,Time Recall @k			
	$k = 10$	$k = 20$	$k = 30$	R^2
Last visit	26.25			—
Most freq.	48.11	60.23	66.00	—
Logistic	36.04	46.32	52.53	0.0726
MLP	38.82	49.09	55.74	0.1221
RNN-1	53.86	65.10	71.24	0.2519
RNN-2	53.61	64.93	71.14	0.2528
RNN-1-IR	54.37	65.68	71.85	0.2492
RNN-2-IR	**54.96**	**66.31**	**72.48**	**0.2534**

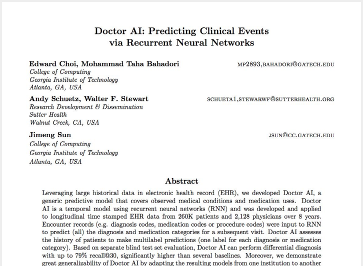
Choi, Doctor AI: Predicting Clinical Event via Recurrent Neural Networks, 2016

Imbalance des classes (Doctor AI)

Marche mieux sur les classes fréquentes



Hommes avec chromosomes XXY



Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori MP2893,BAHADORI@GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Andy Schuetz, Walter F. Stewart SCHUETA1,STEWARWF@SUTTERHEALTH.ORG
Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA

Jimeng Sun JSUN@CC.GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

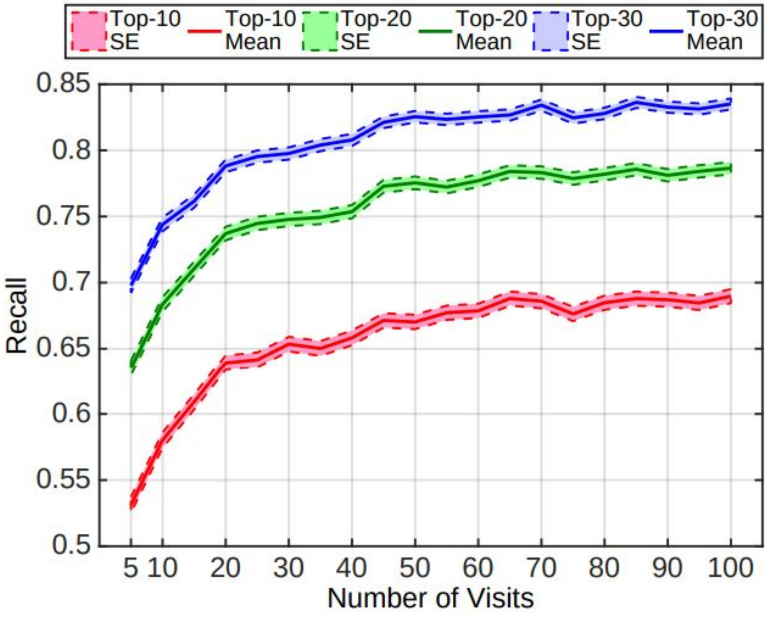
Abstract

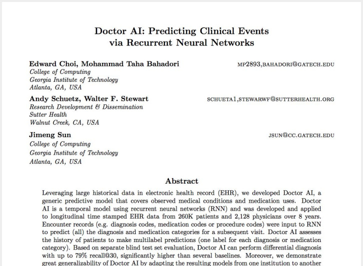
Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another

Choi, Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, 2016

Durée des visites (Doctor AI)

Devient meilleur avec le temps (plus prévisibles)





Doctor AI: Predicting Clinical Events
via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori MP2893,BAHADORI@GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Andy Schuetz, Walter F. Stewart SCHUETA1,STEWARWF@SUTTERHEALTH.ORG
Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA

Jimeng Sun JSUN@CC.GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Abstract

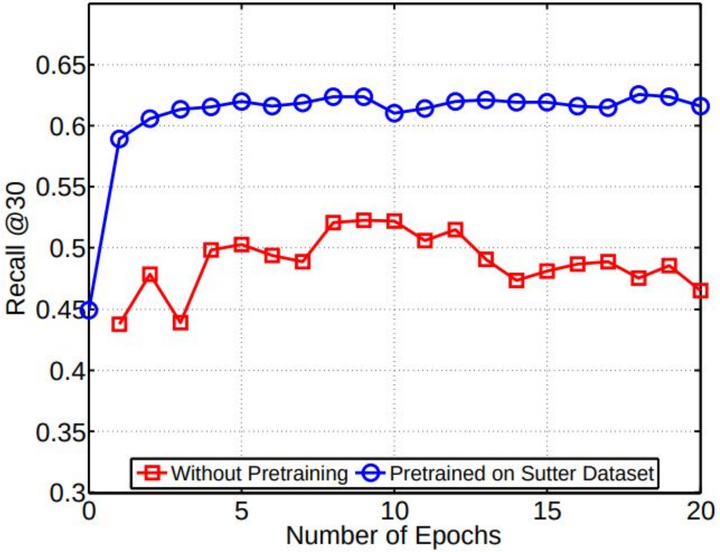
Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another

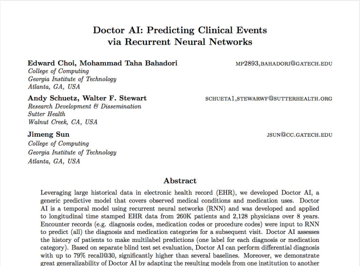
Choi, Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, 2016

Transfert de domaine (Doctor AI)

MIMIC II possède 2,695 patients avec 2+ visites

Démographiques différentes





Doctor AI: Predicting Clinical Events
via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori MP2893,BAHADORI@GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Andy Schuetz, Walter F. Stewart SCHUETA1,STEWARWF@SUTTERHEALTH.ORG
Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA

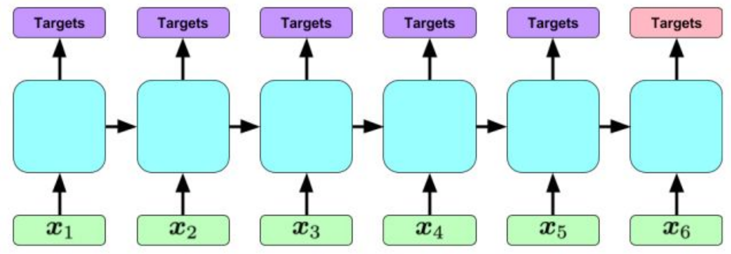
Jimeng Sun JSUN@CC.GATECH.EDU
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Abstract

Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another

Choi, Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, 2016

Diagnostique d'une visite



Entrées: Pression sanguine, pH, température corporelle, etc…, échantillonné à chaque heure.

