

Utilisation de réseaux de confusion pour la reconnaissance de phrases manuscrites en-ligne

Using Confusion Networks to Recognize On-Line Handwritten Sentences

Solen Quiniou¹

Eric Anquetil¹

¹ IRISA - INSA

Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
Solen.Quiniou@irisa.fr

Résumé

Dans cet article, nous nous intéressons à l'intégration d'une représentation des hypothèses de phrases sous forme de réseau de confusion, dans un système de reconnaissance de phrases manuscrites en-ligne. Les probabilités a posteriori des mots, obtenues à partir du réseau de confusion, sont utilisées comme score de confiance afin de détecter d'éventuelles erreurs dans la phrase issue d'un décodage au Maximum A Posteriori sur un graphe de mots. Des classificateurs dédiés (ici, des SVM) sont ensuite appris afin de corriger ces erreurs, en combinant les probabilités a posteriori des mots à d'autres sources de connaissance. Une phase de rejet est aussi introduite dans le processus de détection. Des expérimentations menées sur une base de 320 phrases manuscrites montrent une réduction relative du taux d'erreur sur les mots de 31,3 %, dans le cas de l'extraction manuelle des mots, et une diminution relative de 60 %, lorsque ces mots sont extraits automatiquement.

Mots Clef

Reconnaissance de phrases manuscrites, réseaux de confusion, graphes de mots, algorithme de Viterbi.

Abstract

In this paper we investigate the integration of a representation of sentence hypotheses thanks to a confusion network into an on-line handwritten sentence recognition system. The word posterior probabilities from the confusion network are used as confidence scores to detect potential errors in the output sentence from the Maximum A Posteriori decoding on a word graph. Dedicated classifiers (here, SVMs) are then trained to correct these errors and combine the word posterior probabilities with other sources of knowledge. A rejection phase is also introduced in the detection process. Experiments carried out on 320 handwritten sentences show a 31.3 % relative reduction of the word

error rate when words are manually extracted and a 60 % relative diminution when the words are automatically extracted.

Keywords

Handwritten sentence recognition, confusion networks, word graphs, Viterbi algorithm.

1 Introduction

La plupart des systèmes de reconnaissance de phrases manuscrites utilise des graphes de mots pour représenter différentes hypothèses de phrases [14, 9]. Le décodage standard MAP (Maximum A Posteriori) est alors utilisé pour trouver la phrase ayant la probabilité la plus élevée, en utilisant des informations fournies par le système de reconnaissance ainsi que des informations linguistiques représentées par un modèle de langage. Alors que cette approche MAP vise à minimiser le taux d'erreur sur les phrases, la métrique la plus utilisée pour mesurer les performances d'un système de reconnaissance reste le taux d'erreur sur les mots : il y a ainsi un problème de correspondance entre l'approche utilisée pour la reconnaissance et l'objectif consistant à maximiser le nombre de mots reconnus. Une nouvelle approche, visant à minimiser le taux d'erreur sur les mots, est alors apparue en reconnaissance de parole : elle s'appuie sur des réseaux de confusion [8, 15].

Les réseaux de confusion ont été introduit dans [8]. Ces réseaux sont utilisés pour représenter un ensemble de phrases et s'appuient pour cela sur les probabilités a posteriori des mots. La probabilité a posteriori d'un mot correspond à la somme des probabilités de tous les chemins auxquels ce mot appartient. Ces probabilités a posteriori peuvent ensuite être utilisées pour retrouver la meilleure hypothèse de phrase ou encore en tant qu'indice de confiance sur ses mots [3]. Cette mesure de

confiance peut aussi être combinée à d'autres sources de connaissance dans un réseau de neurones [7] ou dans un SVM [6], pour constituer un meilleur indice de confiance, par exemple.

Dans cet article, nous intégrons une représentation des hypothèses de phrases sous forme de réseau de confusion, dans notre système de reconnaissance de phrases manuscrites, afin d'améliorer ses performances. À notre connaissance, les réseaux de confusion n'ont encore jamais été utilisés pour la reconnaissance de phrases manuscrites. Nous adaptons aussi ces réseaux de confusion afin de pouvoir prendre en compte plusieurs hypothèses de segmentation. Le réseau de confusion est utilisé conjointement au décodage MAP, déjà utilisé dans notre système de reconnaissance [10], pour détecter d'éventuelles erreurs. Les probabilités *a posteriori* des mots ainsi que d'autres sources de connaissance sont ensuite utilisées pour corriger ces erreurs. Nous avons introduit cette nouvelle approche dans [11] sur des phrases segmentées manuellement. Nous l'étendons dans cet article à la segmentation automatique de phrases en mots, en intégrant notamment la gestion des problèmes de sur- et sous segmentations. De plus, une stratégie de rejet est introduite afin d'identifier les mots qui ne pourront pas être corrigés par l'approche présentée. Cette stratégie de rejet pourrait ainsi permettre, d'une part, la mise en évidence de ces mots non reconnus dans une interface de saisie (afin de faciliter leur correction, pour l'utilisateur) et, d'autre part, pourrait donner lieu à une étape supplémentaire de reconnaissance de ces mots rejetés.

