

Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles

Mohamed BOUAZIZ

Soutenance de thèse, le 06 décembre 2017

Directeur de thèse :
Georges LINARÈS

Co-encadrants :
Mohamed MORCHID
Richard DUFOUR

Thèse Cifre

Entreprise EDD

- Gestion des ressources multimédias
 - 80 000 nouveaux documents quotidiens
 - des centaines de flux radio et TV
- Services de veille médiatique
 - Panoramas de l'actualité
 - Notification en temps réel des affaires sensibles



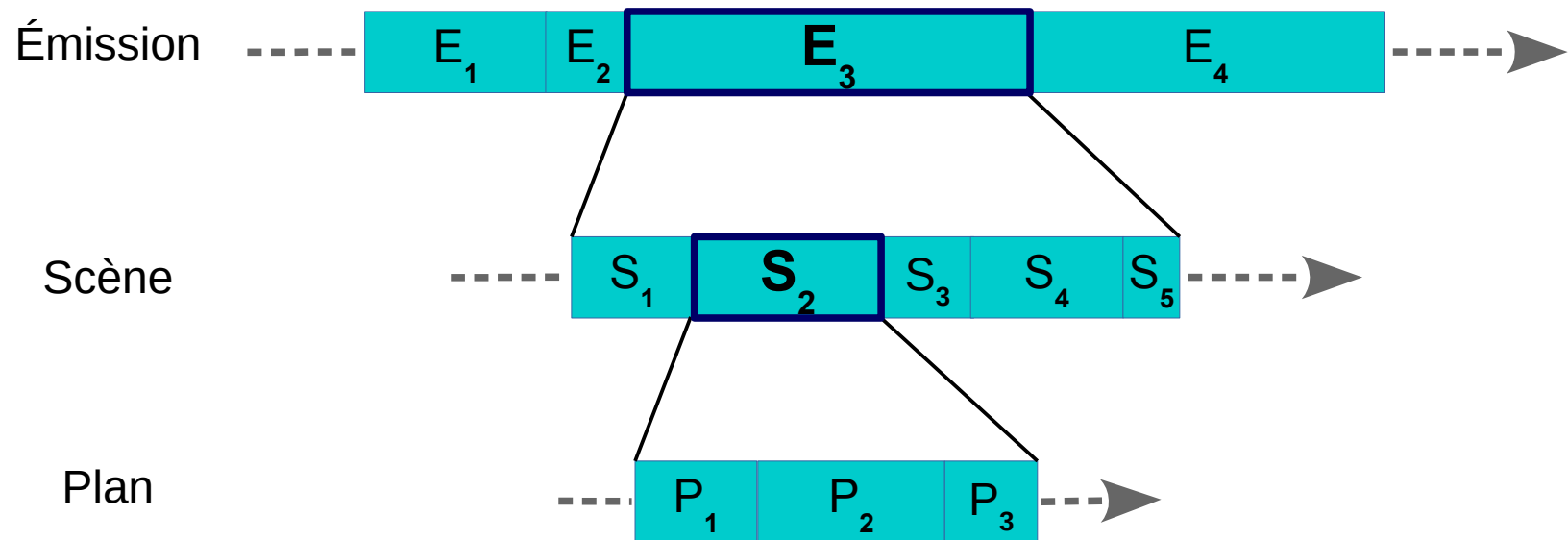
Laboratoire Informatique d'Avignon

- Traitement automatique du langage et de la parole
- Structuration de contenu

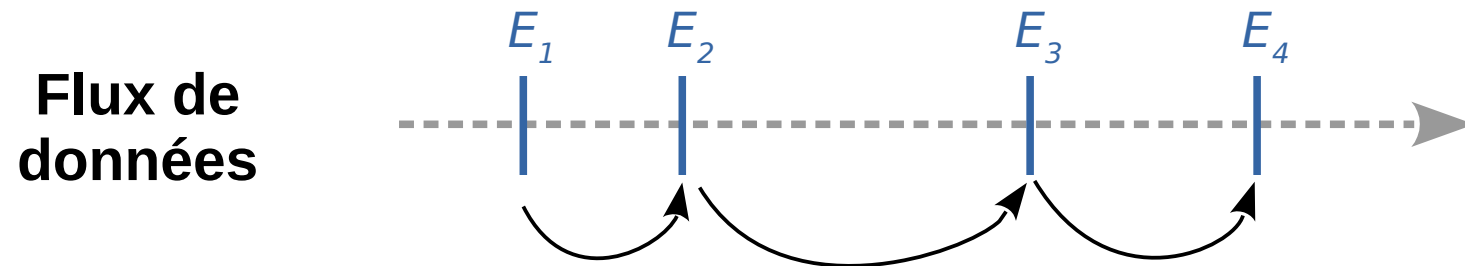


Contenu TV

- Flux de données à différents niveaux de granularité

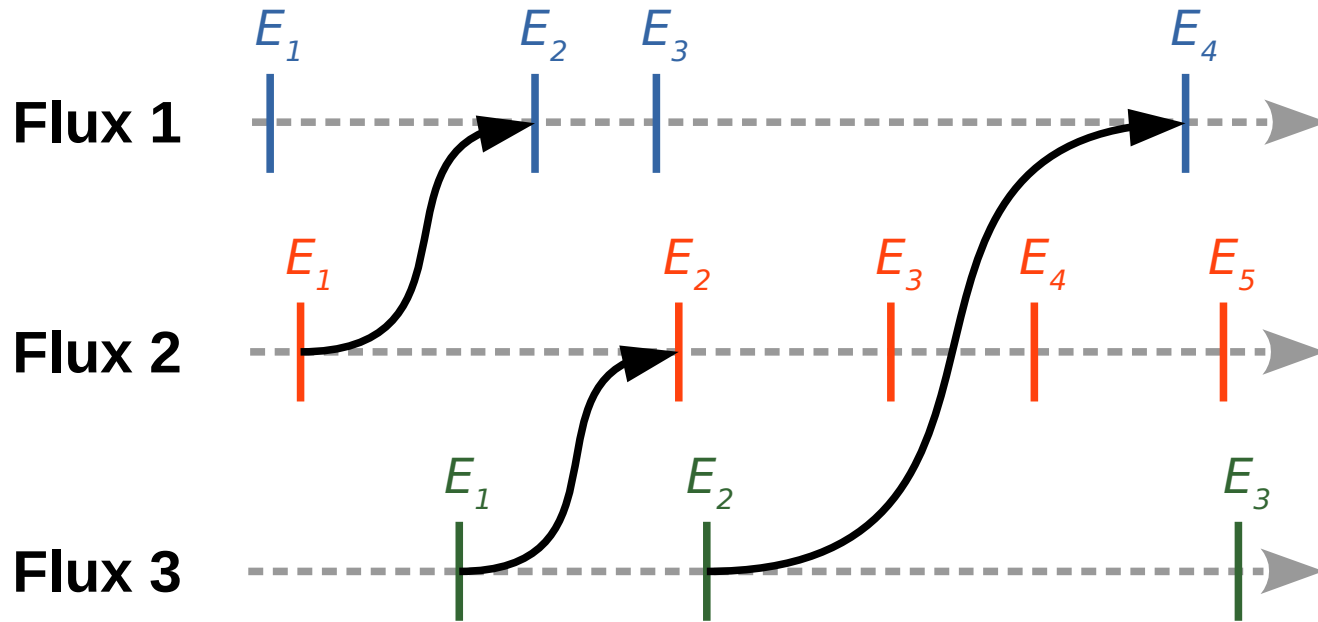


Données séquentielles



- Suite chronologique d'événements
- Relation pouvant exister entre les différents événements
- Traitement automatisé \Rightarrow Méthodes adaptées aux séquences

Flux parallèles



- Plusieurs flux peuvent être émis en parallèle
→ événements asynchrones
- Relations entre les flux (dépendance, concurrence...)

Motivation :

- Exploiter les relations entre les différents flux parallèles pour enrichir les connaissances sur un flux particulier
- Les approches actuelles adaptées aux séquences ne peuvent prendre en compte que des données séquentielles provenant d'un seul flux (**monoflux**)

Objectif :

Concevoir des approches capables d'intégrer simultanément des données séquentielles provenant de plusieurs flux (**multiflux**)

Plan

I. État de l'art : apprentissage automatique pour la classification de séquences

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

IV. LSTM Parallèles (PLSTM)

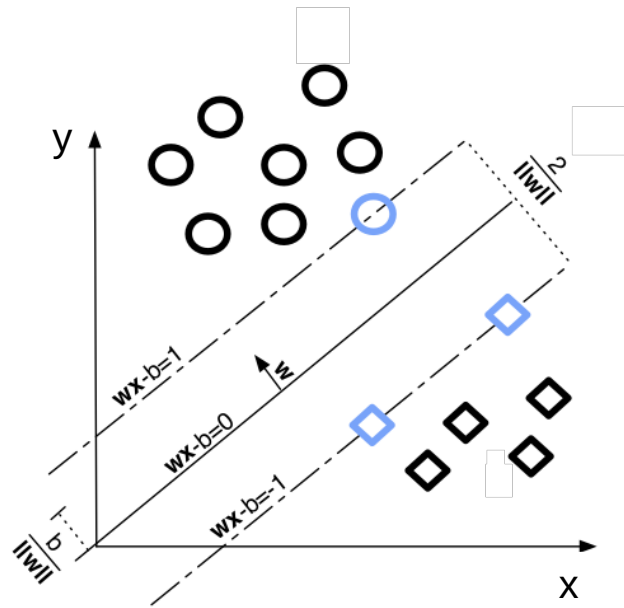
V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

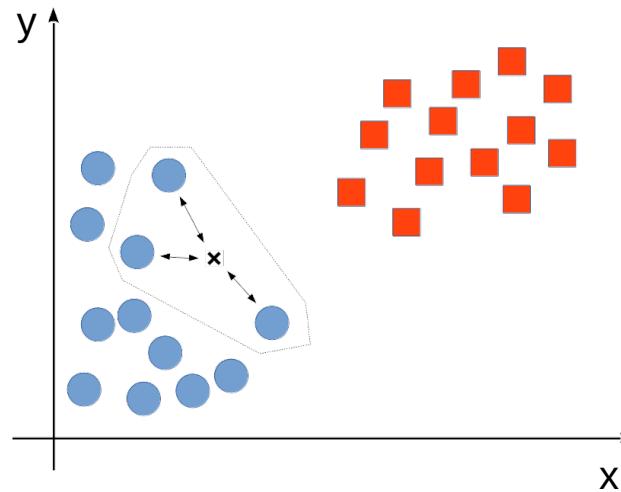
VII. Conclusions et perspectives

I. État de l'art : apprentissage automatique pour la classification de séquences

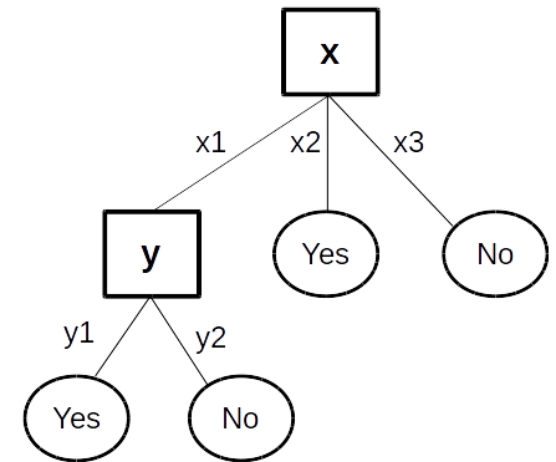
Méthodes « classiques »