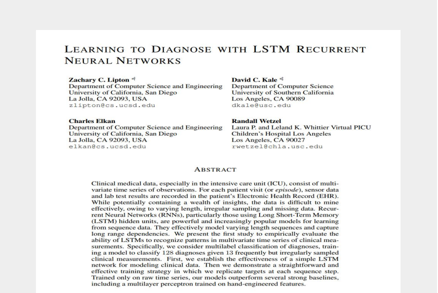
Sorties: 128 codes, ***similaire*** à CIM-9

Données: 10 401 épisodes, durée variables.
Provient de Children's Hospital LA

Classification performance for 128 ICU phenotypes

Model	Micro AUC	Macro AUC	Micro F1	Macro F1	Prec. at 10
Base Rate	0.7128	0.5	0.1346	0.0343	0.0788
Log. Reg., First 6 + Last 6	0.8122	0.7404	0.2324	0.1081	0.1016
Log. Reg., Expert features	0.8285	0.7644	0.2502	0.1373	0.1087
MLP, First 6 + Last 6	0.8375	0.7770	0.2698	0.1286	0.1096
MLP, Expert features	**0.8551**	**0.8030**	**0.2930**	**0.1475**	**0.1170**
LSTM Models with two 64-cell hidden layers					
LSTM	0.8241	0.7573	0.2450	0.1170	0.1047
LSTM, AuxOut (Diagnoses)	0.8351	0.7746	0.2627	0.1309	0.1110
LSTM-AO (Categories)	0.8382	0.7748	0.2651	0.1351	0.1099
LSTM-TR	0.8429	0.7870	0.2702	0.1348	0.1115
LSTM-TR-AO (Diagnoses)	0.8391	0.7866	0.2599	0.1317	0.1085
LSTM-TR-AO (Categories)	0.8439	0.7860	0.2774	0.1330	0.1138
LSTM Models with Dropout (probability 0.5) and two 128-cell hidden layers					
LSTM-DO	0.8377	0.7741	0.2748	0.1371	0.1110
LSTM-DO-AO (Diagnoses)	0.8365	0.7785	0.2581	0.1366	0.1104
LSTM-DO-AO (Categories)	0.8399	0.7783	0.2804	0.1361	0.1123
LSTM-DO-TR	**0.8560**	**0.8075**	**0.2938**	0.1485	**0.1172**
LSTM-DO-TR-AO (Diagnoses)	0.8470	0.7929	0.2735	0.1488	0.1149
LSTM-DO-TR-AO (Categories)	0.8543	0.8015	0.2887	0.1446	0.1161
LSTM-DO-TR (Linear Gain)	0.8480	0.7986	0.2896	**0.1530**	0.1160
Ensembles of Best MLP and Best LSTM					
Mean of LSTM-DO-TR & MLP	0.8611	0.8143	0.2981	0.1553	0.1201
Max of LSTM-DO-TR & MLP	**0.8643**	**0.8194**	**0.3035**	**0.1571**	**0.1218**

Faible gain



LEARNING TO DIAGNOSE WITH LSTM RECURRENT NEURAL NETWORKS

Zachary C. Lipton
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093, USA
zlipton@cs.ucsd.edu

David C. Kale
Department of Computer Science
University of Southern California
Los Angeles, CA 90089
dkale@usc.edu

Charles Elkan
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093, USA
elkan@cs.ucsd.edu

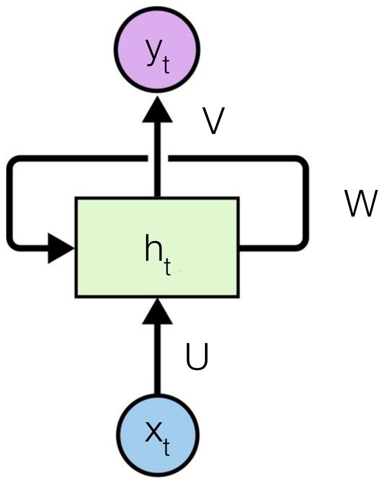
Randall Wetzel
Laura P. and Leland K. Whittier Virtual PICU
Children's Hospital Los Angeles
Los Angeles, CA 90027
rwetzel@chla.usc.edu

ABSTRACT

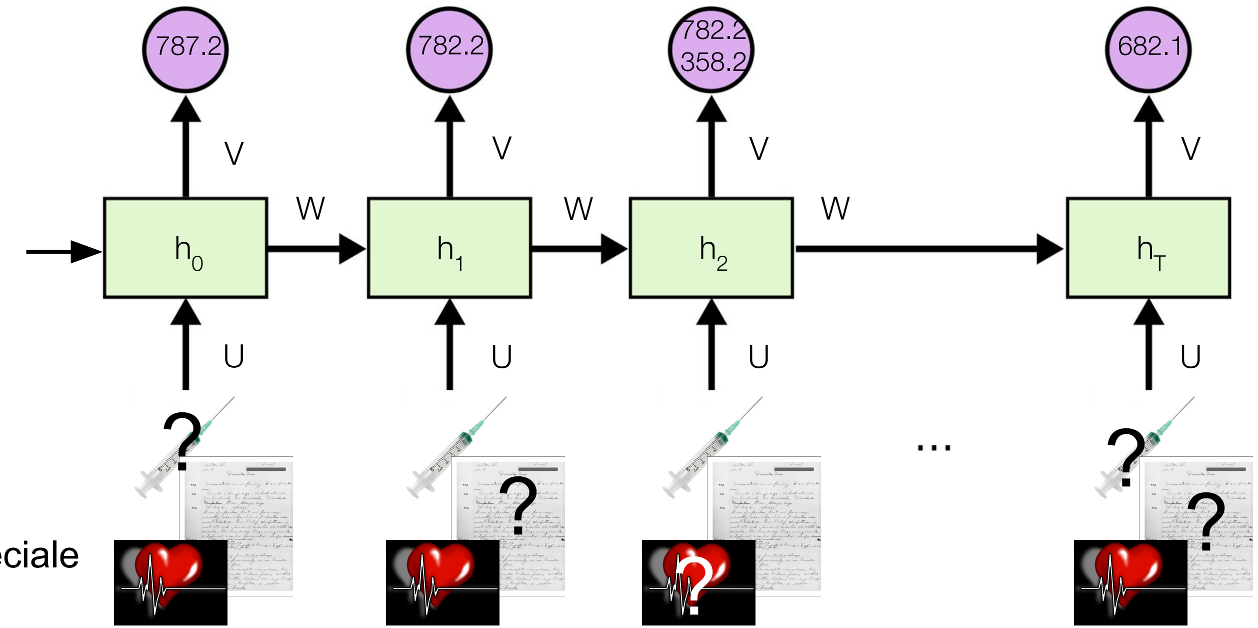
Clinical medical data, especially in the intensive care unit (ICU), consist of multivariate time series of observations. For each patient visit (or *episode*), sensor data and lab test results are recorded in the patient's Electronic Health Record (EHR). While potentially containing a wealth of insights, the data is difficult to mine effectively, owing to varying length, irregular sampling and missing data. Recurrent Neural Networks (RNNs), particularly those using Long Short-Term Memory (LSTM) hidden units, are powerful and increasingly popular models for learning from sequence data. They effectively model varying length sequences and capture long range dependencies. We present the first study to empirically evaluate the ability of LSTMs to recognize patterns in multivariate time series of clinical measurements. Specifically, we consider multilabel classification of diagnoses, training a model to classify 128 diagnoses given 13 frequently but irregularly sampled clinical measurements. First, we establish the effectiveness of a simple LSTM network for modeling clinical data. Then we demonstrate a straightforward and effective training strategy in which we replicate targets at each sequence step. Trained only on raw time series, our models outperform several strong baselines, including a multilayer perceptron trained on hand-engineered features.