La suite de cet article est décomposée de la façon suivante. Le principe du système de reconnaissance de phrases manuscrites est donné dans la section 2. Les deux approches pour la reconnaissance de phrases sont ensuite détaillées dans les sections 3 et 4, en se focalisant plus particulièrement sur les réseaux de confusion. L'approche proposée pour détecter et corriger les erreurs est ensuite présentée dans la section 5. Les résultats expérimentaux sont enfin discutés dans la section 6 et la section 7 tire les conclusions.

2 Système de reconnaissance de phrases manuscrites en-ligne

Le système de reconnaissance de phrases illustré par la figure 1 étend notre système de reconnaissance introduit dans [10].

Le premier module permet l'extraction des parties du signal de la phrase manuscrite correspondant à chacun de ses mots. Cette étape peut être réalisée manuellement lorsque l'on souhaite analyser l'impact des étapes suivantes, indépendamment des possibles erreurs d'extraction des mots. Cette opération peut aussi se faire automatiquement. Dans notre système, l'extraction automatique [12] se base sur la caractérisation des espaces entre couple de traces manuscrites consécutives. Nous appelons *trace* une liste de points ordonnés chronologiquement et capturés entre un posé et

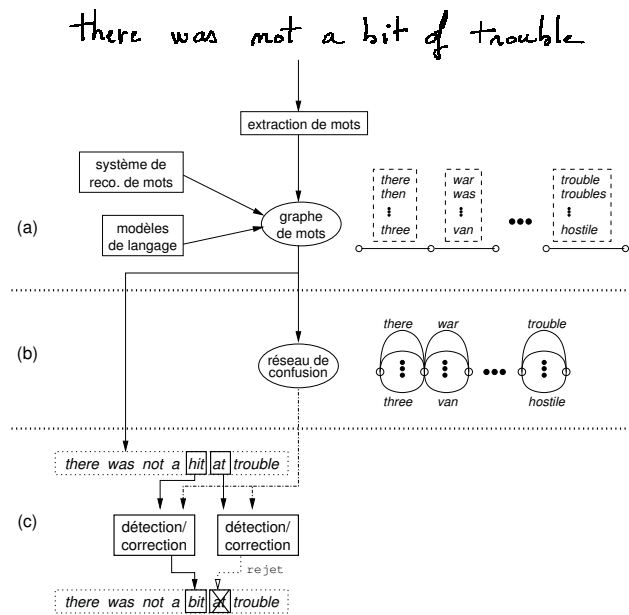


FIG. 1 – Système de reconnaissance de phrases.

un levé de stylo. Un classifieur est ensuite appris (ici, un réseau de neurones à fonction à base radiale) afin d'attribuer, à chaque espace entre traces, une classe parmi les trois suivantes : espace intra-mot (entre deux traces d'un même mot), espace inter-mot (entre deux traces de deux mots consécutifs) et espace inter-ligne (entre deux traces de deux mots écrits sur deux lignes consécutives). Un indice de confiance est aussi associé au résultat du réseau RBFN afin d'évaluer la fiabilité de sa première réponse. Cela permet ainsi de considérer des hypothèses supplémentaires d'extraction de mots lorsque cet indice de confiance n'est pas suffisamment élevé, afin de pouvoir gérer d'éventuelles sur- ou sous-segmentations.

Les hypothèses de mots extraites lors de cette première étape permettent la construction du graphe de mots. Les noeuds de ce graphe représentent les frontières entre deux mots consécutifs et les arcs représentent les mots extraits. Notre système de reconnaissance de mots isolés RESIF-Mot [1] est utilisé pour reconnaître chacun de ces mots manuscrits et donne, pour chacun d'eux, une liste ordonnée de 10 mots candidats (comme illustré en partie (a) de la figure 1). Deux scores, donnés par ce système de reconnaissance de mots, sont ainsi associés à chacun des mots de cette liste : un *score lexical* (dépendant des opérations d'édition durant l'étape de post-traitement lexical [2]) et un *score graphique* (combinant des mesures d'adéquation entre chacun des caractères et son modèle de lettre ainsi que des informations spatiales et statistiques entre les caractères du mot). Un modèle de langage est aussi utilisé dans ce graphe afin de donner des probabilités aux séquences de mots, constituées de mots de listes consécutives. La reconnaissance de phrases peut alors être effec-

tuée sur ce graphe de mots, en utilisant le décodage MAP présenté à la section 3.

La partie (b) de la figure 1 montre le réseau de confusion obtenu à partir du graphe de mots (voir section 4). La partie (c) de la figure 1 illustre enfin l'approche de détection et correction des erreurs. Les probabilités *a posteriori* associées à chacun des mots des listes de mots candidats et calculées grâce au réseau de confusion, sont utilisées pour mesurer la confiance de chacun des mots de la meilleure hypothèse de phrase obtenue par le décodage MAP. Quand une erreur est détectée, une étape de correction permet de retrouver le mot correct, en utilisant la probabilité *a posteriori* de ce mot ainsi que d'autres sources d'informations (voir section 5).

Après avoir présenté notre système de reconnaissance de phrases, nous détaillons, dans les deux sections suivantes, la reconnaissance en utilisant le décodage MAP sur un graphe de mots ainsi que la reconnaissance par consensus sur un réseau de confusion.

