ANALYTIC WORD RECOGNITION WITHOUT SEGMENTATION BASED ON MARKOV RANDOM FIELDS

CHRISTOPHE CHOISY AND ABDEL BELAID
LORIA/CNRS
Campus scientifique, BP 239,
54506 Vandoeuvre-les-Nancy cedex, France
Christophe.Choisy@loria.fr, Abdel.Belaid@loria.fr

In this paper, a method for analytic handwritten word recognition based on causal Markov random fields is described. The words models are HMMs where each state corresponds to a letter; each letter is modelled by a NSHP-HMM (Markov field). Global models are build dynamically, and used for recognition and learning with the *Baum-Welch* algorithm. Learning of letter and word models is made using the parameters reestimated on the generated global models. No segmentation is necessary : the system determines itself the best limits between the letters during learning. First experiments on a real base of french check amount words give encouraging results of 83.4% for recognition.
Keywords : HMM, NSHP-HMM, Cross-learning, Meta-models, *Baum-Welch* Algorithm.

1 Introduction

These recent years, research on writing recognition shows that the 2D approach gives better results that 1D ones ([1,2,3,4,5,6]), because it takes more into acount the basically plane nature of the writing. The 2D estimator can be either a Neural Network (NN) or a planar-HMM (PHMM). The NN can be applied either on letters ([6]) or on segments ([7]); the major default of these classifiers is their lack of elasticity. The PHMM was applied with success in many works ([2,3,4]). This model, based on HMMs has 2D elasticity properties but needs an independance hypothesis between the columns which is not always true in practice.

SAON proposed, in our team, a new model based on Markov fields : the NSHP-HMM([8]). Its architecture, based on an HMM, gives it an horizontal elasticity allowing the adaptation of the length of patterns analysed. Using a 2D neighborhood of pixels, it overcomes the column independance hypothesis of the PHMMs. It applies on binary patterns, with an easy use.

All these models where applied in a model discriminant global approach of words. However, this approach has some limits. Particularly, the NSHP-HMM uses a lot of parameters (cf §.3 p 2). Another classical limit is the restricted and distinct vocabulary. To overcome these limits, an analytic approach is proposed : models for letters are smaller than models for words, and working with letters allows to extend the vocabulary without limits.

Many works on analytic words recognition are based on a segmentation into

In: L.R.B. Schomaker and L.G. Vuurpijl (Eds.), Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition, September 11-13 2000, Amsterdam, ISBN 90-76942-01-3, Nijmegen: International Unipen Foundation, pp 487-492

graphemes ([5,9,10]). This segmentation is generally based on topological criterions and cannot be 100% reliable ([9]). It seems better to let the system decide the best limits of letters. While many works use dynamic time warping algorithms to learn and recognize words, our system is based on the *Baum-Welch* algorithm, that guarantees a local optimum.

The method proposed is a dynamic generation of word models, based on letter models and HMMs in which states represent letters. The reestimation of letter models and transitions between letters is made by cross-learning. This technique, directly derived from the *Baum-Welch* reestimation formulas , consists in crossing the information relative to letters in the different word models. The use of the *Baum-Welch* algorithm allows the system to find the best repartition of parameters in the models, knowing only the label of the words learned.

2 HMM Structure

HMMs are already well known in Automatic Handwriting Recognition. RABINER ([11]) gives the bases of these stochastic models. According to RABINER notation we defined a discrete first order HMM by :
$S = \{s_1, ..., s_N\}$ the N states of the model, denoted at time t by $q_t \in S$
$V = \{v_1, ..., v_M\}$ the M observation symbols, denoted at time t by $O_t \in V$
$A = \{a_{ij}\}_{1 \leq i,j \leq N}$ the matrix of transitions probabilities;
$a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$
$B = \{b_j(k)\}_{1 \leq j \leq N, 1 \leq k \leq M}$ the matrix of observations probabilities;
$b_j(k) = P(O_t = v_k | q_t = s_i)$

Two specific states D and F are introduced : the probabilities to start or end in a state are modelled by the transitions between D and F and this state.

3 Non-Symmetric Half-plane Hidden Markov Model

The NSHP-HMM is a stochastic model of Markov fields type. This model showed very good performances in french check amount words recognition ([8,12]). It runs directly on binary images that are analysed column by column. The use of 2D neighborhoods of pixels enables the system to better take into account the 2D nature of the writing. Its architecture, based on a HMM, allows a horizontal elasticity making its adaptation easy on various image widths.

Each column is observed in one state. Its observation probability is given as the product of elementary probabilities performed for each pixel in the column. The elementary probability is determined according to a neighbor fixed in the half plane analysed before the pixel to overcome the problem of correlation between adjacent columns. Training and recognition methods are described in [8].