Learning to Diagnose with LSTM Recurrent neural Networks, 2015

Données manquantes



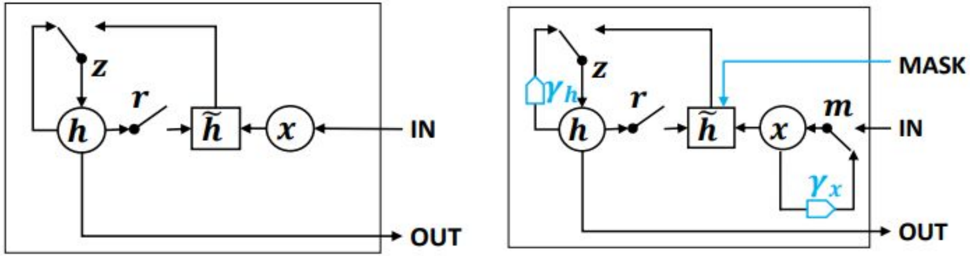
=



- Remplacer par une valeur spéciale
- Prendre la dernière valeur
- Prendre la moyenne

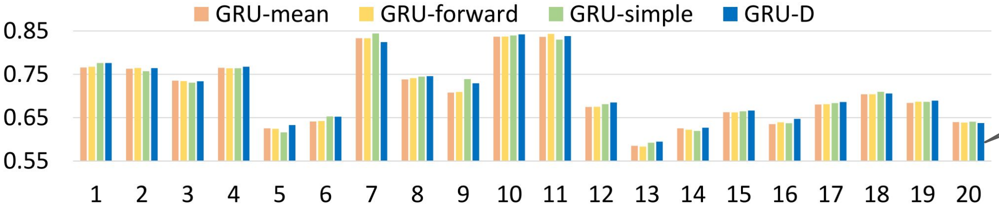
Données manquantes

Ajout de *gates*!



(a) GRU (b) GRU-D

	MIMIC-III
# of samples (N)	19714
# of variables (D)	99
Mean of # of time steps	35.89
Maximum of # of time steps	150
Mean of variable missing rate	0.9621



Faible gain

Figure 9: Performance for predicting 20 ICD-9 diagnosis categories on MIMIC-III dataset. x-axis, ICD-9 diagnosis category id; y-axis, AUC score.

arXiv:1606.01865v2 [cs.LG] 7 Nov 2016

RECURRENT NEURAL NETWORKS FOR MULTIVARIATE TIME SERIES WITH MISSING VALUES

Zhengping Che, Sanjay Purushotham
Department of Computer Science
University of Southern California
Los Angeles, CA 90089, USA
{zche,spurusho}@usc.edu

Kyunghyun Cho, David Sontag
Department of Computer Science
New York University
New York, NY 10012, USA
kyunghyun.cho@nyu.edu, dsontag@cs.nyu.edu

Yan Liu
Department of Computer Science
University of Southern California
Los Angeles, CA 90089, USA
yanliu.cs@usc.edu

ABSTRACT

Multivariate time series data in practical applications, such as health care, geoscience, and biology, are characterized by a variety of missing values. In time series prediction and other related tasks, it has been noted that missing values and their missing patterns are often correlated with the target labels, a.k.a., *informative missingness*. There is very limited work on exploiting the missing patterns for effective imputation and improving prediction performance. In this paper, we develop novel deep learning models, namely GRU-D, as one of the early attempts. GRU-D is based on Gated Recurrent Unit (GRU), a state-of-the-art recurrent neural network. It takes two representations of missing patterns, i.e., *masking* and *time interval*, and effectively incorporates them into a deep model architecture so that it not only captures the long-term temporal dependencies in time series, but also utilizes the missing patterns to achieve better prediction results. Experiments of time series classification tasks on real-world clinical datasets (MIMIC-III, PhysioNet) and synthetic datasets demonstrate that our models achieve state-of-the-art performance and provides useful insights for better understanding and utilization of missing values in time series analysis.

1 INTRODUCTION

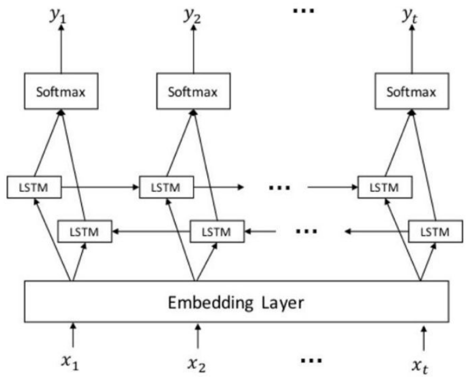
Multivariate time series data are ubiquitous in many practical applications ranging from health care, geoscience, astronomy, to biology and others. They often inevitably carry missing observations due to various reasons, such as medical events, saving costs, anomalies, inconvenience and so on. It has been noted that these missing values are usually *informative missingness* (Rubin, 1976), i.e., the

Recurrent Neural Networks for multivariate time series with missing values, 2016

Notes cliniques

Mr. Smith is a 63-year-old gentleman with coronary artery disease, hypertension, hypercholesterolemia, COPD and tobacco abuse. He reports doing well. He did have some more knee pain for a few weeks, but this has resolved. He is having more trouble with his sinuses. I had started him on Flonase back in December. He says this has not really helped. Over the past couple weeks he has had significant congestion and thick discharge. No fevers or headaches but does have diffuse upper right-sided teeth pain. **He denies any chest pains, palpitations, PND, orthopnea, edema or syncope.** His breathing is doing fine. No cough. He continues to smoke about half-a-pack per day. He plans on **trying the patches again.**

Extraction d’informations à partir de notes cliniques

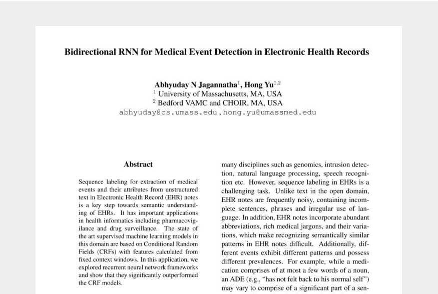


Entrées: Note clinique

Sorties: Chaque mot appartient à un groupe.
Médication (Drug name, Dosage, Duration…)
ou *Maladie (ADE, Indication...)*

Données: 780 notes (moyenne de 786 notes).