3 Graphe de mots et décodage MAP

Le décodage standard MAP (Maximum A Posteriori) a pour but de trouver la phrase la plus probable \hat{W} parmi des hypothèses de phrases $W_k = w_{k,1} \dots w_{k,n_k}$ étant donné un signal S (correspondant à la phrase manuscrite à reconnaître), en combinant des informations graphiques et linguistiques, comme donné par l'équation 1 :

$$\hat{W} = \arg \max_{W_k} \text{score}(S|W_k) + \gamma \log [p(W_k)] + \delta n_k \quad (1)$$

avec $\text{score}(S|W_k)$ le score du signal S estimé par le système de reconnaissance (combinant les scores lexicaux et graphiques des mots) et $p(W_k)$ la probabilité *a priori* de la séquence de mots W_k , donnée par un modèle de langage statistique. Le poids γ associé au modèle de langage (appelé aussi *Grammar Scale Factor*) relativise l'impact du modèle de langage statistique alors que le coefficient δ (aussi nommé *Word Insertion Penalty*) permet le contrôle des délétions et insertions de mots (n_k étant le nombre de mots de la phrase W_k).

Ce décodage est réalisé sur le graphe de mots grâce à l'algorithme de Viterbi [4].

4 Réseau de confusion et consensus

4.1 Motivation

Alors que l'approche standard MAP présentée dans la section précédente vise à trouver la phrase ayant la probabilité la plus élevée, la métrique la plus utilisée pour mesurer les performances d'un système de reconnaissance est le taux d'erreur sur les mots. Il y a ainsi un problème de correspondance entre l'approche utilisée pour la reconnaissance et l'objectif consistant à maximiser le nombre de mots reconnus. Contrairement au décodage MAP, l'approche utilisant des réseaux de confusion cherche à maximiser les

probabilités *a posteriori* des mots. Les noeuds d'un réseau de confusion représentent alors des classes d'équivalence (appelées *ensembles de confusion*) c'est-à-dire des confusions entre hypothèses de mots à une position donnée dans la phrase. Les noeuds adjacents sont reliés par autant d'arcs que d'hypothèses de mots (voir figure 2).

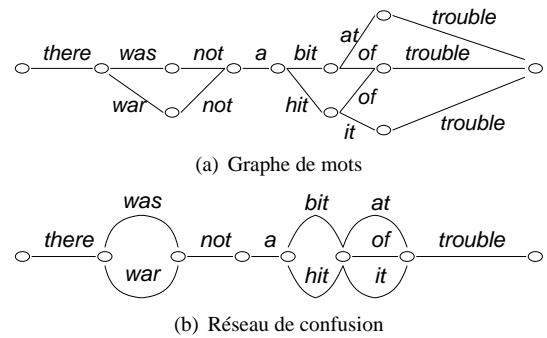


FIG. 2 – Graphe de mots et son réseau de confusion correspondant.

4.2 Algorithme du consensus

Un réseau de confusion est construit en alignant tout d'abord les hypothèses de phrases issues du graphe de mots. Les probabilités *a posteriori* sont alors calculées pour chacun des arcs du graphe. Les classes d'équivalence sont ensuite initialisées de sorte que chaque classe ne contienne que les liens ayant les mêmes temps de début et de fin ainsi que les mêmes étiquettes de mot (une étape d'élagage des arcs peut être réalisée au préalable, en supprimant les arcs dont la probabilité *a posteriori* est trop faible). Le regroupement intra-mot a alors pour but de fusionner les classes ayant la même étiquette de mot mais ayant des temps de début et/ou de fin se chevauchant. Le regroupement inter-mot intervient enfin pour fusionner les classes correspondant à des mots différents, en se basant essentiellement sur leur similarité graphique.

Nous obtenons ainsi le réseau de confusion, comme illustré par la figure 1 (b). L'hypothèse correspondant au taux d'erreur minimal sur les mots peut alors être obtenue en prenant le mot ayant la probabilité *a posteriori* la plus élevée, dans chaque ensemble de confusion, à chaque position de la phrase. Cette hypothèse est appelée *hypothèse consensus*.

Probabilités *a posteriori* des mots. La probabilité *a posteriori* d'un mot est la somme des probabilités *a posteriori* de tous les liens dont les étiquettes correspondent au mot considéré. La probabilité du k^e lien du mot $w_{i,j}$ peut être calculée efficacement en utilisant l'algorithme *forward-backward* et est donnée par l'équation 2 :

$$p(w_{i,j}^{(k)}) = \frac{\alpha(n_{i-1,k}) p(w_{i,j}|w_{i-1,k})^\gamma \text{score}(w_{i+1,k})^\delta \beta(n_{i,j})}{\sum_l \sum_m \sum_r p(w_{l,m}^{(r)})} \quad (2)$$