SVM



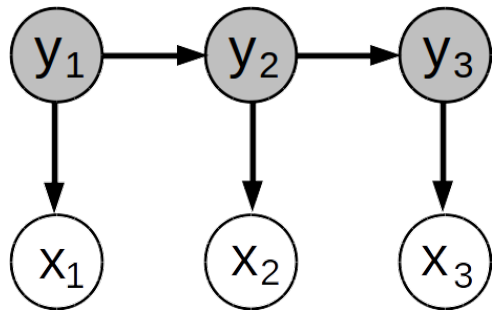
kNN



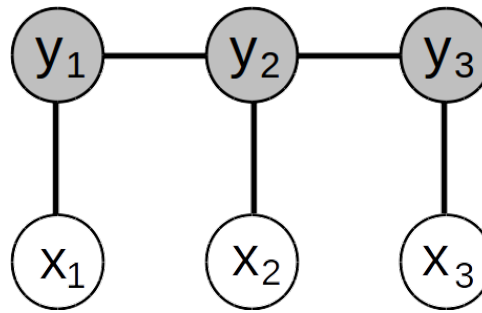
Arbres de décision

- Entrée = caractéristiques prises en compte indépendamment les unes des autres
 - ⇒ Données séquentielles : liens entre les événements ignorés
- ⇒ **Nécessité d'une étape d'extraction de caractéristiques**

Méthodes « adaptées aux séquences »



HMM



CRF

$$P(w_i | w_1 w_2 \cdots w_{i-1})$$

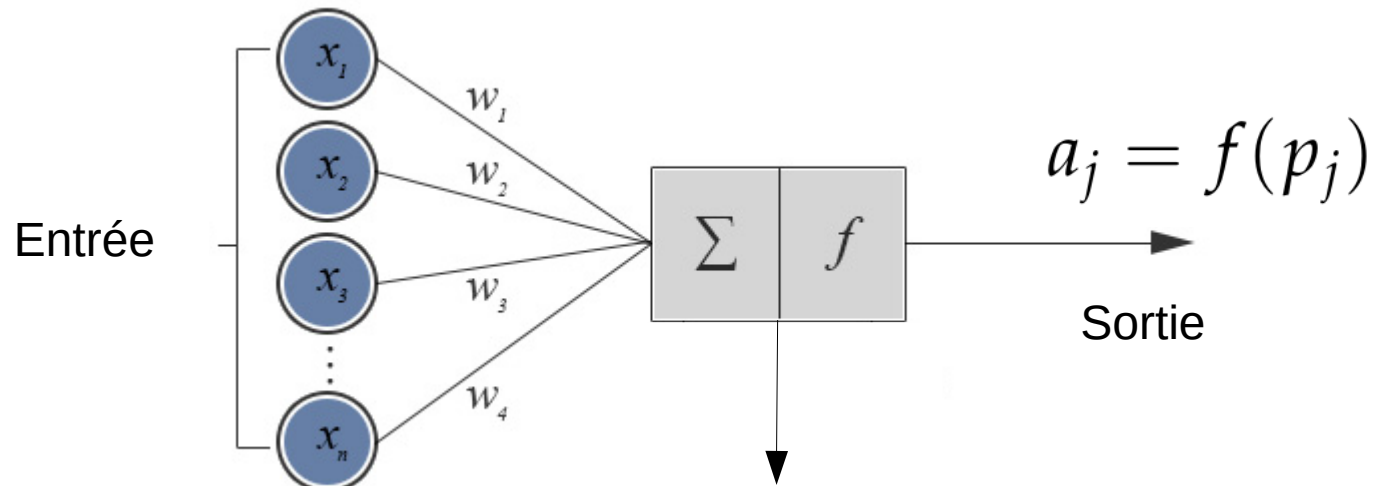
Modèle n-gramme

- Capacité à modéliser les dépendances séquentielles



Modèles à base de réseaux de neurones

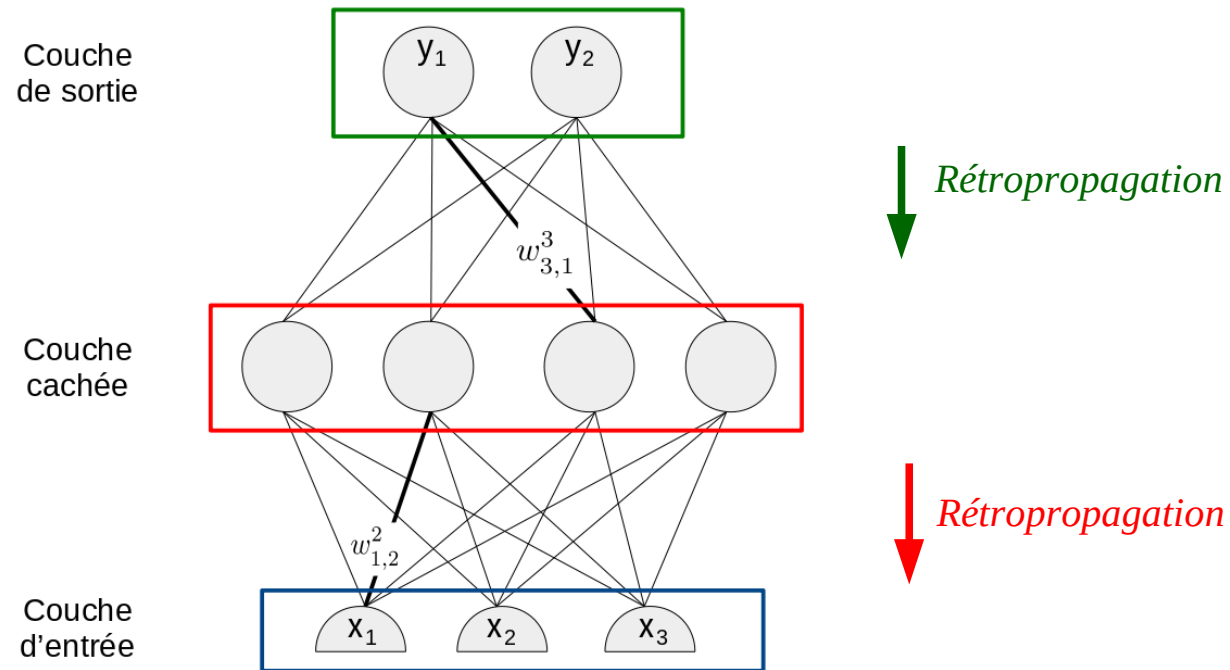
Concepts de base : neurone formel (perceptron)



$$p_j = \sum_{i=1}^N w_{ij}a_i + b_j$$

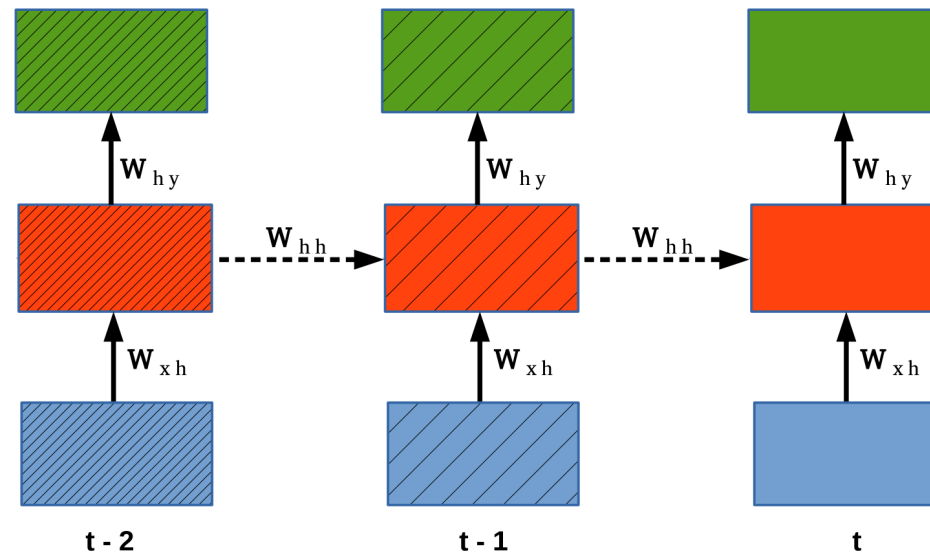
- Utilisé pour :
 - La régression (approximation d'une valeur)
 - La classification (prédiction d'une classe)

Concepts de base : Perceptron Multicouches (MLP)



- ✓ Couches cachées : représentations latentes des données
 - Apprentissage : rétropropagation du gradient

Réseaux de neurones récurrents (RNN)



- Suite de réseaux MLP

- Sortie à l'instant t :
$$h_t = \mathcal{H}(\mathbf{W}_{xh}x_t + \mathbf{W}_{hh}h_{t-1} + b_h)$$
$$y_t = \mathbf{W}_{hy}h_t + b_y$$

✓ Modélisation des relations latentes entre les événements

✗ Séquences longues : « Dissipation du gradient »

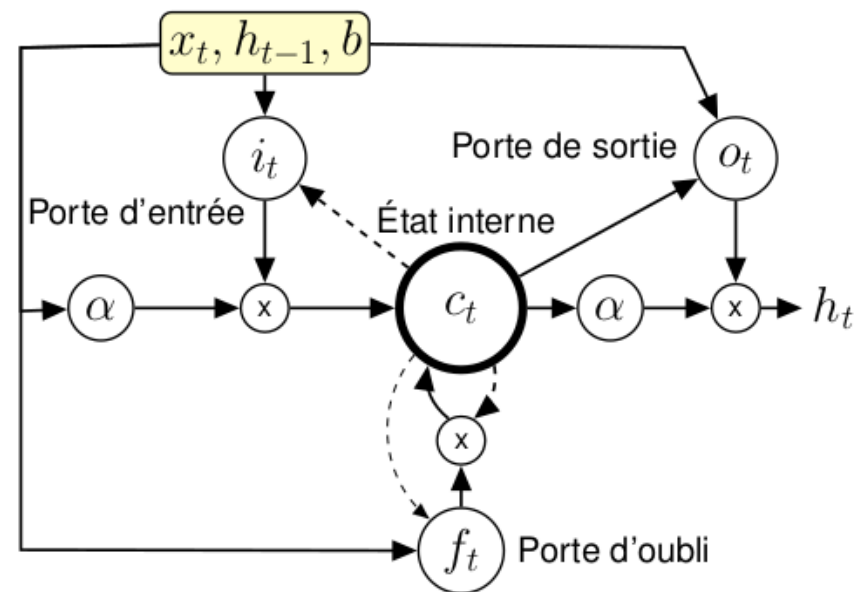
Long Short-Term Memory (LSTM)

- Une extension des RNN :
 - ✓ Modulation de la propagation de l'information
 - ✓ Séquences longues : prévenir la « dissipation du gradient »
- Couche cachée = cellule à mémoire