Models	**Recall**	**Precision**	**F-score**
CRF-nocontext	0.6562	0.7330	0.6925
CRF-context	0.6806	0.7711	0.7230
LSTM-sentence	0.8024	0.7803	0.7912
GRU-sentence	0.8013	0.7802	0.7906
LSTM-document	0.8050	0.7796	0.7921
GRU-document	**0.8126**	**0.7938**	**0.8031**



Bidirectional RNN for Medical Event Detection in Electronic Health Records

Abhyuday N Jagannatha[1], Hong Yu[1,2]
[1] University of Massachusetts, MA, USA
[2] Bedford VAMC and CHOIR, MA, USA
abhyuday@cs.umass.edu, hong.yu@umassmed.edu

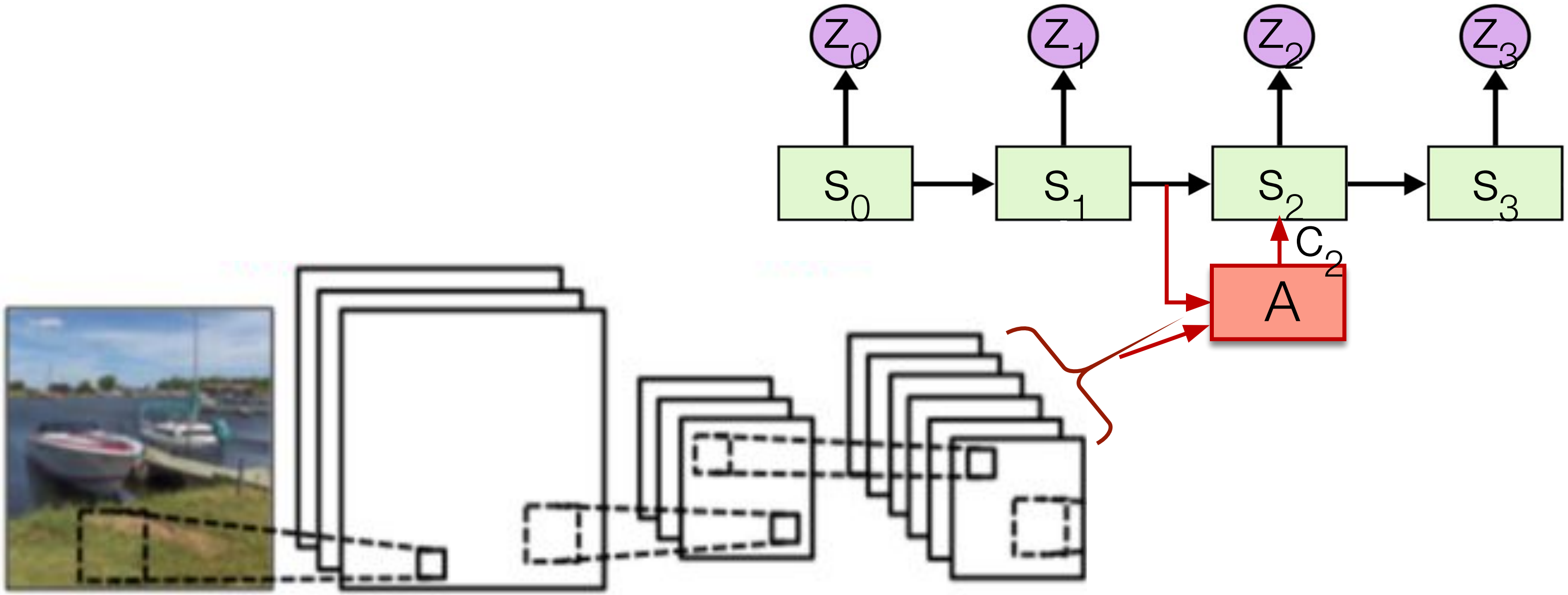
Abstract

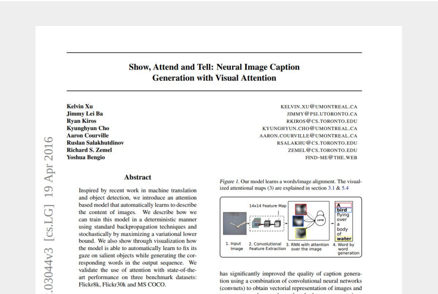
Sequence labeling for extraction of medical events and their attributes from unstructured text in Electronic Health Record (EHR) notes is a key step towards semantic understanding of EHRs. It has important applications in health informatics including pharmacovigilance and drug surveillance. The state of the art supervised machine learning models in this domain are based on Conditional Random Fields (CRFs) with features calculated from fixed context windows. In this application, we explored recurrent neural network frameworks and show that they significantly outperformed the CRF models.

many disciplines such as genomics, intrusion detection, natural language processing, speech recognition etc. However, sequence labeling in EHRs is a challenging task. Unlike text in the open domain, EHR notes are frequently noisy, containing incomplete sentences, phrases and irregular use of language. In addition, EHR notes incorporate abundant abbreviations, rich medical jargons, and their variations, which make recognizing semantically similar patterns in EHR notes difficult. Additionally, different events exhibit different patterns and possess different prevalences. For example, while a medication comprises of at most a few words of a noun, an ADE (e.g., “has not felt back to his normal self”) may vary to comprise of a significant part of a sen-

Jagannatha, Bidirectional RNN for Medical Event Detection in Electronic Health Records, 2016

Génération de légendes





Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Kelvin Xu KELVIN.XU@UMONTREAL.CA
Jimmy Lei Ba JIMMY@PSI.UTORONTO.CA
Ryan Kiros RKIROS@CS.TORONTO.EDU
Kyunghyun Cho KYUNGHYUN.CHO@UMONTREAL.CA
Aaron Courville AARON.COURVILLE@UMONTREAL.CA
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Richard S. Zemel ZEMEL@CS.TORONTO.EDU
Yoshua Bengio FIND-ME@THE.WEB