avec $w_{i,j}^{(k)}$ le lien correspondant à la k^e hypothèse du mot $w_{i,j}$, $\alpha(n_{i-1,k})$ la probabilité *forward* du noeud $n_{i-1,k}$, $\beta(n_{i,j})$ la probabilité *backward* du noeud $n_{i,j}$, $p(w_{i,j}|w_{i-1,k})$ la probabilité du bigramme $w_{i-1,k}w_{i,j}$ (donnée par le modèle de langage) et $score(w_{i+1,k})$ le score donné par notre système de reconnaissance de mots au mot $w_{i+1,k}$ et qui combine les scores lexical et graphique présentés dans la section 2 (comme notre système de reconnaissance n'est pas probabiliste, ce score est ramené dans l'intervalle [0 :1]). La probabilité *forward* du noeud $n_{i,k}$ correspond à la somme des probabilités de tous les chemins entre le noeud initial du graphe et ce noeud, alors que la probabilité *backward* correspond à la somme des probabilités de tous les chemins entre ce noeud $n_{i,k}$ et le noeud final. Les poids γ et δ sont utilisés pour pondérer la probabilité du modèle de langage et le score de reconnaissance, respectivement, et sont optimisés sur un ensemble de validation.

La probabilité *a posteriori* d'un mot est donnée par l'équation 3 :

$$p(w_{i,j}) = \sum_k p(w_{i,j}^{(k)}). \quad (3)$$

Dans le cas de l'extraction manuelle, les ensembles de confusion s'obtiennent facilement puisqu'ils correspondent chacun à la liste de mots candidats, pour chacun des arcs du graphe de mots. Lorsque l'extraction des mots est réalisée automatiquement, plusieurs hypothèses d'extraction de mots peuvent être considérées afin de parer aux éventuelles sur- et sous-segmentations des mots. Nous étendons ainsi l'algorithme du consensus, pour prendre en compte ces nouvelles segmentations. Nous présentons cette extension dans la section suivante.

4.3 Extension pour la gestion des sur- et sous-segmentations

Lors de l'extraction automatique des mots des phrases, nous prenons en compte plusieurs hypothèses d'extraction des mots, afin de pallier aux sur- et sous-segmentations qui peuvent se produire. Nous avons ainsi à aligner un mot avec plusieurs mots, lors de la constitution des ensembles de confusion, comme illustré par la figure 3.

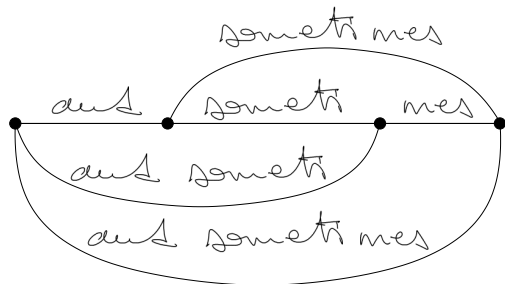


FIG. 3 – Exemple d'ensemble de confusion.

Les ensembles de confusion sont ainsi entièrement déterminés par la phase d'extraction des mots et correspondent maintenant à l'ensemble des arcs entre deux noeuds de telle façon qu'il n'existe aucun arc entre un des noeuds d'un ensemble de confusion donné et un des noeuds de l'ensemble de confusion précédent ou suivant.

Cette modification des ensembles de confusion entraîne des modifications dans le calcul des probabilités *a priori* des arcs ainsi que dans le calcul de l'hypothèse consensus. En effet, lors du calcul des probabilités *a priori* des arcs, il faut maintenant prendre en compte le fait qu'un arc peut avoir plusieurs arcs prédécesseurs et plusieurs arcs successeurs. Nous avons ainsi étendu le calcul des probabilités *forward* et *backward*, afin de prendre cela en compte. Enfin, le choix du mot dans chaque ensemble de confusion, afin de constituer l'hypothèse consensus, est lui aussi étendu au choix du meilleur chemin dans chacun des ensembles de confusion. La probabilité de chacun de ces chemins est alors le produit des probabilités *a posteriori* de chacun de ses mots.

Nous avons présenté deux approches pour la reconnaissance de phrases. L'approche basée sur le consensus s'appuie sur le calcul des probabilités *a posteriori* des mots du réseau. Ces probabilités peuvent aussi être utilisées comme indice de confiance sur les mots. Dans la section suivante, nous décrivons notre approche utilisant les probabilités *a posteriori* des mots comme indice de confiance sur les mots de la phrase résultat de la reconnaissance par approche MAP, afin de détecter et corriger des erreurs de reconnaissance.

5 Détection et correction d'erreurs grâce aux probabilités *a posteriori*

Les étapes de détection et de correction éventuelle sont illustrées dans la partie (c) de la figure 1. Le réseau de confusion est utilisé pour mesurer la confiance de chacun des mots de la meilleure hypothèse de phrase obtenue par l'approche MAP, sur le graphe de mots. Quand une erreur est détectée, une étape de correction vise à trouver le mot correct, en utilisant sa probabilité *a posteriori* ainsi que d'autres connaissances.

Dans les sous-sections suivantes, nous décrivons notre approche pour détecter et corriger les mots reconnus. Deux types d'erreurs sont ainsi considérées et, pour chacun de ces types, un classifieur dédié est appris afin de corriger le type d'erreur associé. Une stratégie de rejet est finalement présentée dans la dernière sous-section.

5.1 Non-correspondance avec le mot ayant la probabilité *a posteriori* la plus élevée

Pour la détection de ce premier type d'erreur, le mot considéré est détecté comme potentiellement non reconnu si sa probabilité *a posteriori* n'est pas la plus élevée dans son ensemble de confusion correspondant. En d'autres termes,

le mot de la phrase obtenue par l'approche MAP et celui de l'hypothèse consensus sont différents l'un de l'autre.

