UNIVERSITAT POLITÈCNICA DE VALÈNCIA
ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA INFORMÀTICA
DEPARTAMENT DE SISTEMES INFORMÀTICS I COMPUTACIÓ

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Quantificació de les millores en la segmentació del cos central del text manuscrit utilitzant aprenentatge supervisat

Projecte Final de Carrera - Enginyeria Informàtica

Joan Puigcerver i Pérez

Supervisat per:
Dr. Moisés Pastor i Gadea

17 de desembre de 2012

Als meus pares per mai haver-me dit "prou".
A la meva germana per suportar-me
fent-li la punyeta.
A Anna per alegrar-me els dies i
acompanyar-me arreu del món.
A Adrià i Fernando per tots els riures
durant aquests cinc anys.
I a Moisés per la seva confiança
i oportunitats que han resultat
en aquest treball i m'han obert portes
a llocs increïbles.

Índex

Capítol 1

Introducció

1.1 Motivació

Des de fa unes poques dècades, els ordinadors permeten emmagatzemar una àmplia varietat d'informació de manera fiable i replicada per a que puga sobreviure al pas del temps, mantenir-la organitzada per a que siga fàcilment utilitzable i fer-la accessible arreu del món i quasi universalment. Però durant segles, l'única forma de transmetre el coneixement i emmagatzemar-lo de manera més o menys segura ha sigut mitjançant l'escriptura. Precisament el fet de mantenir el coneixement en llibres, manuscrits primer i impresos després, que permeten la seva preservació i més fàcil difusió, ha sigut una de les principals bases de tot el desenvolupament del coneixement humà, especialment científic i tecnològic, arreu del món i és el principal motor pel qual avui gaudim d'unes millors condicions de vida que les dels nostres avantpassats.

Malauradament, els llibres manuscrits i impresos no sempre han tingut èxit en la seva missió de preservar el coneixement. Sols cal recordar el desastrós incendi de l'antiga Biblioteca d'Alexandria que provocà que milers d'obres d'autors de l'antiguitat es perderen per sempre. Ajudar a la preservació de la informació continguda en els llibres manuscrits i ajudar també a la cerca del contingut en aquests, són dos dels motius que van fer nàixer el reconeixement de text manuscrit (HTR, de l'anglès *Handwriting Text Recognition*) a principis del segle XX.

Tot i el progrés aconseguit en els últims anys pel Reconeixement de Text

Manuscrit, aquest encara té molts problemes per resoldre. Molts d'ells causats perquè la gran variabilitat que presenta el text, en quant a l'estil d'escriptura, no aporta cap informació rellevant per a la classificació dels símbols representats a les imatges i dificulta el seu reconeixement. Per exemple, un mateix autor no escriu un mateix símbol sempre de la mateixa forma, ni de la mateixa grandària i ni tan sols amb la mateixa orientació. Per això, un dels components fonamentals de qualsevol sistema de reconeixement de l'escriptura és la normalització d'aquest text, un procés que tracta de reduir aquesta variabilitat.

Aquest projecte compara i quantifica les diferències entre dues alternatives per solucionar un dels problemes que forma part d'aquest procés de normalització: la segmentació del del cos central del text manuscrit. El cos central d'una línia de text manuscrit és aquella porció de la línia on resideix el cos central de cadascun dels símbols que formen el text. Les dues alternatives estudiades per a aquesta segmentació del cos central estan basades en un enfocament heurístic del problema, on un algorisme amb unes regles pre-establertes determina quina és la regió del cos central, i una altra basada en tècniques d'aprenentatge supervisat, on un humà ha segmentat manualment el cos central d'un conjunt d'imatges de mostra i ha entrenat el sistema per a que intente segmentar de manera semblant les noves imatges. Els detalls es veuran en el capítol 2.

1.2 Reconeixement de formes

El reconeixement de formes (de l'anglès *Pattern Recognition*) és una branca de l'aprenentatge automàtic que té com a objectiu fer que una màquina (un ordinador) tinga la capacitat de discernir entre diferents objectes del seu entorn. El sistema pot percebre el seu entorn a partir de diferents sensors com poden ser càmeres fotogràfiques o de vídeo, micròfons, sensors làser, de temperatura, etc. L'objectiu és, a partir dels senyals obtinguts per aquests sensors, descobrir i atorgar un significat als diferents objectes representats en els senyals (típicament, assignar-los a una categoria) [DH73].

Existeixen dos grans grups de problemes de reconeixement de formes. El primer, com s'havia dit, consisteix en assignar una categoria a algun objecte representat per una imatge. Per exemple, donada una imatge que conté un símbol, decidir quin és aquest símbol d'entre un conjunt de possibilitats (reconèixer el símbol a la imatge). És diu d'aquests problemes que tenen un

aprenentatge supervisat perquè al sistema se l'entrena amb senyals d'entrada prèviament etiquetats, de manera que puga aprendre quines són les propietats del senyal que determinen la seva categoria.