.03044v3 [cs.LG] 19 Apr 2016

Abstract

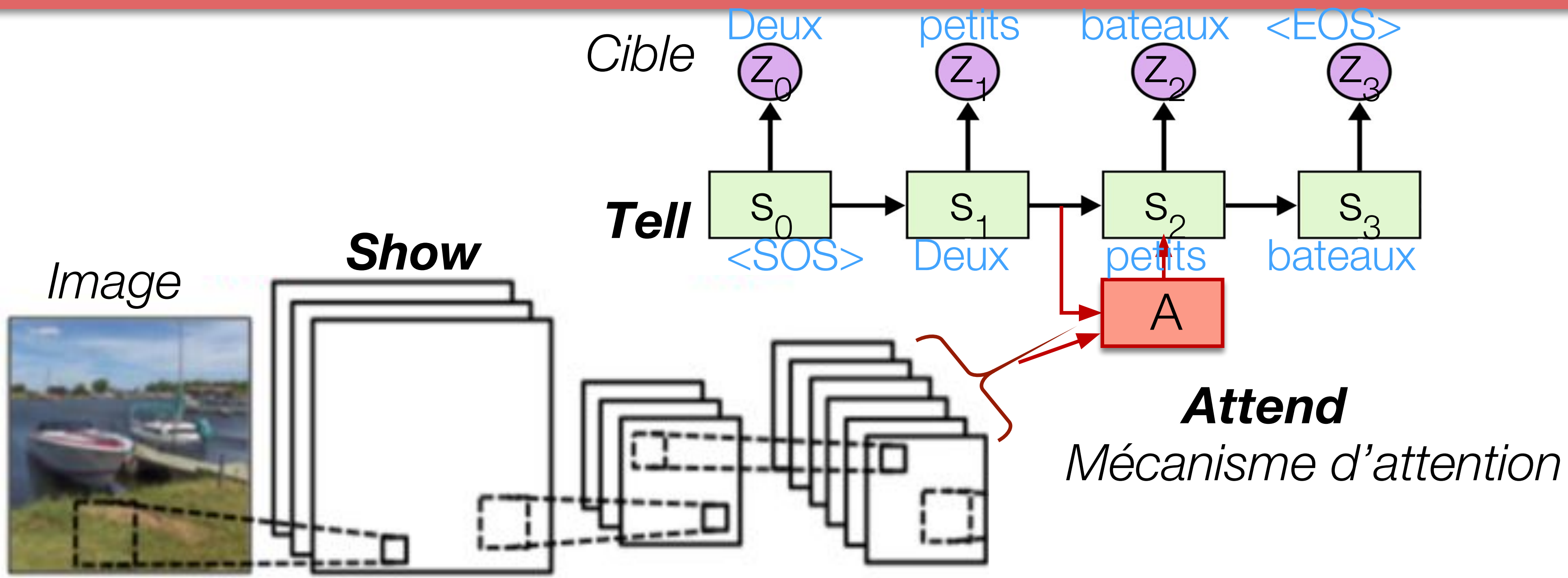
Inspired by recent work in machine translation and object detection, we introduce an attention based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation techniques and stochastically by maximizing a variational lower bound. We also show through visualization how the model is able to automatically learn to fix its gaze on salient objects while generating the corresponding words in the output sequence. We validate the use of attention with state-of-the-art performance on three benchmark datasets: Flickr8k, Flickr30k and MS COCO.

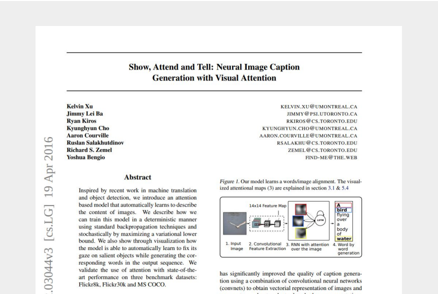
Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in section 3.1 & 5.4

has significantly improved the quality of caption generation using a combination of convolutional neural networks (convnets) to obtain vectorial representation of images and

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2016
Diapo de l'École d'automne 2018, César Laurent

Show, Attend and Tell





.03044v3 [cs.LG] 19 Apr 2016

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Kelvin Xu KELVIN.XU@UMONTREAL.CA
Jimmy Lei Ba JIMMY@PSI.UTORONTO.CA
Ryan Kiros RKIROS@CS.TORONTO.EDU
Kyunghyun Cho KYUNGHYUN.CHO@UMONTREAL.CA
Aaron Courville AARON.COURVILLE@UMONTREAL.CA
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Richard S. Zemel ZEMEL@CS.TORONTO.EDU
Yoshua Bengio FIND-ME@THE.WEB

Abstract

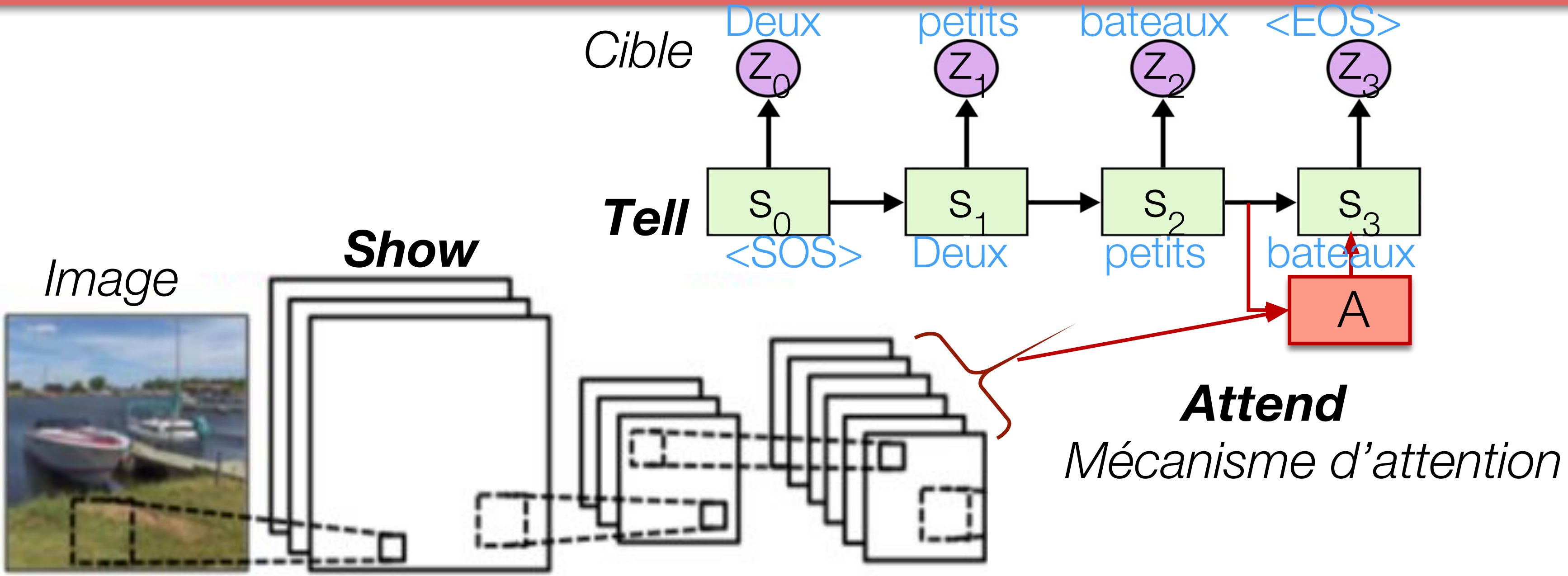
Inspired by recent work in machine translation and object detection, we introduce an attention based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation techniques and stochastically by maximizing a variational lower bound. We also show through visualization how the model is able to automatically learn to fix its gaze on salient objects while generating the corresponding words in the output sequence. We validate the use of attention with state-of-the-art performance on three benchmark datasets: Flickr8k, Flickr30k and MS COCO.