\mathcal{H}



$$\begin{aligned}i_t &= \sigma(\mathbf{W}_{xi}x_t + \mathbf{W}_{hi}h_{t-1} + \mathbf{W}_{ci}c_{t-1} + b_i) \\f_t &= \sigma(\mathbf{W}_{xf}x_t + \mathbf{W}_{hf}h_{t-1} + \mathbf{W}_{cf}c_{t-1} + b_f) \\c_t &= f_t c_{t-1} + i_t \tanh(\mathbf{W}_{xc}x_t + \mathbf{W}_{hc}h_{t-1} + b_c) \\o_t &= \sigma(\mathbf{W}_{xo}x_t + \mathbf{W}_{ho}h_{t-1} + \mathbf{W}_{co}c_t + b_o) \\h_t &= o_t \tanh(c_t)\end{aligned}$$



Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

IV. LSTM Parallèles (PLSTM)

V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

VII. Conclusions et perspectives

Information du genre d'émission

- Utile pour la sélection et l'orientation de certains traitements automatiques
 - segmentation en thèmes (journaux télévisés, magazine de débat)
 - extraction de moments forts (émissions de sport ou de divertissement)
- Contenue dans les guides de programmes, mais de mauvaise qualité
 - taxonomies ambiguës : *Société, Politique, Animalier, Voyage*
 - programmes incomplets (ex : programme de nuit)

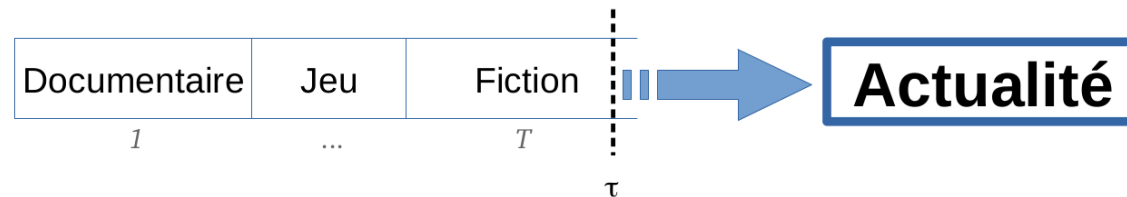
⇒ **Prédire automatiquement le genre d'émission**

Classification en genre à partir du contenu

- **Indices visuels** : fréquence des plans, détection d'objets...
- **Indices acoustiques** : taux de l'énergie, événements sonores...
- **Indices linguistiques** : mesure de confiance, TF-IDF...

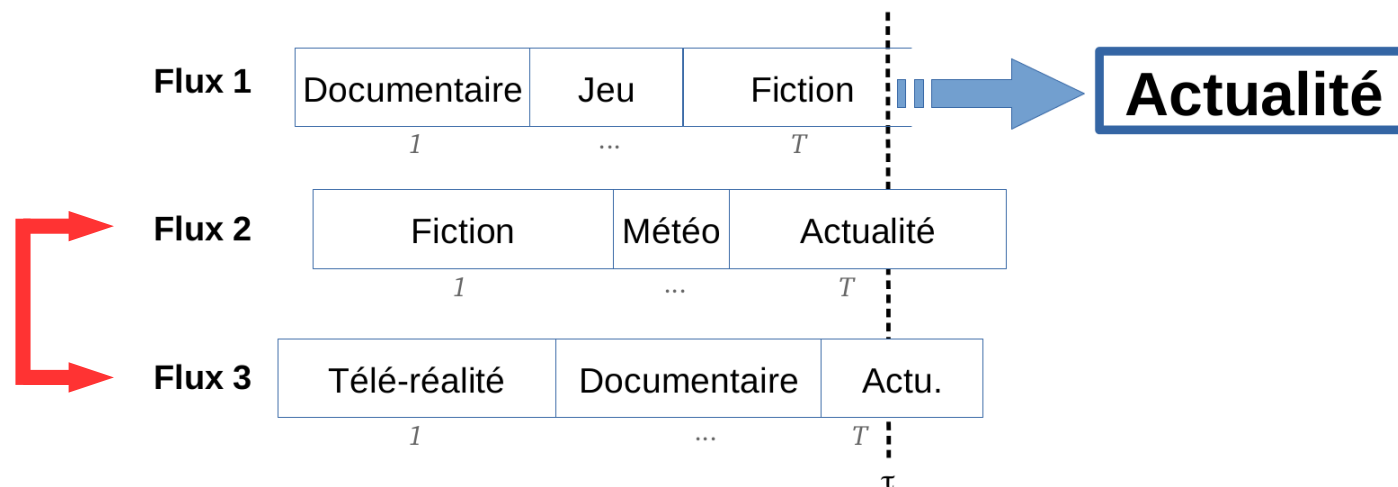
Prédiction du genre d'émission

- Style éditorial \Rightarrow chronologie des genres d'émission
- Utiliser l'historique des genres des T dernières émissions diffusées pour prédire le genre de l'émission suivante
- Séquences **monoflux** \Rightarrow classification de séquences



Prédiction du genre d'émission

- Relations entre les chaînes TV
 - séquençement de genres dans les chaînes parallèles : des informations supplémentaires ?
- Prédiction du genre dans le cadre **multiflux** : exploitation des séquences parallèles



Taxonomie de genres :

- ✗ Taxonomies existantes trop détaillées (INA, Médiamétrie...)
- Taxonomie proposée : 11 genres distincts entre eux
 - ▶ Actualité
 - ▶ Météo
 - ▶ Dessin animé
 - ▶ Fiction
 - ▶ Documentaire
 - ▶ Téléachat
 - ▶ Télé-réalité
 - ▶ Musique
 - ▶ Magazine
 - ▶ Jeux
 - ▶ Autres

Corpus d'historique de genres d'émission

- 4 chaînes françaises :
 - Chaîne principale : **M6** (généraliste)
 - Chaînes parallèles :
 - ▶ TF1 (généraliste)
 - ▶ France 5 et TV5 Monde (semi-thématiques)
- 3 années d'historique d'émissions :
 - 2013-2014 : **Apprentissage** (70%) + **Développement** (30%)
 - 2015 : **Test**
- Informations supplémentaires :
 - titre
 - horodatage (tranche horaire, jour de la semaine...)
 - résumé

II. Tâche et cadre expérimental

Corpus d'historique de genres d'émission

| Genres | Apprentissage | Développement | Test |
|--------------|---------------|---------------|-------------|
| Météo | 2691 | 1153 | 1683 |
| Fiction | 1890 | 810 | 1444 |
| Actualité | 913 | 392 | 663 |
| Magazine | 981 | 421 | 451 |
| Musique | 461 | 197 | 330 |
| Téléachat | 421 | 180 | 307 |
| Jeu | 476 | 204 | 284 |
| Dessin animé | 361 | 155 | 205 |
| Autres | 277 | 119 | 129 |
| Télé-réalité | 83 | 36 | 76 |
| Documentaire | 29 | 13 | 14 |
| Total | 8583 | 3680 | 5586 |



Distribution des genres d'émission pour M6

Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

IV. LSTM Parallèles (PLSTM)

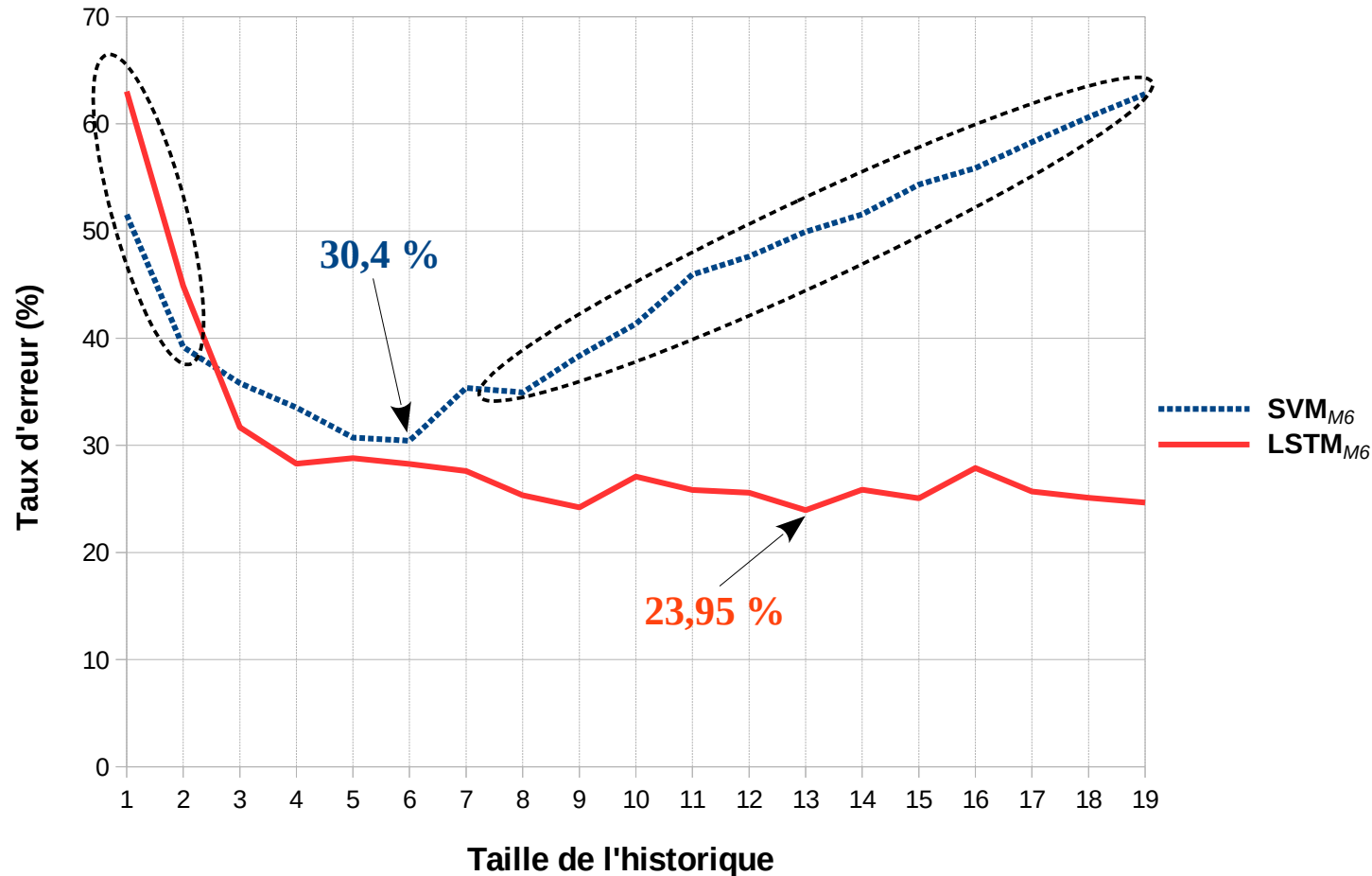
V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

VII. Conclusions et perspectives



III. Classification de séquences provenant d'un seul flux

LSTM vs SVM



- **LSTM** plus **efficace** et plus **stable** (21 % de gain relatif)
- moins efficace avec les séquences courtes

Bilan

| Méthode | Taux d'erreur (%) | Séquences longues |
|--------------------------|-------------------|---|
| SVM_{M6} | 30,4 |  |
| LSTM_{M6} | 23,95 |  |