Afin de corriger ces erreurs potentielles, un classifieur dédié est appris pour choisir entre ces deux mots : le mot reconnu et celui ayant la probabilité *a posteriori* la plus élevée. Trois caractéristiques sont considérées pour chacun de ces mots, à savoir sa probabilité *a posteriori* ainsi que ses scores lexical et graphique normalisés (voir la section 2 pour la description de ces derniers scores).

5.2 Mot ayant une probabilité *a posteriori* non fiable

Le deuxième type d'erreurs détectées concerne les mots dont la probabilité *a posteriori* est la plus élevée dans leur ensemble de confusion mais cette probabilité est inférieure à un certain seuil (ici, 0,8) : elle est ainsi considérée comme non fiable.

Pour la phase de correction correspondante, un classifieur dédié est appris pour discriminer les deux meilleurs mots de l'ensemble de confusion (classés selon leurs probabilités *a posteriori* décroissantes). Les mêmes caractéristiques que précédemment sont utilisées pour chacun de ces deux mots.

5.3 Rejet des mots absents

Dans les deux stratégies de correction présentées, nous supposons que le mot correct se trouve parmi les deux mots donnés en entrée de chacun des classifieurs dédiés. En fait, le mot correct peut soit être parmi les autres mots de l'ensemble de confusion, soit ne pas être dans l'ensemble de confusion. Cette simplification a été faite car la majorité des mots corrects se trouve parmi ces deux mots. Elle est aussi effectuée pour palier à un problème de manque de données puisqu'en considérant plus de mots en entrée des classifieurs dédiés, il est nécessaire d'avoir plus de données d'apprentissage afin d'obtenir des classifieurs fiables.

La stratégie de rejet proposée est donc utilisée pour discriminer les mots pour lesquels le mot correct se trouve parmi les deux mots considérés, des mots pour lesquels le mot correct n'est aucun des deux. Un nouveau classifieur avec 6 entrées (correspondant aux trois caractéristiques précédentes, pour chacun des deux mots) et 2 sorties (acceptation et rejet) est alors appris pour chaque stratégie de détection.

Nous allons comparer ces stratégies dans la section suivante, après avoir présenté les données que nous utilisons pour nos expérimentations.

6 Expérimentations et résultats

6.1 Données linguistiques et manuscrites

Le modèle de langage utilisé est un modèle bigramme construit sur le corpus Brown [5], avec l'outil SRILM [13]. Ce corpus est composé de 52 954 phrases anglaises (soit 1 002 675 mots). Parmi celles-ci, 46 836 phrases (soit

900 108 mots) ont été utilisées pour l'apprentissage du modèle de langage. Nous avons utilisé le lexique associé qui est constitué de 13 748 mots.

Les données manuscrites sont des phrases saisies à partir des 2 598 phrases restantes du corpus Brown. L'ensemble d'apprentissage inclut 398 phrases (soit 6 420 mots) écrites par 17 scripteurs. Cet ensemble d'apprentissage est utilisé pour régler les paramètres des algorithmes de Viterbi et du Consensus ainsi que pour apprendre les classifieurs utilisés pour la détection et la correction des erreurs. L'ensemble de test, quant à lui, inclut 320 phrases (soit 5 071 mots) écrites par 10 scripteurs. Les scripteurs de l'ensemble de test sont différents de ceux de l'ensemble d'apprentissage.

6.2 Réseaux de confusion pour la reconnaissance de phrases

La table 1 compare les résultats obtenus en utilisant soit l'approche MAP sur le graphe de mots, soit l'approche du Consensus sur le réseau de confusion, pour chercher la meilleure phrase. Nous rappelons le taux de reconnaissance obtenu sans utiliser de modèle de langage lors de la reconnaissance (système de base). Nous comparons aussi les résultats obtenus en utilisant soit l'extraction manuelle soit l'extraction automatique des mots des phrases considérées.

TAB. 1 – Comparaison des deux approches de reconnaissance et des deux stratégies d'extraction des mots.

Extraction	Approche	Taux de reco.	Baisse rel. taux erreur
Manuelle	Base	81,3 %	-
	MAP	89,9 %	45,7 %
	Consensus	89,5 %	44,0 %
Auto.	Base	77,3 %	-
	MAP	87,8 %	45,7 %
	Consensus	85,7 %	37,0 %

En ce qui concerne tout d'abord l'extraction manuelle des mots, l'amélioration obtenue avec l'approche MAP est plus élevée que celle obtenue avec l'approche Consensus, relativement au système de base : la baisse relative du taux d'erreur sur les mots est ainsi de 45,7 % sur l'ensemble de test. Le fait que l'approche Consensus donne des résultats un peu moins bons que l'approche MAP peut être expliqué par le fait que l'approche MAP permet déjà d'atteindre un taux de reconnaissance élevé. En effet, il a été montré dans [3] que la corrélation entre le taux d'erreur sur les mots et celui sur les phrases est d'autant plus forte que le taux d'erreur sur les mots est bas. Ainsi, le bénéfice du passage de l'approche MAP à l'approche Consensus est amoindri. De plus, comme notre système de reconnaissance de mots n'est pas probabiliste, l'actuelle combinaison des scores lexical et graphique avec le modèle de langage peut ne pas être optimale.