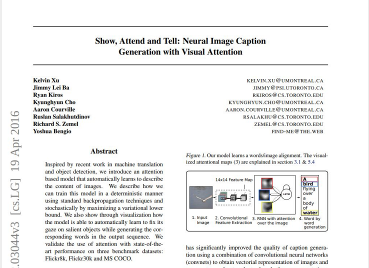
Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in section 3.1 & 5.4

has significantly improved the quality of caption generation using a combination of convolutional neural networks (convnets) to obtain vectorial representation of images and

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2016
Diapo de l'École d'automne 2018, César Laurent

Show, Attend and Tell





.03044v3 [cs.LG] 19 Apr 2016

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Kelvin Xu	KELVIN.XU@UMONTREAL.CA
Jimmy Lei Ba	JIMMY@PSI.UTORONTO.CA
Ryan Kiros	RKIROS@CS.TORONTO.EDU
Kyunghyun Cho	KYUNGHYUN.CHO@UMONTREAL.CA
Aaron Courville	AARON.COURVILLE@UMONTREAL.CA
Ruslan Salakhutdinov	RSALAKHU@CS.TORONTO.EDU
Richard S. Zemel	ZEMEL@CS.TORONTO.EDU
Yoshua Bengio	FIND-ME@THE.WEB

Abstract

Inspired by recent work in machine translation and object detection, we introduce an attention based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation techniques and stochastically by maximizing a variational lower bound. We also show through visualization how the model is able to automatically learn to fix its gaze on salient objects while generating the corresponding words in the output sequence. We validate the use of attention with state-of-the-art performance on three benchmark datasets: Flickr8k, Flickr30k and MS COCO.

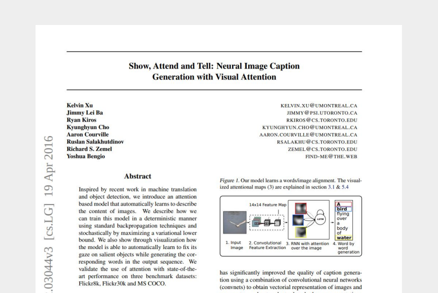
Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in section 3.1 & 5.4

14x14 Feature Map; 1. Input Image; 2. Convolutional Feature Extraction; 3. RNN with attention over the image; 4. Word by word generation; A bird flying over a body of water

has significantly improved the quality of caption generation using a combination of convolutional neural networks (convnets) to obtain vectorial representation of images and

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2016
Diapo de l'École d'automne 2018, César Laurent

Show, Attend and Tell



.03044v3 [cs.LG] 19 Apr 2016

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Kelvin Xu KELVIN.XU@UMONTREAL.CA
Jimmy Lei Ba JIMMY@PSI.UTORONTO.CA
Ryan Kiros RKIROS@CS.TORONTO.EDU
Kyunghyun Cho KYUNGHYUN.CHO@UMONTREAL.CA
Aaron Courville AARON.COURVILLE@UMONTREAL.CA
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Richard S. Zemel ZEMEL@CS.TORONTO.EDU
Yoshua Bengio FIND-ME@THE.WEB

Abstract

Inspired by recent work in machine translation and object detection, we introduce an attention based model that automatically learns to describe the content of images. We describe how we can train this model in a deterministic manner using standard backpropagation techniques and stochastically by maximizing a variational lower bound. We also show through visualization how the model is able to automatically learn to fix its gaze on salient objects while generating the corresponding words in the output sequence. We validate the use of attention with state-of-the-art performance on three benchmark datasets: Flickr8k, Flickr30k and MS COCO.

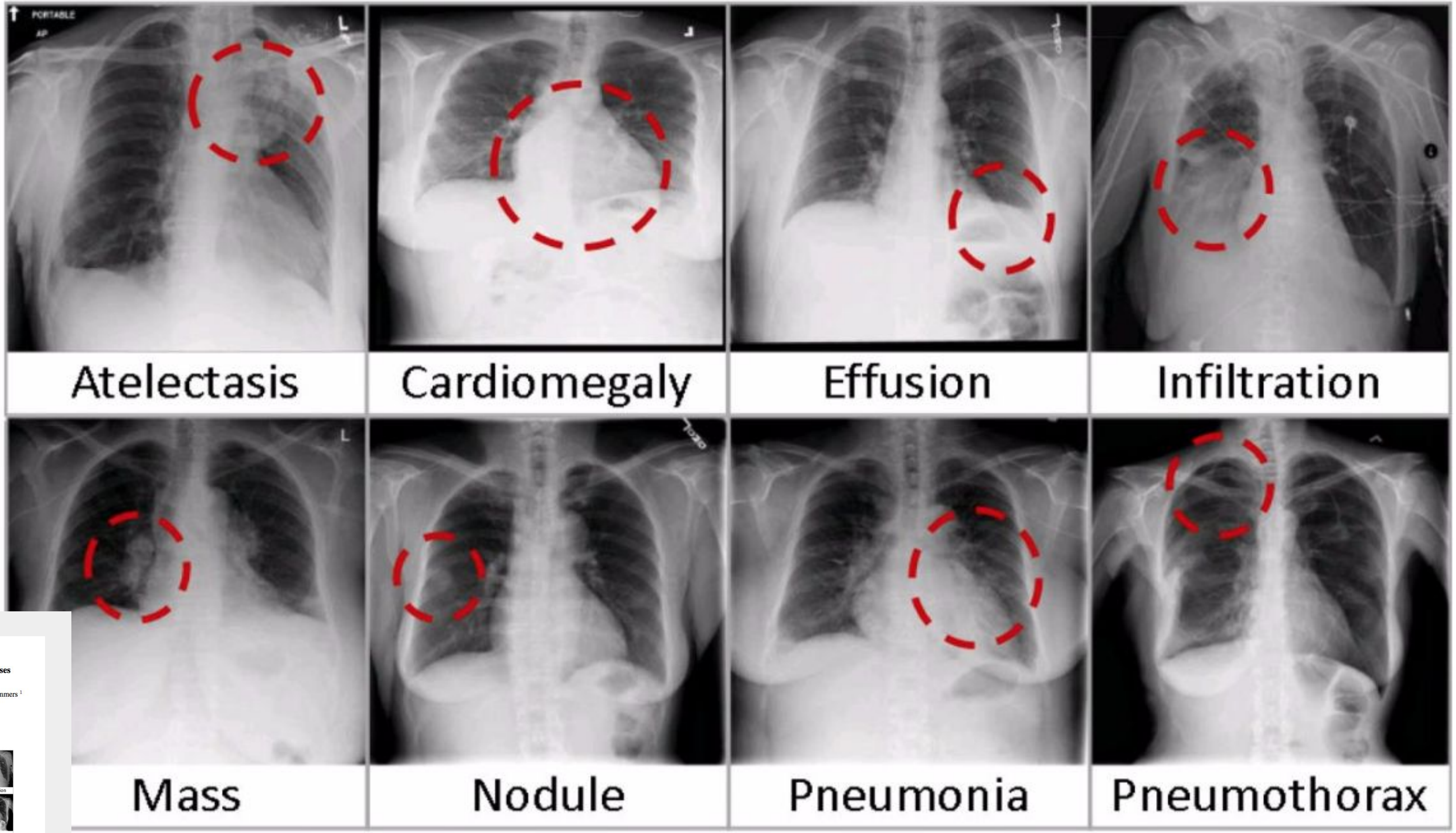
Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in section 3.1 & 5.4

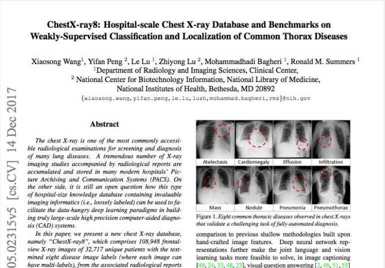
14x14 Feature Map
A bird flying over a body of water
1. Input Image 2. Convolutional Feature Extraction 3. RNN with attention over the image 4. Word by word generation

has significantly improved the quality of caption generation using a combination of convolutional neural networks (convnets) to obtain vectorial representation of images and