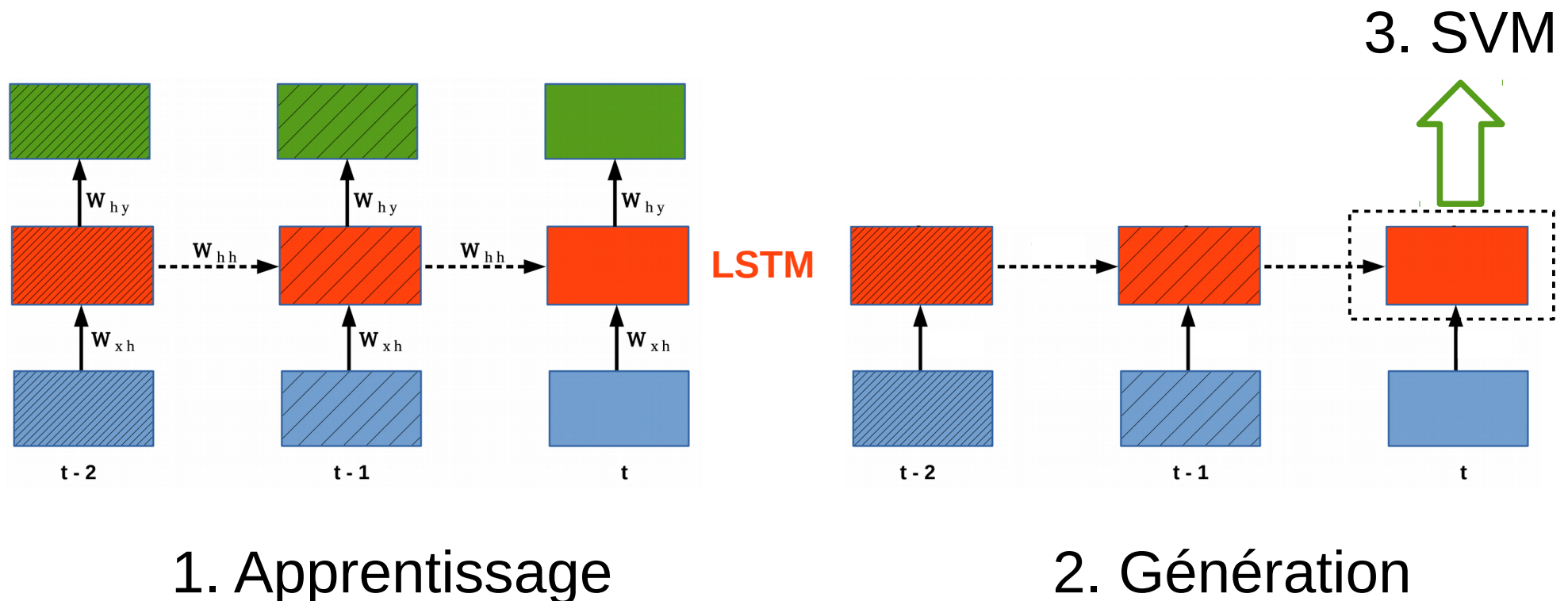
⇒ Méthodes **adaptées aux séquences** plus efficaces que les méthodes **classiques**

III. Classification de séquences provenant d'un seul flux

Représentations vectorielles de séquences + SVM (SE-SVM)

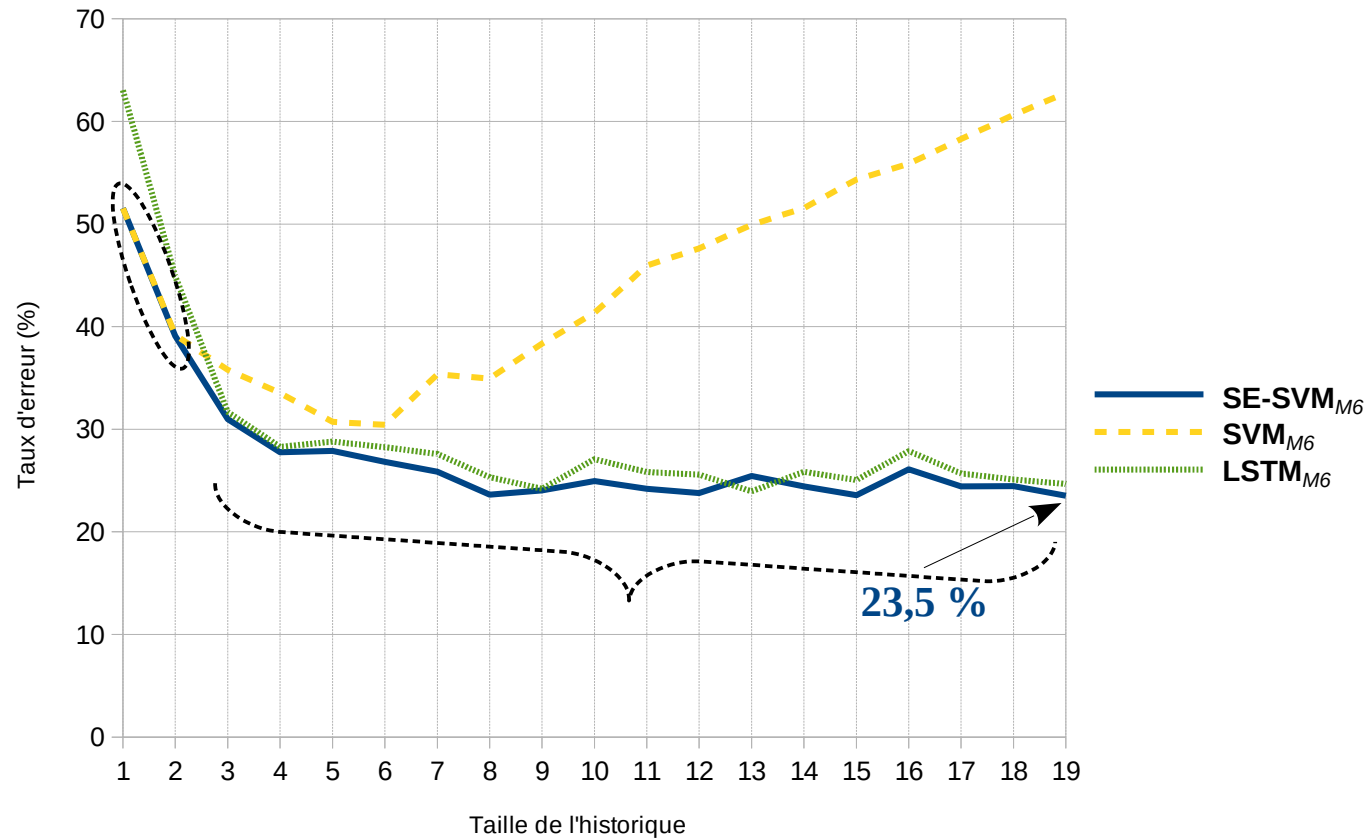
- Valeurs produites par la couche cachée : représentations latentes

⇒ **vecteurs de caractéristiques** utilisables par des méthodes classiques



III. Classification de séquences provenant d'un seul flux

Représentations vectorielles de séquences + SVM (SE-SVM)



→ Combinaison des avantages des **SVM** (historiques **courts**) et des **LSTM** (historiques **longs**)

Bilan

| Méthode | Taux d'erreur (%) | Séquences longues |
|-----------------------------|-------------------|-------------------|
| SVM _{M6} | 30,4 | ✗ |
| LSTM _{M6} | 23,95 | ✓ |
| SE-SVM _{M6} | 23,5 | ✓ |

Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

IV. LSTM Parallèles (PLSTM)

V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

VII. Conclusions et perspectives

Formulation théorique

■ Parallel RNN (PRNN)

$$h_t^N = \mathcal{H}(\mathbf{W}_{x^N h^N} x_t^N + \mathbf{W}_{h^N h^N} h_{t-1}^N + b_h^N)$$

.....

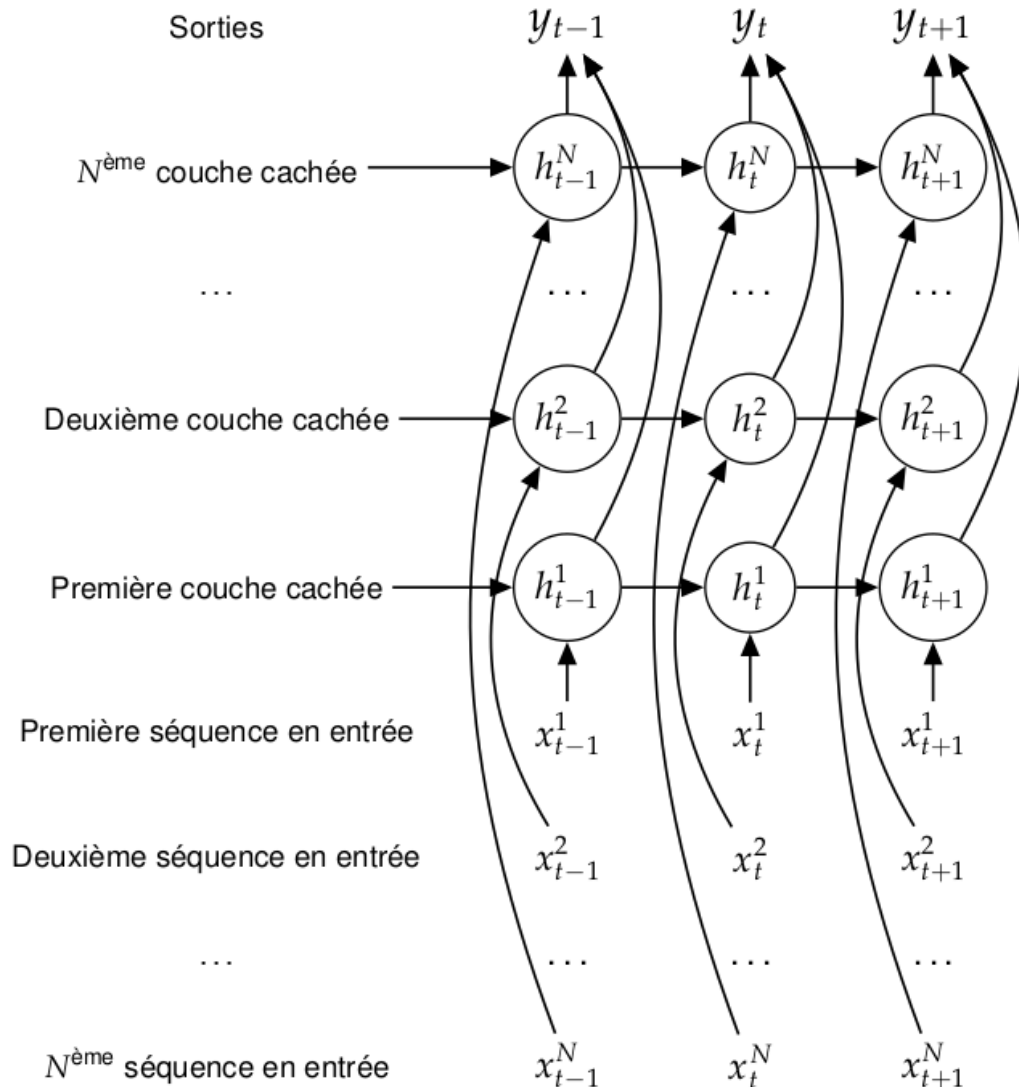
$$h_t^2 = \mathcal{H}(\mathbf{W}_{x^2 h^2} x_t^2 + \mathbf{W}_{h^2 h^2} h_{t-1}^2 + b_h^2)$$

$$h_t^1 = \mathcal{H}(\mathbf{W}_{x^1 h^1} x_t^1 + \mathbf{W}_{h^1 h^1} h_{t-1}^1 + b_h^1)$$