The parameters of the NSHP-HMM are the height of the columns analysed, the size of the neighborhood (order of the model), the number of states of the HMM.

4 Word Modeling

For word modeling we use meta-HMMs in which each meta-state represents a letter .

Starting from a meta-model, we build a global NSHP-HMM by connecting the NSHP-HMM associated to the meta-model letter states. i^x is the state i of the model associated to the meta-state x, D^x and F^x the initial and final states of this model. D_m and F_m are the specific states associated to the meta-model. Each sequence of state of type $i^x \rightarrow F^x \rightarrow D^y \rightarrow j^y$ is replaced by one transition $i^x \rightarrow j^y$, whose value is the product of the transitions between these states :
$P(j^y|i^x) = P(j^y|D^y) * P(D^y|F^y) * P(F^y|i^x)$
Following the same idea, we obtain :
$P(i^x|D_m) = P(i^x|D^x) * P(D^x|D_m)$
$P(F_m|i^x) = P(F_m|F^x) * P(F^x|i^x)$

The model obtained is a NSHP-HMM which can be applied as a global model. The meta-states must not loop on themselves because this would build transitions between states in the HMM associated, and erase those existing.

5 Cross-learning

Cross-learning consists in crossing the informations of the words to determine the informations corresponding to each letter. The reestimation of the letter models are made using the reestimation of word models. This method is derived from the Baum-Welch training, considering that each state of each word model also belongs to a letter model. The reestimation of the transition between specific and normal states of the letters models is made through transitions generated between the letters.

Let m be a meta-model of a word, with S_m normal states and D_m and F_m the specific states. At each meta-state $x \in S_m$ is associated a NSHP-HMM of a letter with S^x normal states and the specific state D^x and F^x.

During the construction of the global model, the specific states D^x and F^x are removed. The reestimation of the transitions $a_{D^x i^x}$ and $a_{i^x F^x}$ is made using generated transitions. For a meta-model m, K is the number of images analysed, O^k is the kth image, T_k is the number of columns of the image k, $P_k = P(O^k|m)$.
For a model associated to the meta-state x :

- the transition $a_{D^x i^x}$ is used to build the transitions $a_{j^y i^x}$ $y \neq x$ and $a_{D_m i^x}$
- the transition $a_{i^x F^x}$ is used to build the transitions $a_{i^x j^y}$ $y \neq x$ and $a_{i^x F_m}$

- the internal transitions $a_{i^x j^x}$ are left unchanged.

The principle of the cross-reestimation is to gather this information for all the models associated with a same letter in the various meta-models. For the internal transitions the Baum-Welch formula can be applied directly by summing the paths containing the transition over all occurences of a letter model in all the word models (the reestimation of observation probabilities follows the same principle). For the transitions $a_{D^x i^x}$ and $a_{i^x F^x}$ this sum is made through the sum of the paths containing the transitions built with these.

6 Meta-model Reestimation

Global models are build from meta-models. These are HMMs and the transitions between the meta-state from the informations of the generated models can be reestimated. Indeed, for $x, y \in S$, we obtain by construction :
$a_{xy} = a_{F^x D^y}, a_{D_m x} = a_{D_m D^x}, a_{xF_m} = a_{F^x F_m}$.

- the transition a_{xy} $(x \neq y)$ is used to build the transitions $a_{i^x j^y}$
- the transition $a_{D_m x}$ is used to build the transitions $a_{D_m i^x}$
- the transition a_{xF_m} is used to build the transitions $a_{i^x F_m}$

As for the cross-learning, according to the Baum-Welch formulas, the reestimation of a meta-transition is made by summing on all the paths containing the transition using this.

7 First Experiments

The system was tested on a base of 7031 french bank check words given by the SRTP [a] (vocabulary of 26 words). The parameters of the NSHP-HMM for the letter models are : height of 20 pixels, 3 pixels for the neighborhoods; the number of normal states for the NSHP-HMM corresponding to a letter is $\overline{n}/2 + 1$, where $\overline{n}$ is the average number of columns of samples for the letter. The meta-models synthetise the frequent errors found in the words.

Two preprocessing steps are applied to reduce the variability of the writing. The first is a slant correction, as proposed in [12]. The second normalizes the height of the words by normalizing the 3 writing bands in 3 equal vertical parts.