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2016
Diapo de l’École d’automne 2018, César Laurent

ChestX-ray 8





05.02315v5 [cs.CV] 14 Dec 2017

ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

Xiaosong Wang[1], Yifan Peng [2], Le Lu [1], Zhiyong Lu [2], Mohammadhadi Bagheri [1], Ronald M. Summers [1]
[1]Department of Radiology and Imaging Sciences, Clinical Center,
[2] National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, Bethesda, MD 20892
{xiaosong.wang,yifan.peng,le.lu,luzh,mohammad.bagheri,rms}@nih.gov

Abstract

The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). On the other side, it is still an open question how this type of hospital-size knowledge database containing invaluable imaging informatics (i.e., loosely labeled) can be used to facilitate the data-hungry deep learning paradigms in building truly large-scale high precision computer-aided diagnosis (CAD) systems.

In this paper, we present a new chest X-ray database, namely "ChestX-ray8", which comprises 108,948 frontal-view X-ray images of 32,717 unique patients with the text-mined eight disease image labels (where each image can have multi-labels), from the associated radiological reports

Figure 1. *Eight common thoracic diseases observed in chest X-rays that validate a challenging task of fully-automated diagnosis.*

comparison to previous shallow methodologies built upon hand-crafted image features. Deep neural network representations further make the joint language and vision learning tasks more feasible to solve, in image captioning [49, 24, 33, 48, 23], visual question answering [2, 46, 51, 55]

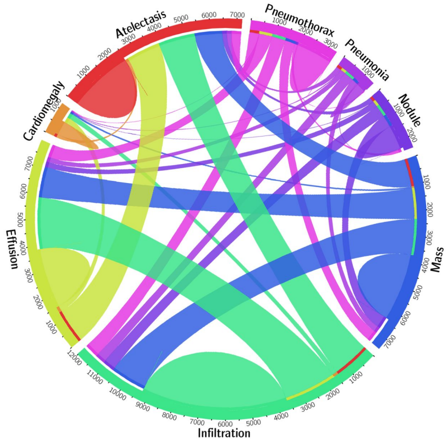
Wang, ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, 2017

ChestX-ray 8

Setting	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
AlexNet	0.6458	0.6925	0.6642	0.6041	**0.5644**	0.6487	0.5493	0.7425
GoogLeNet	0.6307	0.7056	0.6876	0.6088	0.5363	0.5579	0.5990	0.7824
VGGNet-16	0.6281	0.7084	0.6502	0.5896	0.5103	0.6556	0.5100	0.7516
ResNet-50	**0.7069**	**0.8141**	**0.7362**	**0.6128**	0.5609	**0.7164**	**0.6333**	**0.7891**

AUCs for chaque classes d’un modèle à plusieurs sorties

- 32,717 patients
- 108,948 X-Rays
- Classes extraite de ***notes cliniques***



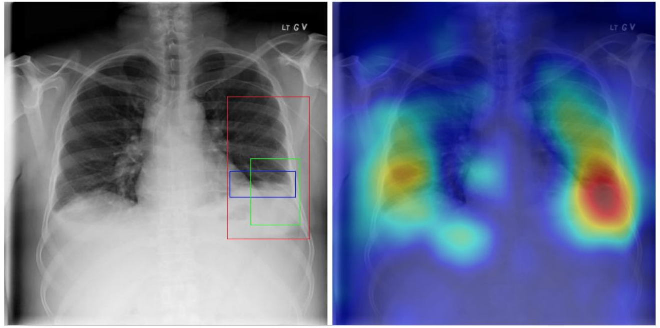
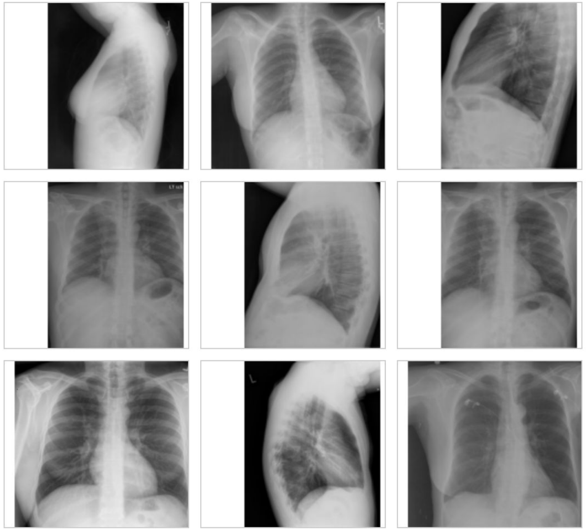
Radiology report	Keyword	Localization Result
findings include: 1. left basilar atelectasis/consolidation. 2. prominent hilum (mediastinal adenopathy). 3. left pic catheter (tip in atriocaval junction). 4. stable, normal appearing cardiomediastinal silhouette. impression: small right pleural effusion otherwise stable abnormal study including left basilar infiltrate/atelectasis, prominent hilum, and position of left pic catheter (tip atriocaval junction).	Effusion; Infiltration; Atelectasis	

Table 8. *A sample of chest x-ray radiology report, mined disease keywords and localization result from the "Atelectasis" Class. Correct bounding box (in green), false positives (in red) and the ground truth (in blue) are plotted over the original image.*

Rapports de l’hôpital universitaire de l'Indiana

Images de radiographie pulmonaire du réseau hospitalier de l'Université de l'Indiana



Indiana University Chest X-ray Collection

Kohli MD, Rosenman M - (2013)

Affiliation: Indiana University

ABSTRACT

Comparison: None.

Indication: Positive TB test

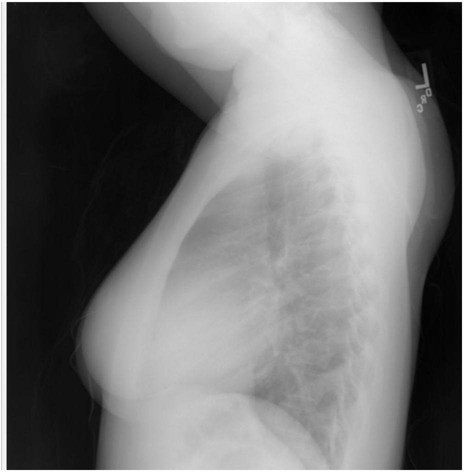
Findings: The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.

Impression: Normal chest x-XXXX.