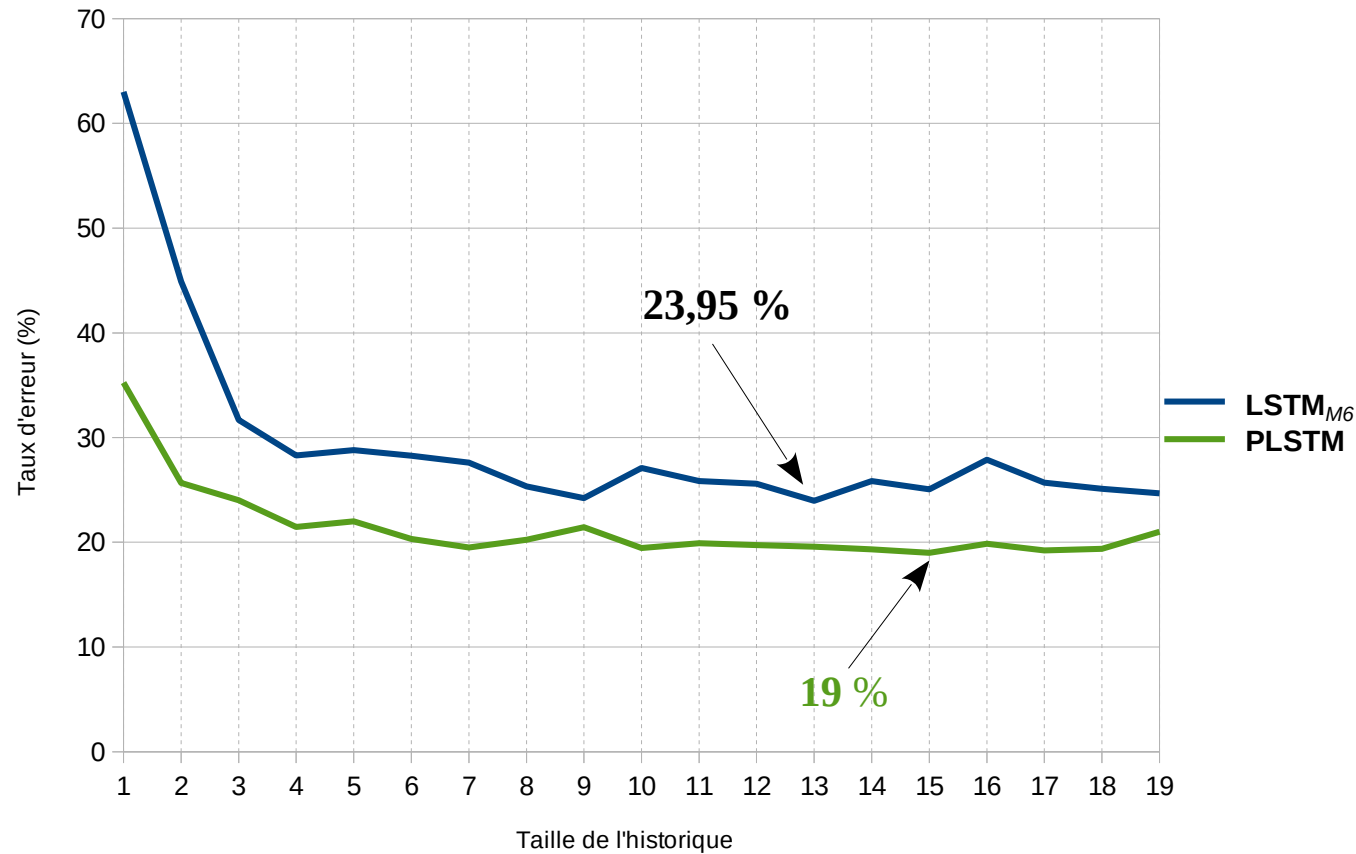
$$y_t = \sum_{n=1}^N \mathbf{W}_{h^n y} h_t^n + b_y$$

■ PLSTM = PRNN

en remplaçant H par la
fonction composite du LSTM



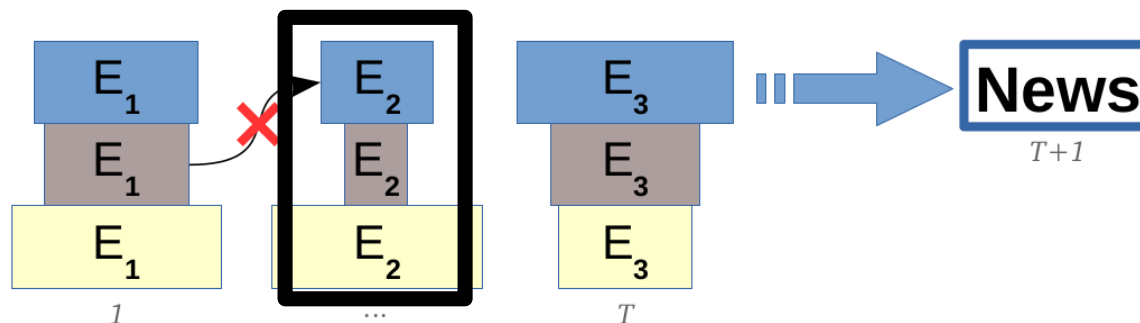
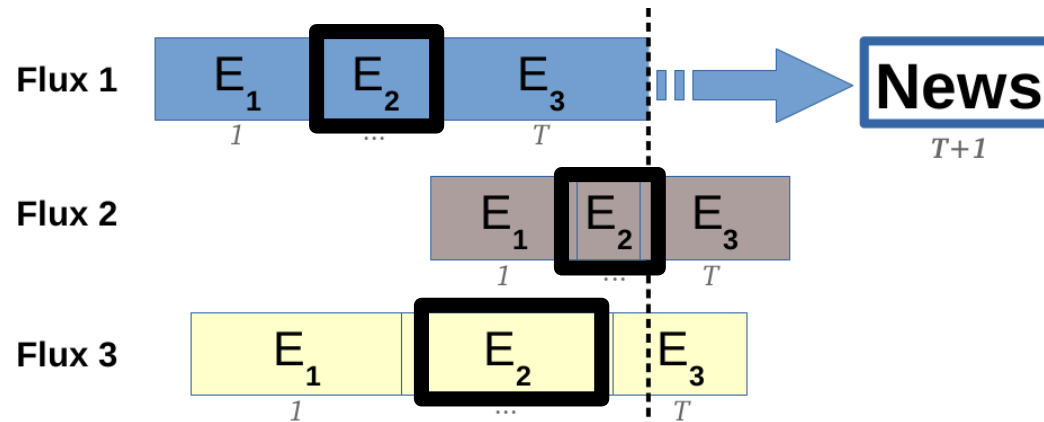
PLSTM vs LSTM



- **PLSTM** : **20 %** de gain relatif
- ⇒ efficace dans l'intégration des séquences parallèles.

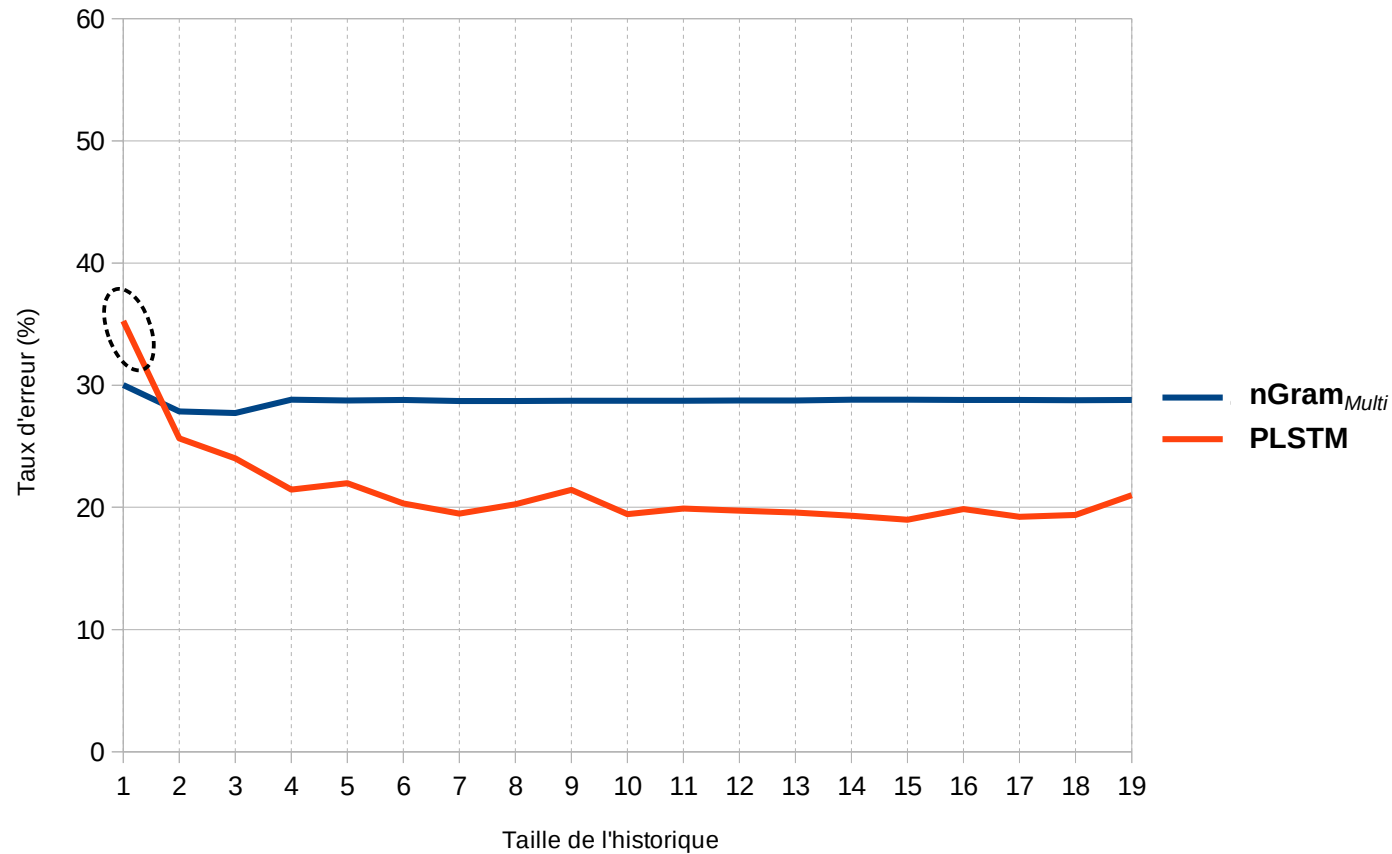
IV. LSTM Parallèles (PLSTM)

n-gramme : combinaison de séquences parallèles



IV. LSTM Parallèles (PLSTM)

PLSTM vs ngramme multiflux



- nGram_{Multi} : ne peut pas tirer profit des séquences longues
- ⇒ PLSTM : efficacité confirmée
- difficultés avec les historiques courts