En utilisant l'extraction automatique des mots, les taux de reconnaissance obtenus sont inférieurs. Nous constatons aussi que la différence entre les résultats obtenus avec l'approche MAP et avec l'approche Consensus est plus importante. Cela est dû au problème des sur- et sous-segmentations résultant de l'extraction automatique des mots. En effet, les phrases courtes ont tendance à être favorisées car leur probabilité est souvent inférieure à celle de phrases plus longues. Ainsi, dans un ensemble de confusion de la forme de celui présenté à la figure 3, le chemin passant par l'arc correspondant à l'ensemble de traces « *and sometimes* » sera favorisé, au détriment du chemin passant par les arcs correspondants aux ensembles de traces *and* puis *sometimes*. Contrairement à l'approche Consensus, l'approche MAP est moins pénalisée par ces éventuels problèmes de segmentation puisque le nombre de mots des phrases est intégré dans le calcul de la probabilité des phrases (voir section 3).

Dans les expérimentations suivantes, nous utilisons les probabilités *a posteriori* des mots (calculées sur le réseau de confusion) pour évaluer la confiance en la phrase résultat, obtenue par l'approche MAP. Les classifieurs utilisés pour corriger les erreurs ainsi détectées sont des *Support Vector Machines* (SVM) avec des noyaux gaussiens. Nous avons choisi d'utiliser ces classifieurs pour leur capacité à traiter les cas où les classes sont déséquilibrées (en termes de nombre d'exemples d'apprentissage) et parce que ce sont des classifieurs fiables et efficaces.

6.3 Détection et correction d'erreurs

La détection des erreurs des deux types, présentés dans les sections 5.1 et 5.2, permet la sélection de 10,7 % des mots de l'ensemble de test, dans le cas de l'extraction manuelle des mots, et de 9,7 % des mots pour l'extraction automatique de ceux-ci. Ces mots ainsi sélectionnés correspondent à 62,5 % des erreurs commises par l'approche MAP, pour l'extraction manuelle, et 42,3 % des erreurs commises par l'approche MAP, dans le cas de l'extraction automatique des mots.

La table 2 compare le taux de reconnaissance obtenu en utilisant l'approche MAP seule à celui obtenu en utilisant l'approche présentée, permettant la correction des erreurs détectées. Les taux de reconnaissance des mots sont donnés par rapport à tous les mots de l'ensemble de test ainsi que parmi les mots sélectionnés (c'est-à-dire les mots détectés comme potentiellement en erreur) et parmi les mots présents (c'est-à-dire les mots sélectionnés et tels que le mot à reconnaître est l'un des deux mots en entrée du classifieur utilisé pour la correction). Nous comparons aussi les résultats obtenus lorsque l'extraction des mots est manuelle ou automatique.

La stratégie de détection et de correction permet d'améliorer le taux de reconnaissance, que l'extraction des mots soit manuelle ou automatique, les taux de reconnaissance étant respectivement 90,4 % et 88,3 %. En réalité, le taux

TAB. 2 – Taux de reconnaissance pour la correction globale des erreurs.

Extraction	Approche	Tous mots	Mots sélec.	Mots présents
Manuelle	MAP	89,9 %	40,8 %	69,3 %
	Dét./corr.	90,4 %	45,9 %	78,7 %
Auto.	MAP	87,8 %	43,3 %	73,2 %
	Dét./corr.	88,3 %	48,1 %	81,3 %

de reconnaissance atteignable n'est pas de 100 % et correspond au cas où toutes les erreurs détectées auraient été corrigées (puisque aucune amélioration n'est possible dans le cas où de potentielles erreurs ne sont pas détectées). Ainsi, pour l'extraction manuelle, le taux de reconnaissance atteignable est de 91,8 % : la réduction du taux d'erreur, grâce à l'approche de détection et correction, par rapport à ce taux est alors de 31,3 %. Dans le cas de l'extraction automatique des mots, la réduction du taux d'erreur est plus importante puisqu'elle est de 60 %, le taux de reconnaissance atteignable étant alors 89,2 %.

Dans la section suivante, nous détaillons les résultats obtenus, pour chacun des deux types d'erreurs détectées et corrigées.

6.4 Détails sur les types d'erreurs corrigés

Détection basée sur les mots différents. Cette première stratégie permet la détection de 3,0 % des mots de l'ensemble de test, que ce soit pour l'extraction manuelle ou automatique des mots. Dans le cas de l'extraction manuelle, cela représente 18,1 % des erreurs alors que cela en représente 13,2 %, pour l'extraction automatique. Dans le cas de l'extraction automatique des mots, nous nous limitons au cas des ensembles de confusion « simples », dans lesquels un mot ne peut pas être aligné avec plusieurs mots.

La table 3 donne le taux de reconnaissance sur les mots sélectionnés et présents, pour l'approche MAP et pour l'approche corrigeant les erreurs lorsque le mot reconnu par l'approche MAP n'appartient pas à l'hypothèse consensus. La diminution des erreurs, parmi les mots présents, est alors respectivement de 43,9 % et de 15,4 %, pour les extractions manuelle et automatique des mots.

