

CSE561-Probabilistic Graphical Models : Assignment 1

Ashutosh Gupta and Sagar Verma

March 28, 2016

1 Problem Description

1.1 Optical Word Recognition

Optical word recognition problem is the identification of word from a given image of characters. OWR is an extension of Optical character recognition. Through various computer vision techniques individual characters can be obtained from an image with certain probability for each character for a single image. To construct meaningful word from those characters we need prior probabilities of occurrence of one character after another. This whole problem can be represented as a graphical model and maximum probable explanation (MAP) of the model gives the probable word the image is showing.

In given problem we have 1000 image ids and for each image id we have 10 characters and their factors. These factors represent the certainty with which each character is the content of that image. Transition factors between every two characters is given. Skip factor is given if two same image ids means same characters which is 5.0 otherwise 1.0. Given these factors, the probability to the character variables of a word w according to model will be given by:

$$p(chars) = \frac{1}{Z} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \prod_{i=0, \dots, |w|-2} \psi_t(char(i), char(i+1)) \prod_{i,j | img(i)=img(j)} \psi_s(char(i), char(j)) \right)$$

where ψ_o is OCR factor, ψ_t is transition factor and ψ_s is skip factor.

2 Solution Approach

2.1 Undirected Graphical Model

Assume a word consisting n characters and it formed of n images. Graphically this can be represented as below

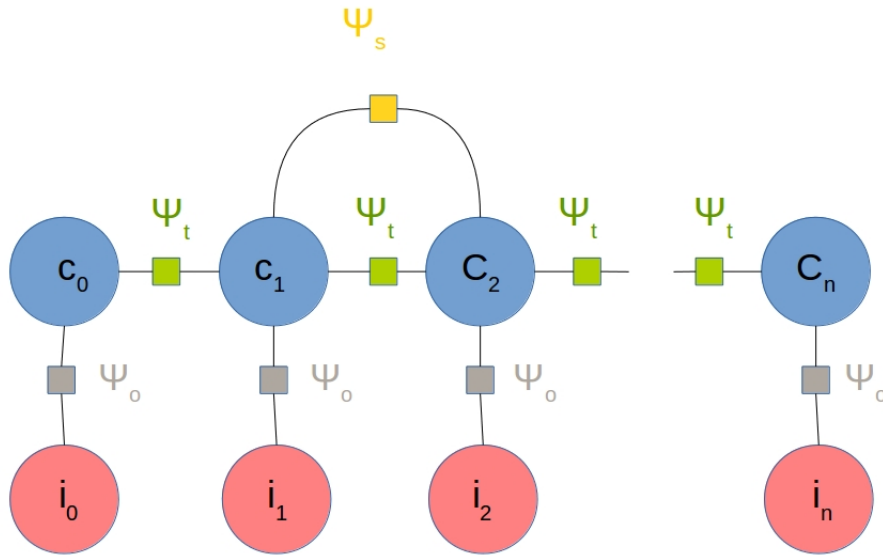


Figure 2.1: OWR problem as Undirected Graphical Model

2.1.1 OCR Model

OCR model contains only OCR factors i.e. ψ_o . Most probable explanation can be calculated as,

$$\begin{aligned} \tilde{p}(chars) &= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \tilde{p}(c_0, c_1, \dots, c_n) \\ &= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \right) \end{aligned}$$

Partition function, z can be calculated as,

$$z = \sum_{\psi_o} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \right)$$

Probability of word can be calculated as,

$$p(chars) = \frac{1}{z} \tilde{p}(chars)$$

2.2 Transition Model

Transition model contains both OCR factors and transition factors. Most probable explanation can be calculated as,

$$\begin{aligned} \tilde{p}(chars) &= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \tilde{p}(c_0, c_1, \dots, c_n, i_0, i_1, \dots, i_n) \\ &= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \prod_{i=0, \dots, |w|-2} \psi_t(char(i), char(i+1)) \right) \end{aligned}$$

Partition function, z can be calculated as,

$$z = \sum_{\psi_o \psi_t} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \prod_{i=0, \dots, |w|-2} \psi_t(char(i), char(i+1)) \right)$$

Probability of word can be calculated as,

$$p(chars) = \frac{1}{z} \tilde{p}(chars)$$

2.3 Combined Model

Combined model contains OCR factors, transition factors and Skip factors. Most probable explanation can be calculated as,

$$\begin{aligned}
\tilde{p}(chars) &= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \tilde{p}(c_0, c_1, \dots, c_n, i_0, i_1, \dots, i_n) \\
&= \operatorname{argmax}_{c_0, c_1, \dots, c_n} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \prod_{i=0, \dots, |w|-2} \psi_t(char(i), char(i+1)) \right. \\
&\quad \left. \prod_{i, j | img(i)=img(j)} \psi_s(char(i), char(j)) \right)
\end{aligned}$$

Partition function, z can be calculated as,

$$\begin{aligned}
z &= \sum_{\psi_o \psi_t \psi_s} \left(\prod_{\forall i} \psi_o(img(i), char(i)) \prod_{i=0, \dots, |w|-2} \psi_t(char(i), char(i+1)) \right. \\
&\quad \left. \prod_{i, j | img(i)=img(j)} \psi_s(char(i), char(j)) \right)
\end{aligned}$$

Probability of word can be calculated as,

$$p(chars) = \frac{1}{z} \tilde{p}(chars)$$

3 Results Obtained

Exhaustive inference can be done by calculating probabilities for all possible words for given set of images. To find most probable explanation(MPE) or MAP, junction tree algorithm is used.

Model	Char Accuracy	Word Accuracy	Log Likelihood
OCR	0.53921	0.08653	-7.54175
Transition	0.66274	0.259615	-6.88575
Combined	0.71176	0.355769	-6.102387

Table 3.1: Model Accuracy

Table 3.1 shows character-wise accuracy, word-wise accuracy and average dataset log-likelihood for all three models. From the results we can see that combined model performs best out of the three models and OCR model performs the worst.

4 Further Fun

Varying the potentials has no effect on the word predicted or the accuracy calculated. Exhaustive inference for OCR model can be calculated using matrix multiplication which is faster than generating all possible words and calculating probabilities for each of them.