Bilan

| Méthode | Taux d'erreur (%) | Séquences longues |
|------------------------------|-------------------|-------------------|
| SVM_{M6} | 30,4 | ✗ |
| LSTM_{M6} | 23,95 | ✓ |
| SE-SVM_{M6} | 23,5 | ✓ |
| nGram_{Multi} | 27,7 | ✗ |
| PLSTM | 19 | ✓ |

Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

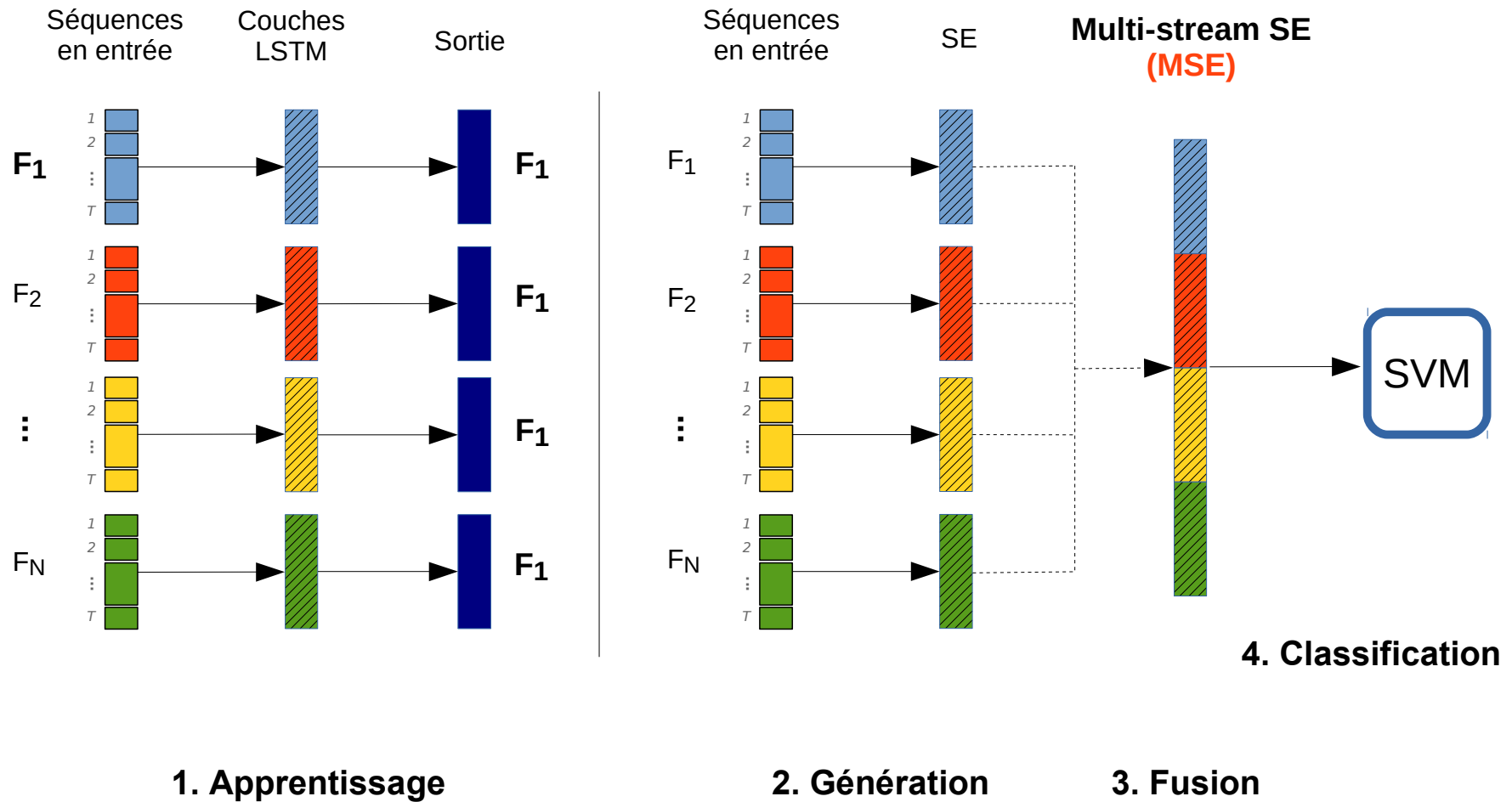
IV. LSTM Parallèles (PLSTM)

V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

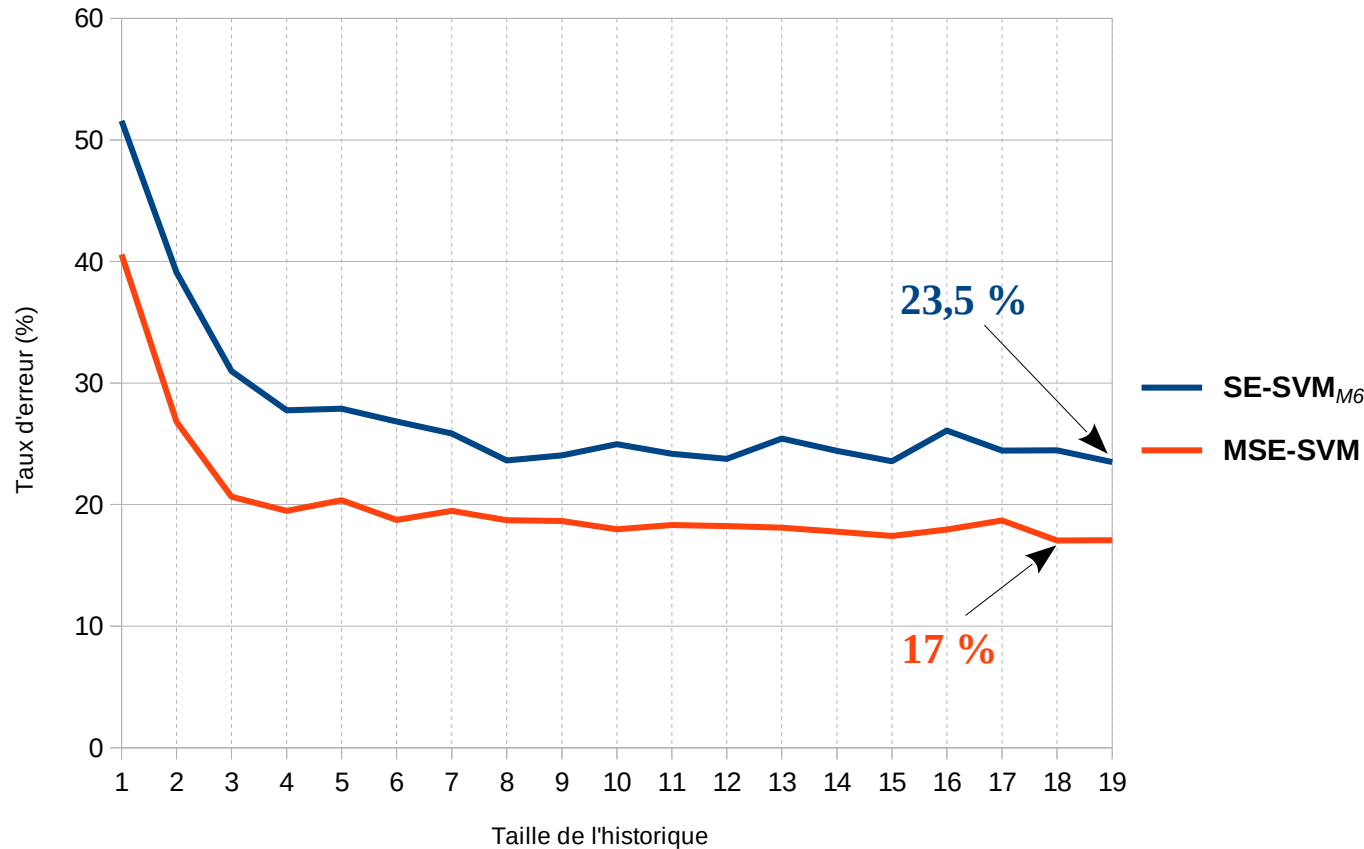
VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

VII. Conclusions et perspectives

Formulation théorique

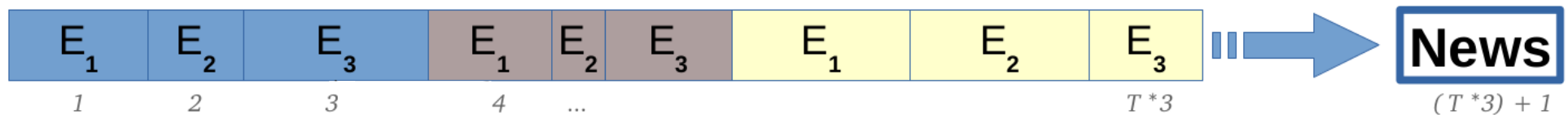
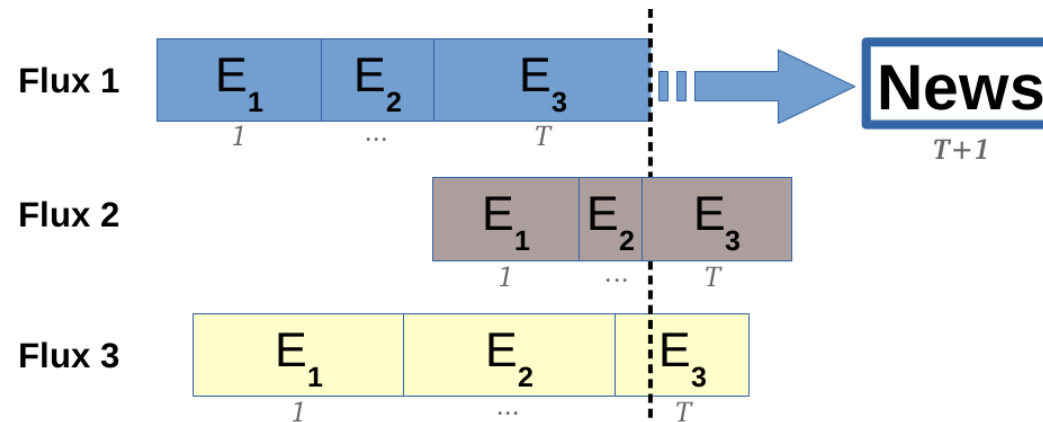