TAB. 3 – Taux de reconnaissance pour la décision basée sur les mots différents.

Extraction	Approche	Mots sélec.	Mots présents
Manuelle	MAP	38,0 %	58,2 %
	Dét./corr.	50,0 %	76,5 %
Auto.	MAP	43,4 %	70,0 %
	Dét./corr.	53,0 %	84,6 %

Détection basée sur les mots non fiables. Cette deuxième stratégie permet la détection de 7,7 % des mots de l'ensemble de test, pour l'extraction manuelle des mots, et de 6,8 % des mots, pour leur extraction automatique. Dans le cas de l'extraction manuelle, cela représente 44,4 % des erreurs alors que cela en représente 29,1 %, pour l'extraction automatique.

La table 4 donne le taux de reconnaissance sur les mots sélectionnés et présents, pour l'approche MAP et pour l'approche corrigeant les erreurs lorsque le mot reconnu a une probabilité *a posteriori* jugée non fiable (inférieure à 0,8). Les erreurs, parmi les mots présents, sont diminuées de 17,5 %, dans chacun des cas d'extraction des mots.

TAB. 4 – Taux de reconnaissance pour la décision basée sur les mots non-fiables.

Eextraction	Approche	Mots sélec.	Mots présents
Manuelle	MAP	41,8 %	74,2 %
	Dét./corr.	44,4 %	78,7 %
Auto.	MAP	43,2 %	74,9 %
	Dét./corr.	45,7 %	79,3 %

6.5 Rejet des mots absents

Comme nous l'avons présenté dans la section 5.3, il peut être intéressant de distinguer les erreurs que l'on pourra corriger de celles qui ne pourront l'être. Nous traitons ici la distinction de ces deux catégories d'erreur, dans le cas de l'extraction manuelle des mots. Dans ce cas, 93,3 % des mots à reconnaître qui apparaissent dans le réseau de confusion et qui sont détectés par la stratégie présentée dans la section 5.1 correspondent à l'un des deux mots présentés en entrée du classifieur utilisé pour la correction. De la même façon, 88,4 % des mots présents détectés par la seconde stratégie (voir section 5.2) correspondent à l'un des deux mots en entrée du classifieur correspondant. Les autres mots sélectionnés par chacune des deux stratégies sont, quant à eux, à rejeter.

La figure 4 représente la courbe ROC pour différents classifieurs de rejet, pour chacune des stratégies de détection. Ces courbes montrent le compromis entre les mots corrects qui n'ont pas été rejetés (TAR) et les mots à rejeter mais qui ne l'ont pas été (FAR).

Le point M correspond au classifieur choisi pour le rejet des mots sélectionnés par la première stratégie : il permet le rejet de 12,0 % de ces mots (avec un TAR de 99,0 %). Pour les mots sélectionnés par la deuxième stratégie, nous avons considéré deux classifieurs de rejet (représentés par les points U1 et U2, sur la figure 4). Avec le premier classifieur de rejet, 8,4 % des mots sélectionnés sont rejetés (avec un TAR de 98,6 %) alors que le deuxième classifieur entraîne le rejet de 14,5 % des mots sélectionnés (avec un TAR de 98,2 %).

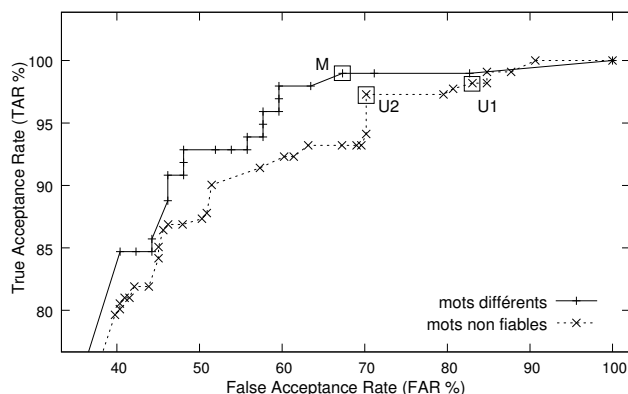


FIG. 4 – Courbes ROC pour le rejet des mots sélectionnés, pour chacune des stratégies de détection.

La table 5 compare les taux de reconnaissance, de rejet et d'erreur, lorsque le rejet est utilisé en complément de la détection et de la correction des erreurs. Nous constatons que le taux de reconnaissance est légèrement inférieur à celui obtenu sans utiliser de rejet (voir section 6.3). Néanmoins, les mots rejetés sont distingués des mots en erreur. Nous obtenons ainsi une réduction de 11,4 % des erreurs, lorsque nous utilisons les classifieurs de rejet M et U1, et une diminution des erreurs de 15,5 %, avec les classifieurs de rejet M et U2. Ces premiers résultats sont encourageants pour la suite de nos travaux.

TAB. 5 – Taux de reconnaissance, de rejet et d'erreur pour le rejet des mots absents.