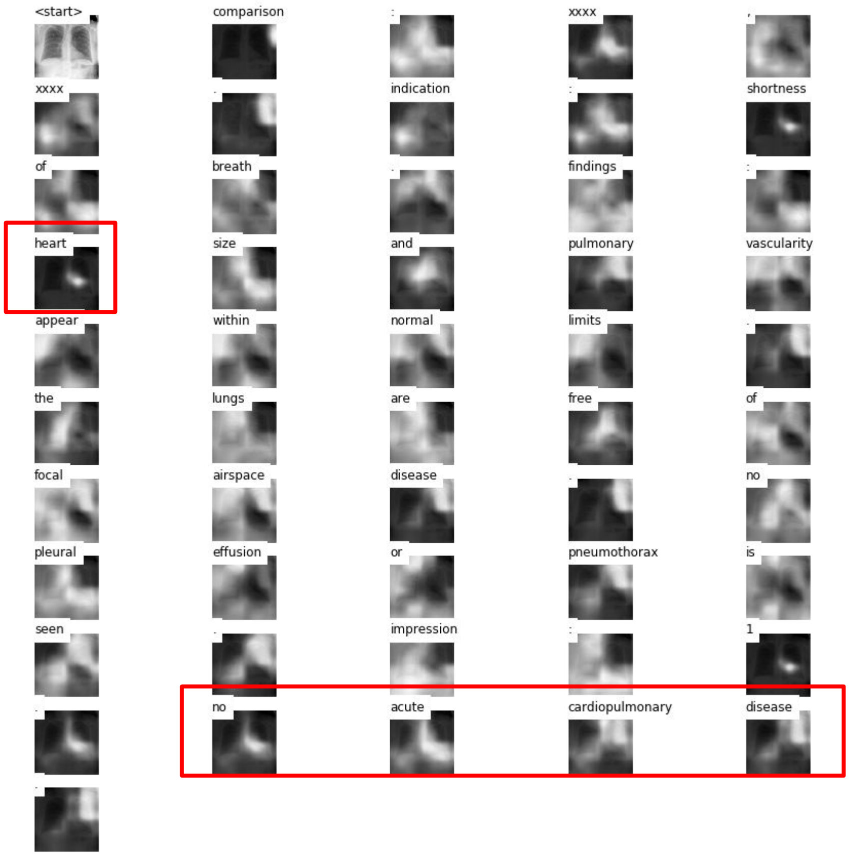
NOTE: The data are drawn from multiple hospital systems.

Show MeSH

Related in: MedlinePlus Request Collection



Génération de légendes sur des notes cliniques



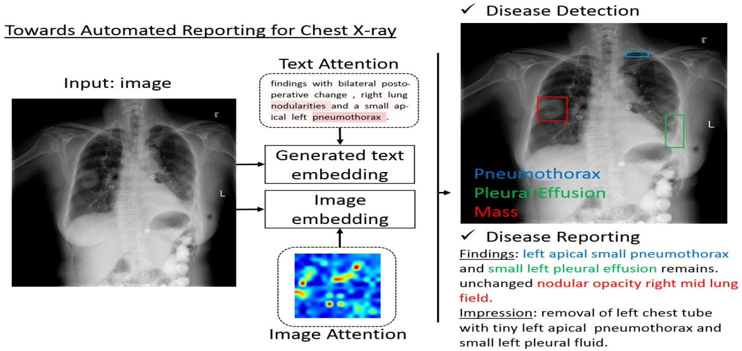
Code source: https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Image-Captioning

Génération de légendes sur des notes cliniques

Super générique!

Code source: https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Image-Captioning

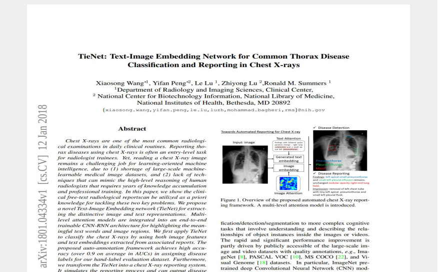
Génération de légendes ET classification



	Captioning [37]	TieNet I+GR
BLEU-1	0.2391	0.2860
BLEU-2	0.1248	0.1597
BLEU-3	0.0861	0.1038
BLEU-4	0.0658	0.0736
METEOR	0.1024	0.1076
ROUGE-L	0.1988	0.2263

Attention!
Implication dangereuse!

Disease	OpenI R /	I+R /	I [35] /	I+GR /	#
Atelectasis	.981 /	.976 /	.702 /	.774 /	293
Cardiomegaly	.944 /	.962 /	.803 /	.847 /	315
Effusion	.968 /	.977 /	.890 /	.899 /	140
Infiltration	.981 /	.984 /	.585 /	.718 /	57
Mass	.959 /	.903 /	.756 /	.723 /	14
Nodule	.967 /	.960 /	.647 /	.658 /	102
Pneumonia	.983 /	.994 /	.642 /	.731 /	36
Pneumothorax	.960 /	.960 /	.631 /	.709 /	22
Consolidation	.969 /	.989 /	.790 /	.855 /	28
Edema	.984 /	.995 /	.799 /	.879 /	40
Emphysema	.849 /	.868 /	.675 /	.792 /	94
Fibrosis	.985 /	.960 /	.744 /	.791 /	18
PT	.948 /	.953 /	.691 /	.749 /	52
Hernia	– /	– /	– /	– /	0
NoFinding	.933 /	.936 /	– /	.747 /	2789
AVG	.960 /	.965 /	.719 /	.779 /	–
#wAVG	.957 /	.966 /	**.741 /**	**.798 /**	–



Wang, TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays, 2018

	I	J	K	L
Image Sample cases				
P	**Emphysema**	**Infiltration**	**Mass Consolidation**	**No finding Nodule**
Original report	minimal-moderate left neck and upper chest subcutaneous emphysema unchanged or minimally decreasing. moderate-marked right chest , minimal-moderate right neck and upper abdomen subcutaneous emphysema.	findings : interval development of left upper lobe patchy nodular infiltrate inferiorly . unchanged radiopaque catheter compatible with vp shunt . stable catheter overlying the stomach . costophrenic angles are clear . cardiac and mediastinal borders are within normal limits of size . impression : interval development of left upper lobe patchy nodular infiltrate	findings : a single ap view of the chest demonstrates unchanged or minimally increasing , dependent positioning , right lung mass/consolidation. the cardiac and mediastinal contours are stable . impression : 1. unchanged or minimally increasing , dependent positioning , right lung mass/consolidation . 2. no evidence of developing infiltrate the visualized left lung	findings : lungs are well aerated with no evidence of infiltrate . cardiac and mediastinal borders are within normal limits of size . .. . impression : no evidence of infiltrate
Generated Report	reason for study : s/p vats clinical informati on : aplastic anemia chest 1 view : chest x-ray performed on the same day . the heart and mediastinum are normal . the subcutaneous emphysema is seen in the right neck and neck on the right . there is unchanged subcutaneous emphysema seen on the right .	reason for exam (entered by ordering clinician into cris): r/o acute , r/opulmonary disease interval changes no interval change and seen areavailable of the heart . there are a nodular interstitial infiltrateor scarring as well as mild degenerative changes of the spine .	findings include 1. left lung consolidation . 2. left lung consolidation .	Findings:heartand mediastinum unchanged . lungs unchanged , no evidence of acute infiltrates . nodule projecting on posterior ribposteriorly . osseous structures intact . impression : stable chest .

Questions?