MSE-SVM vs SE-SVM

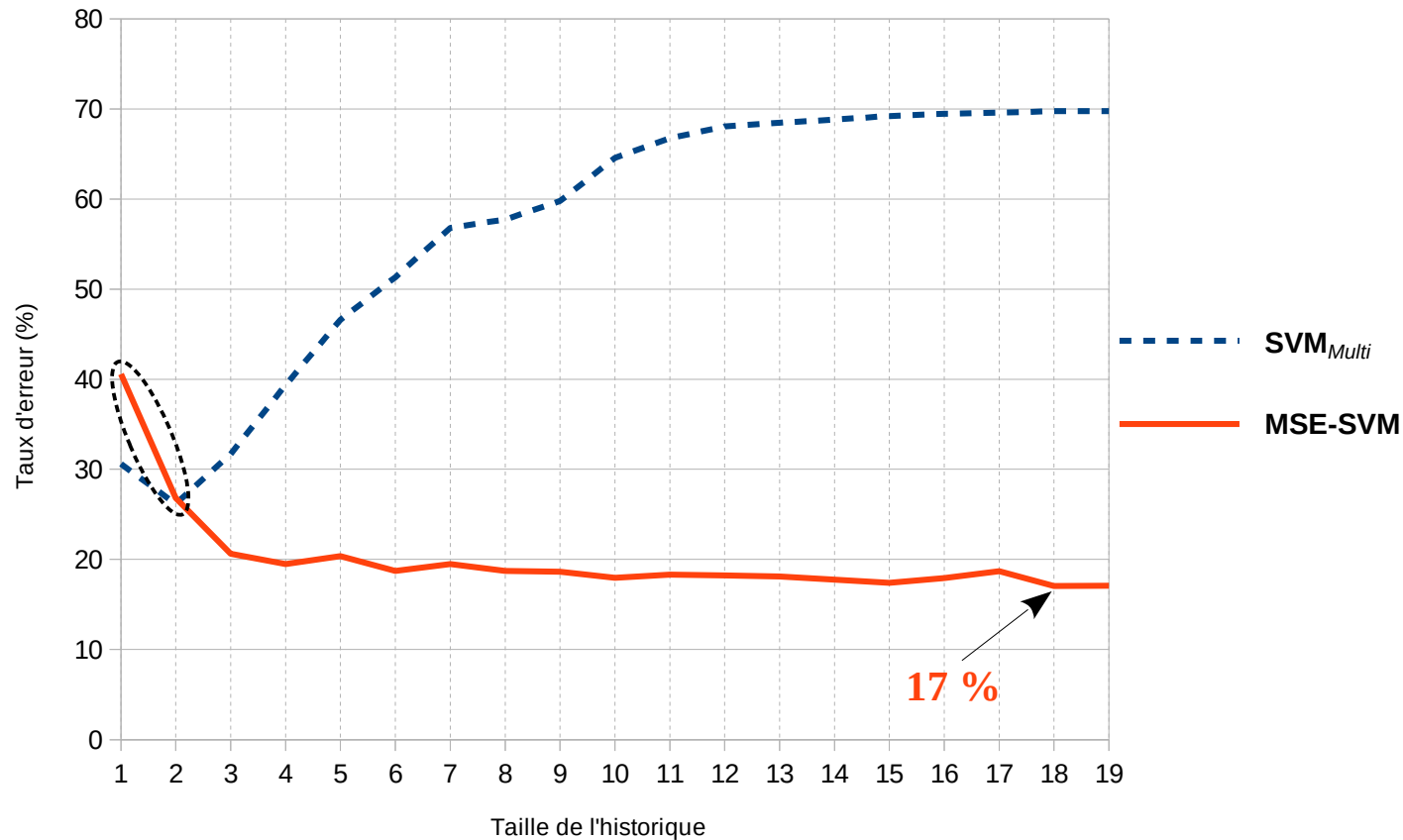


→ **MSE-SVM** : **27 %** de gain relatif
⇒ efficace dans l'intégration des séquences parallèles.

SVM : combinaison de séquences parallèles

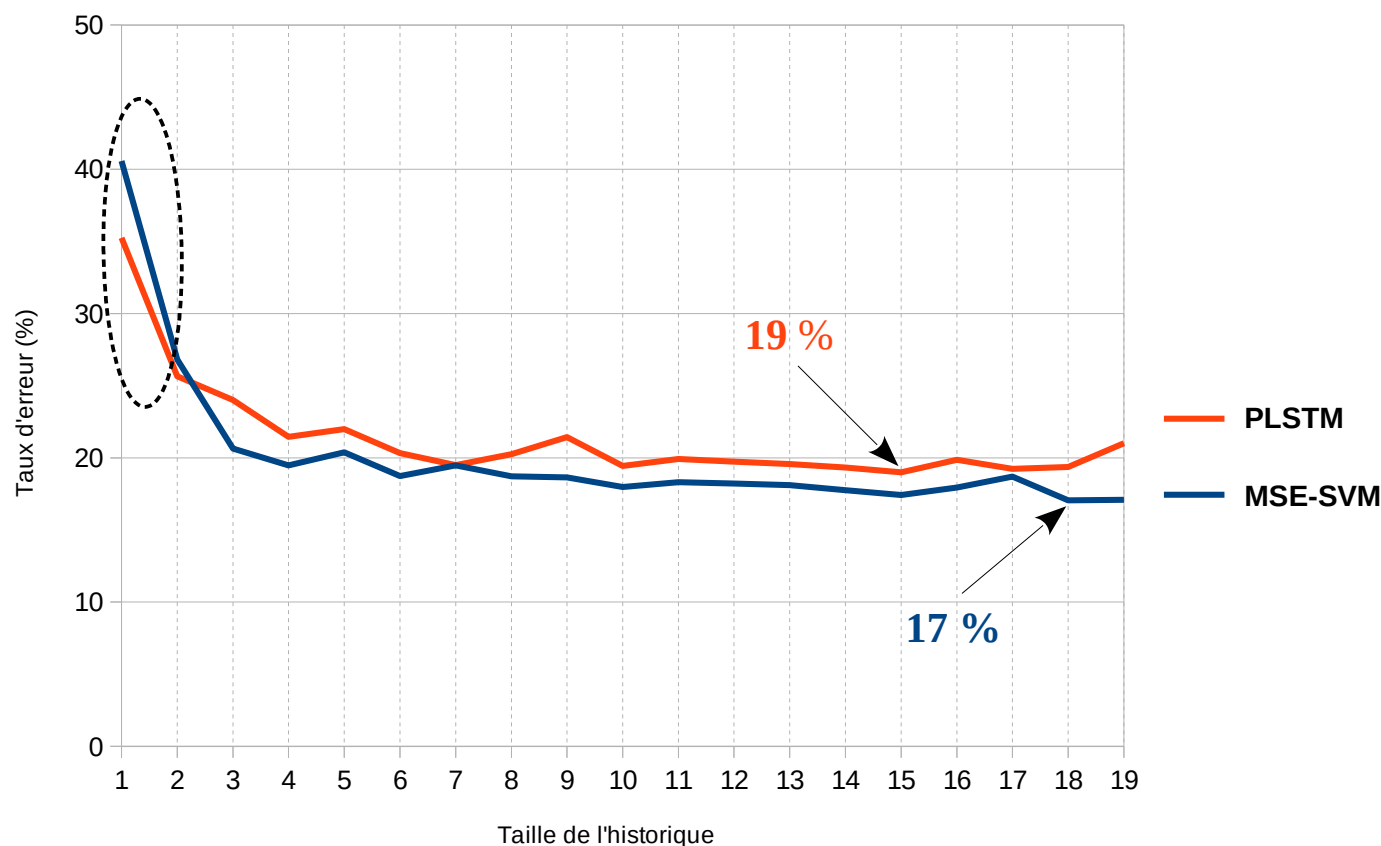


MSE-SVM vs SVM multiflux



- SVM_{multi} : ne peut pas exploiter les séquences longues
- ⇒ MSE-SVM : efficacité confirmée
- difficultés avec les historiques courts

MSE-SVM vs PLSTM



- **MSE-SVM** est légèrement meilleure (**10 %** de gain relatif)
- Pour les deux approches : **difficulté** avec les historiques **courts** (plus prononcée pour **MSE-SVM**)

Bilan

| Méthode | Taux d'erreur (%) | Séquences longues |
|-----------------|-------------------|-------------------|
| SVM_{M6} | 30,4 | ✗ |
| $LSTM_{M6}$ | 23,95 | ✓ |
| $SE-SVM_{M6}$ | 23,5 | ✓ |
| $nGram_{Multi}$ | 27,7 | ✗ |
| PLSTM | 19 | ✓ |
| SVM_{Multi} | 26,2 | ✗ |
| MSE-SVM | 17 | ✓ |



Faiblesse : séquences courtes (manque de connaissances)

Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

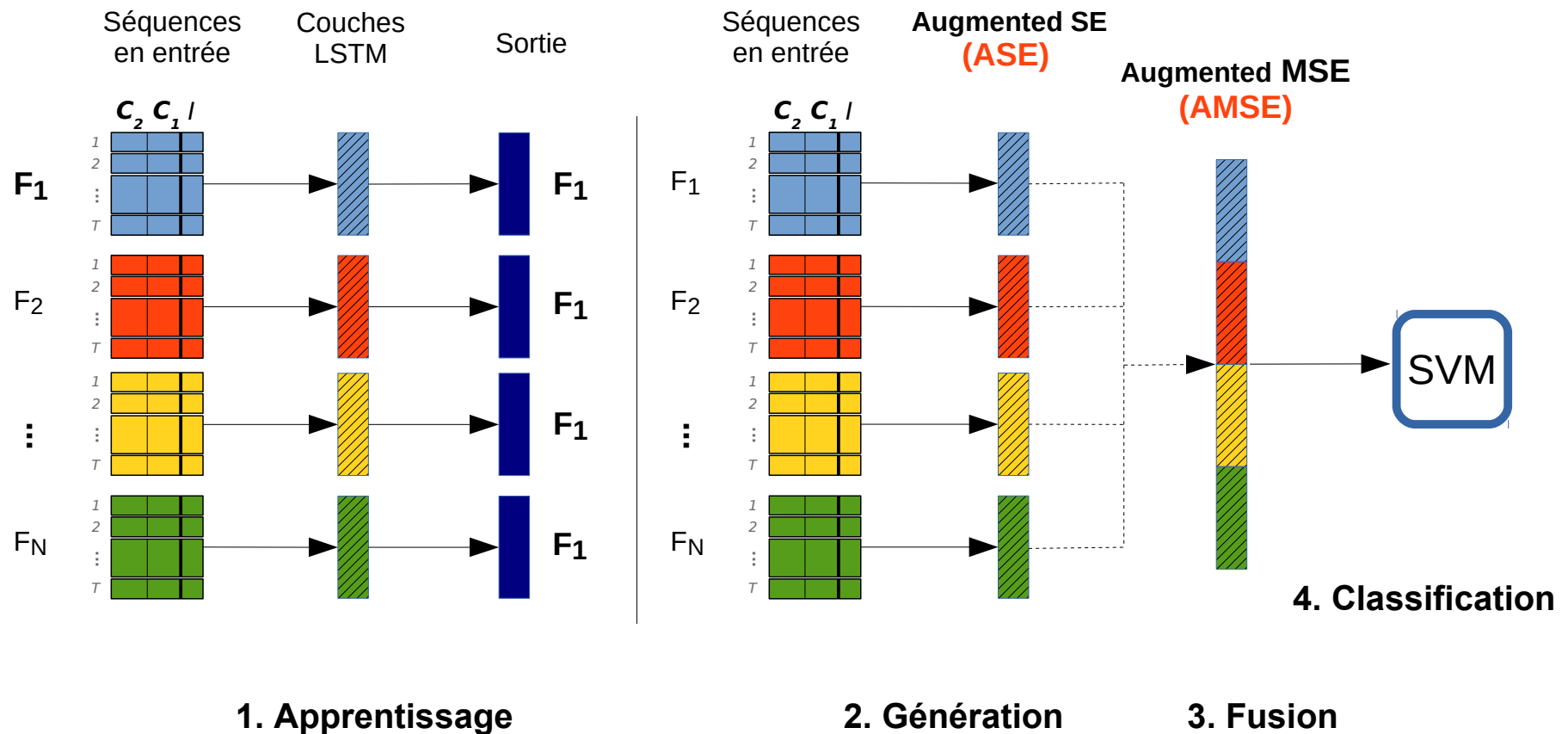
IV. LSTM Parallèles (PLSTM)