Stratégie	Taux reconnaissance	Taux rejet	Taux erreur
Rejets M+U1	90,3 %	1,1 %	8,6 %
Rejets M+U2	90,3 %	1,5 %	8,2 %

7 Conclusion et perspectives

Dans cet article, nous avons présenté l'intégration d'une représentation des hypothèses de reconnaissance de phrases par un réseau de confusion, dans un système de reconnaissance de phrases manuscrites dans lequel un graphe de mots était déjà utilisé. Nous avons de plus adapté ce réseau de confusion afin de pouvoir prendre en compte des hypothèses de sur- et sous-segmentations des mots. Les probabilités *a posteriori* calculées sur ce réseau de confusion ont ensuite été utilisées comme indice de confiance sur les mots de la phrase reconnue par l'approche MAP sur le graphe de mots. Deux types d'erreurs ont ainsi été mis en évidence. Afin de corriger chacun de ces types d'erreurs, un mécanisme de correction utilisant des SVM a été mis en place. Cette étape de correction a permis la diminution du nombre d'erreurs sur les mots de la phrase obtenue par l'approche MAP, dans le cas de l'extraction manuelle des

mots de la phrase mais aussi dans le cas de l'extraction automatique de ces mots. Enfin, une étape de rejet a été intégrée afin d'identifier les mots détectés comme potentiellement non reconnus mais ne pouvant être corrigés lors de l'étape de correction. Ces mots pourront ensuite être mis en évidence ou être soumis à un nouveau traitement.

Lors de l'extraction automatique des mots et leur reconnaissance en utilisant l'approche Consensus, nous avons constaté que les mots appartenant à des phrases sous-segmentées étaient favorisés. Dans nos travaux futurs, nous nous intéresserons à la prise en compte du nombre de mots dans les phrases, lors du calcul des probabilités *a posteriori* des mots, afin de remédier à ce problème. Nous étendrons aussi la détection et la correction des erreurs dans le cas où les phrases obtenues par chacune des deux approches (MAP et Consensus) n'ont pas la même taille. En effet, dans ce cas, il faut aligner un mot d'une des phrases avec plusieurs mots de l'autre phrase : cela se traduit par des ensembles de confusion comme celui présenté à la figure 3. Au lieu de choisir entre les deux mots différemment reconnus, il faut en fait choisir entre deux chemins : l'idée serait alors d'étendre la classification de deux mots à celle de deux chemins. Ce type d'erreur représentant 23,6 % des erreurs de reconnaissance des mots, cela laisse envisager d'intéressantes perspectives d'amélioration. Enfin, nous étendrons aussi la stratégie de rejet présentée pour pouvoir l'utiliser dans le cas de l'extraction automatique des mots.

Références

- [1] S. Carbonnel and E. Anquetil. Lexical Post-Processing Optimization for Handwritten Word Recognition. In *7th ICDAR*, pages 477–481, 2003.
- [2] S. Carbonnel and E. Anquetil. Lexicon Organization and String Edit Distance Learning for Lexical Post-Processing in Handwriting Recognition. In *9th IWFHR*, pages 462–467, 2004.
- [3] G. Evermann and P.C. Woodland. Large Vocabulary Decoding and Confidence Estimation Using Word Posterior Probabilities. In *25th ICASSP*, pages 1655–1658, 2000.
- [4] G.D. Forney. The viterbi Algorithm. *Proceedings of the IEEE*, 61(3) :268–278, 1973.
- [5] W.N. Francis and H. Kucera. *Brown Corpus Manual*. Brown University, 1979.
- [6] D. Hillard and M. Ostendorf. Compensating for Word Posterior Estimation Bias in Confusion Networks. In *31st ICASSP*, pages 1153–1156, 2006.
- [7] T. Kemp and T. Schaaf. Estimating Confidence Using Word Lattices. In *5th Eurospeech*, pages 827–830, 1997.
- [8] L.L. Mangu. *Finding Consensus in Speech Recognition*. PhD thesis, Johns Hopkins University, 2000.
- [9] F. Perraud, C. Viard-Gaudin, E. Morin, and P.-M. Lallican. Statistical Language Models for On-Line Handwriting Recognition. *IEICE Transactions on Information and Systems*, E88-D(8) :1807–1814, 2005.
- [10] S. Quiniou and E. Anquetil. A Priori and A Posteriori Integration and Combination of Language Models in an On-line Handwritten Sentence Recognition System. In *10th IWFHR*, pages 403–408, 2006.
- [11] S. Quiniou and E. Anquetil. Use of a Confusion Network to Detect and Correct Errors in an On-line Handwritten Sentence Recognition System. In *9th ICDAR*, pages 382–386, 2007.
- [12] S. Quiniou, F. Bouteruche, and E. Anquetil. Word Extraction for the Recognition of On-Line Handwritten Sentences. In *13th IGS*, pages 52–55, 2007.
- [13] A. Stolcke. SRILM - An Extensible Language Modeling Toolkit. In *7th ICSLP*, pages 901–904, 2002.
- [14] A. Vinciarelli, S. Bengio, and H. Bunke. Offline Recognition of Unconstrained Handwritten Texts using HMMs and Statistical Language Models. *IEEE Transactions on PAMI*, 26(6) :709–720, 2004.
- [15] J. Xue and Y. Zhao. Improved Confusion Network Algorithm and Shortest Path Search from Word Lattice. In *30th ICASSP*, pages 853–856, 2005.