A test was performed to validate the cross-learning principle : the interest of this method is that all the models and the meta-models can theoretically be learnt in the

[a] Service de Recherche Technique de la Poste : French Post Research Team

same time knowing only the labels of the words. The base was split approximatively in 66% (4626 words) for cross-learning and 34% (2405 words) for recognition tests. Word meta-models and letter models with equal probabilities of transitions and observations are generated and the cross-learning is applied at several steps. The results are reported for several learning steps in Table 1. The results show the efficiency of the cross-learning to gather the informations of letters from various words without initialization of the system.

Table 1. Average word recognition rates for different numbers of learning steps

cross-learning	top 1	top 2	top 3	top 5
5 steps	80.96%	88.48%	91.43%	94.47%
10 steps	83.12%	89.23%	92.35%	95.14%
15 steps	83.41%	89.31%	92.02%	94.84%
20 steps	82.83%	89.15%	91.56%	94.43%

This approach allows the reduction of the complexity of the system proposed by SAON ([8,12]). The number of floating point operations is proportional to the number of states. The global approach proposed by SAON has a high number of states, based on the mean size of words. Our approach considers the mean size of letters reducing the number of states by a factor 7; this divides the floating point operations necessary for a word analysis by 7.

At the same time, we observe that a neighborhood of size 4 is too high for the letter recognition. We choose a size of 3 which divides by two the number of parameters to estimate for each model. The combination of these factors allows a reduction of a factor 14 for the number of parameters to estimate.

8 Conclusion

We proposed a new approach for analytic word recognition based on a dynamic generation of global models. This approach divides the number of parameters of the system of SAON by 14. The first tests give encouraging results of 83.4%. The learning of letter models is made between the words models, and the *Baum-Welch* algorithm ensures the optimal learning in good conditions. More tests need to be made with bigger databases in order to evaluate in a better condition such an approach.

The word models are dynamically generated, corresponding to meta-models. At first remark, we can say that this method can easily be extended at entire amounts with another level of meta-models. This extension requires we can find the best path between words. This problem is the same with a generalisation of our method to unconstrained vocabulary recognition. For such a task, we need to find the sequence of states in the meta-model that best describes the word analysed. Some studies are necessary to find the best method to get the path in the meta-models.

References

1. H. S. Park and S. W. Lee. An HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition. In *IEEE Proceedings of ICPR'96*, volume 2, pages 320–324, 1996.
2. R. Bippus. 1-Dimensional and Pseudo 2-Dimensional HMMs for the Recognition of German Literal Amounts. In *Fourth International Conference on Document Analysis and Recognition (ICDAR'97)*, volume 2, pages 487–490, Ulm, Germany, Aug. 1997.
3. O. E. Agazzi and S. Kuo. Hidden Markov Model Based Optical Character Recognition in the Presence of Deterministic Transformation. *Pattern Recognition*, 26(12):1813–1826, February 1993.
4. M. Gilloux. Reconnaissance de chiffres manuscrits par modèle de Markov pseudo-2D. In *Actes du* $3^{ème}$ *Colloque National sur l'Écrit et le Document*, pages 11–17, Rouen, France, 1994.
5. M. Gilloux, B. Lemarié, and M. Leroux. A Hybrid Radial Basis Function Network/Hidden Markov Model Handwritten Word Recognition System. In *Third International Conference on Document Analysis and Recognition (ICDAR'95)*, pages 394–397, Montréal, 1995.
6. J. C. Simon, O. Baret, and N. Gorski. A System for the Recognition of Handwritten Literal amounts of checks. In *Internal Association for Pattern Recognition Workshop on Document Analysis System (DAS'94), Kaiserlautern, Germany*, pages 135–155, September 1994.
7. B. Lemarié, M. Gilloux, and M. Leroux. Un modèle neuro-markovien contextuel pour la reconnaissance de l'écriture manuscrite. In *Actes 10ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle*, Rennes, France, 1996.
8. G. Saon. *Modèles markoviens uni- et bidimensionnels pour la reconnaissance de l'écriture manuscrite hors-ligne*. PhD thesis, Université Henri Poincaré - Nancy I, Vandœuvre-lès-Nancy, 1997.
9. Mou-Yen Chen, Amlan Kundu, and Jian Zhou. Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type Stochastic Network. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 16(5):481–497, 1994.
10. F. Kimura M. Shridhar, Gilles Houle. Handwritten word recognition using lexicon free and lexicon directed word recognition algorithms. In *Fourth International Conference on Document Analysis and Recognition (ICDAR'97)*, Ulm, Germany, Aug. 1997.
11. L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), February 1989.
12. G. Saon and A. Belaïd. Off-line Handwritten Word Recognition Using A Mixed HMM-MRF Approach. In *Fourth International Conference on Document Analysis and Recognition (ICDAR'97)*, volume 1, pages 118–122, Ulm, Germany, Aug. 1997.