V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

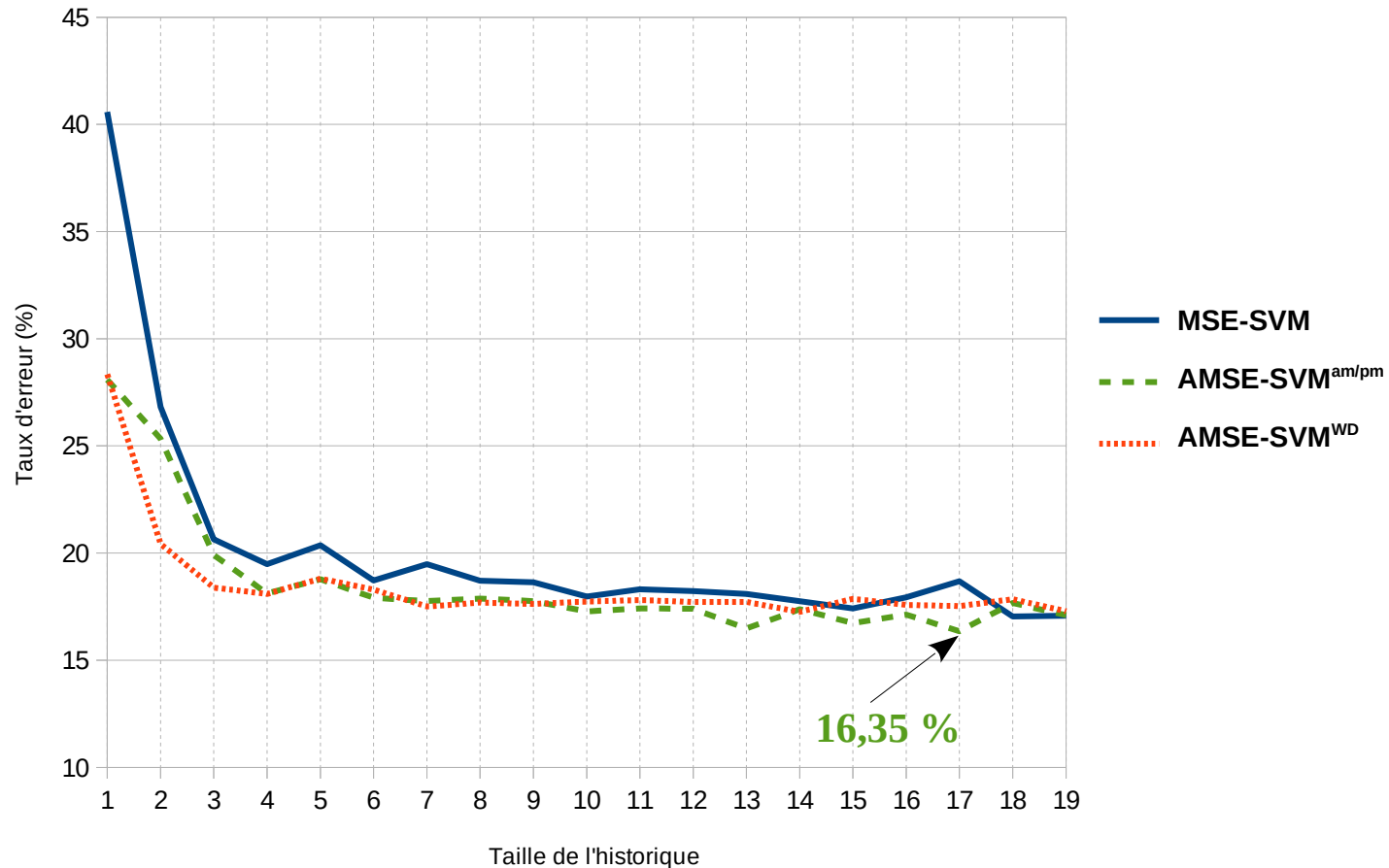
VII. Conclusions et perspectives

Formulation théorique



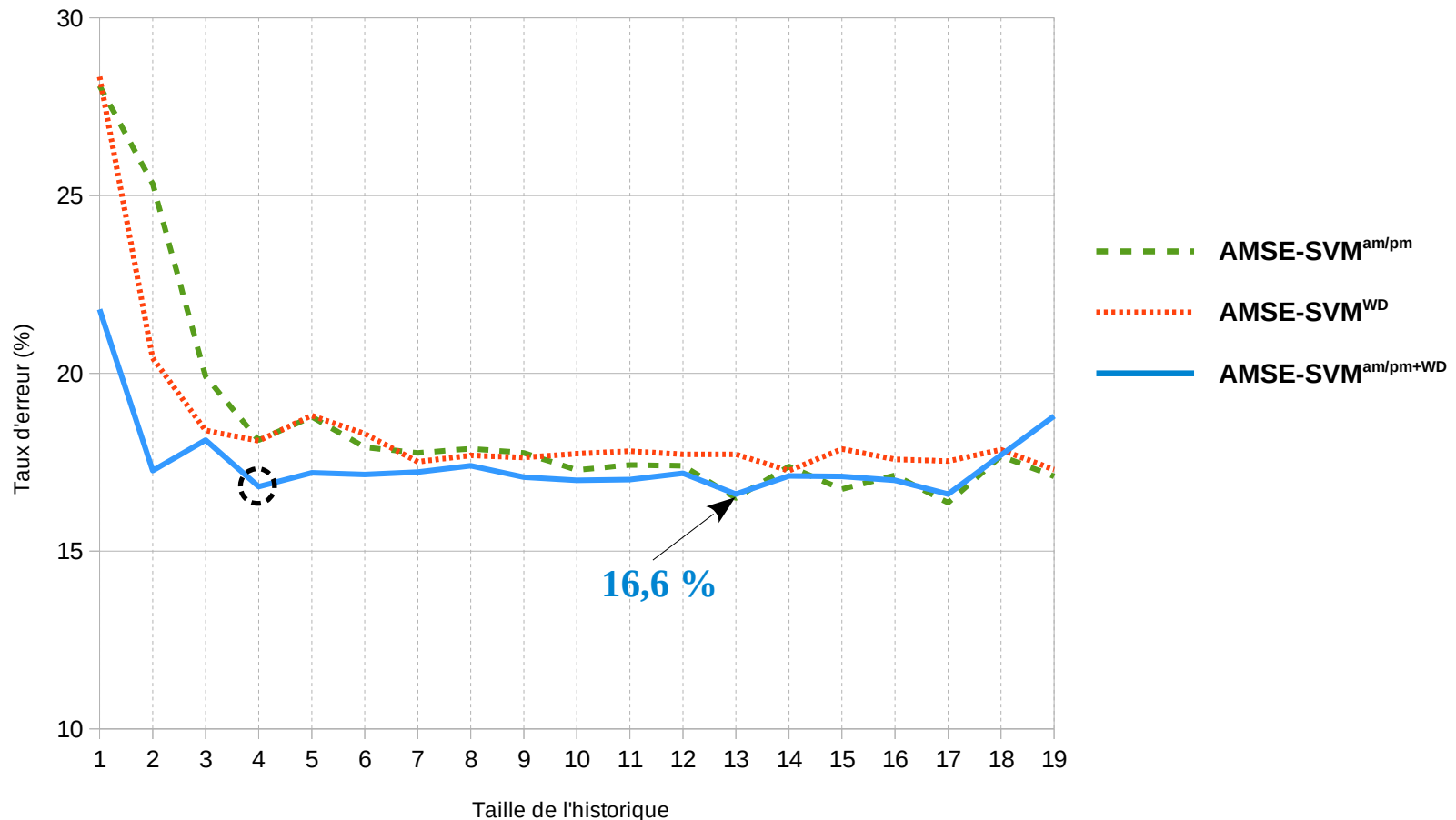
VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

Les AMSE unicontextuelles



- Amélioration plus visible avec les historiques courts (jusqu'à **30 %** de gain relatif)
- Peu d'amélioration avec les séquences longues (**4 %** de gain relatif)

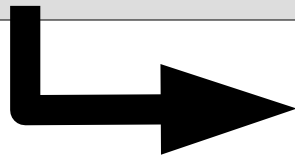
Les AMSE bicontextuelles



- Amélioration avec les historiques courts (jusqu'à **23 %** de gain relatif)
 - Stagnation précoce (pas d'amélioration avec les séquences longues)
- ⇒ les historiques longs incorporent déjà les informations contextuelles.

Bilan

| Méthode | Taux d'erreur (%) | Séquences longues |
|-----------------------|-------------------|-------------------|
| SVM_{M6} | 30,4 | ✗ |
| $LSTM_{M6}$ | 23,95 | ✓ |
| $SE-SVM_{M6}$ | 23,5 | ✓ |
| $nGram_{Multi}$ | 27,7 | ✗ |
| PLSTM | 19 | ✓ |
| SVM_{Multi} | 26,2 | ✗ |
| MSE-SVM | 17 | ✓ |
| $AMSE-SVM^{am/pm+WD}$ | 16,6 | ± |



Contribution : séquences courtes

Plan

I. État de l'art

II. Tâche et cadre expérimental

III. Classification de séquences provenant d'un seul flux

IV. LSTM Parallèles (PLSTM)

V. Représentations vectorielles de séquences parallèles + SVM (MSE-SVM)

VI. MSE-SVM : ajout d'informations issues du contexte (AMSE-SVM)

VII. Conclusions et perspectives

Conclusions

- **Motivation** : Exploitation des données séquentielles multiflux
- Conception d'un cadre expérimental adapté
- 2 propositions principales
 - **PLSTM** : extension du modèle LSTM
 - **MSE-SVM** : « combinaison » des modèles LSTM et SVM
- ✓ Efficacité dans l'intégration des séquences parallèles
- ✓ Capacité à tirer profit des séquences longues
- ✗ Difficultés avec les séquences courtes

Conclusions

- Solution : extension de l'approche MSE-SVM
⇒ **AMSE-SVM**
- ✓ Apport principal : séquences courtes
- ✗ Avec les informations contextuelles choisies :
pas d'amélioration sur les séquences longues

Perspectives

- Utilisation d'autres informations contextuelles : caractéristiques intrinsèques d'une émission (nom, durée, thème général...)
- Prédiction de la suite des genres des émissions suivantes : Encodeur-Décodeur multiflux ?
- Prédiction simultanée pour tous les flux donnés en entrée

Perspectives

- Conditionner les couches LSTM des flux parallèles selon la sortie de la couche LSTM du flux principal.
- Application dans d'autres contextes de prédiction d'événements
 - Prévisions météorologiques
 - Évolution des actions des sociétés cotées en bourse

Merci pour votre attention