En el segon grup, no és disposa de les categories assignades al senyal durant l'etapa d'entrenament. L'objectiu d'aquest grup és descobrir propietats en el senyal d'entrada. Per exemple, assignar una categoria automàtica a cadascun dels senyals d'entrada (*clustering*) o trobar un model (típicament probabilístic) que permeta representar les dades d'entrenament.

Són moltes les aplicacions que poden ser interpretats com un problema de reconeixement de formes, és per això, i gràcies a l'avanç en la tecnologia informàtica, que l'interès en aquest camp a crescut notablement en les darreres dècades. Algunes de les aplicacions del reconeixement de formes són:

- Aplicacions sobre el llenguatge humà: Reconeixement automàtic de text, reconeixement automàtic de la parla, traducció automàtica, etc.
- Aplicacions sobre imatges: Reconeixement facial, detecció i classificació d'objectes, etc.
- Medicina: Detecció de tumors, classificació de cromosomes, etc.
- Física: Detecció i classificació de cossos celestes, detecció de partícules subatòmiques, etc.

En el cas de l'aprenentatge supervisat, que és el més estès i el que s'utilitzarà en una de les tècniques descrites en aquest treball per a la segmentació del cos central (veure secció 2.2), el procés de reconeixement es fa seguint el següent procediment (veure figura 1.1).

1. **Preprocessament**: Al senyal se li apliquen diferents transformacions que tenen com a objectiu eliminar la informació no rellevant per a la classificació.
2. **Extracció de característiques**: El resultat del preprocessament sol ser un senyal amb un alt nombre de dimensions i això dificulta l'estimació dels models per a la classificació. Per això, es tracta de reduir aquest nombre de dimensions extraient el que s'anomenen "característiques" del senyal (informació més rellevant per a representar-lo), que no són més que una representació del senyal en un espai menor de dimensions.

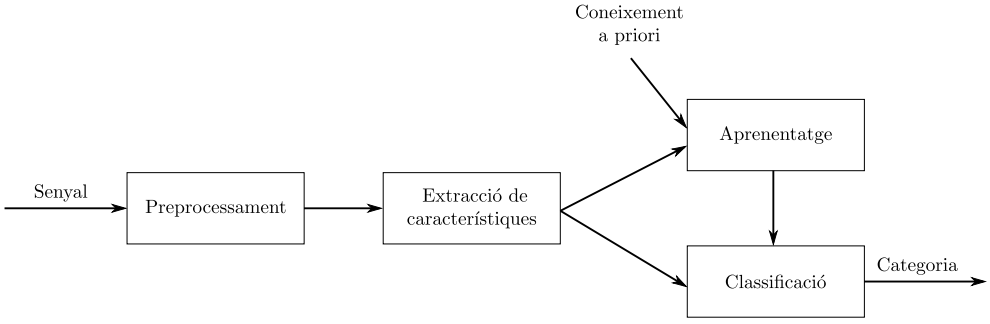


Figura 1.1: Esquema d'un sistema de reconeixement de formes.

3. **Aprenentatge**: S'aprèn un model per a classificar el senyal amb la informació aportada per les característiques del senyal d'entrada i el coneixement a priori sobre les dades i la tasca (categoria del senyal, probabilitats de les categories, etc). Aquesta fase sols és utilitzada durant l'aprenentatge del model.

4. **Classificació**: Una vegada l'aprenentatge està complet, s'utilitzen les característiques del senyal i el model obtingut en el pas anterior per a predir la categoria d'una nova observació del senyal.

L'objectiu de l'aprenentatge supervisat pot interpretar-se com l'estimació d'una funció g amb perfil $g : X \rightarrow C$, on X és l'espai de característiques del senyal i C el conjunt de categories; de manera que es minimitzen els errors de classificació per a un conjunt de dades d'entrenament $D = \{(x_1, c_1), (x_2, c_2), \ldots, (x_n, c_n)\}$ on x_i és són les característiques obtingudes d'una mostra del senyal d'entrada i c_i la seva categoria assignada.

La millor manera de construir aquesta funció és assignant la categoria $\hat{c}$ que maximitza la probabilitat de que les característiques del senyal observat representen aquesta categoria [DH73].

$$g(x) = \underset{\forall c \in C}{\operatorname{argmax}}\, p(c|x) = \hat{c} \tag{1.1}$$

L'equació 1.1 pot escriure's de nou utilitzant el Teorema de Bayes segons l'equació 1.2 [BP63].

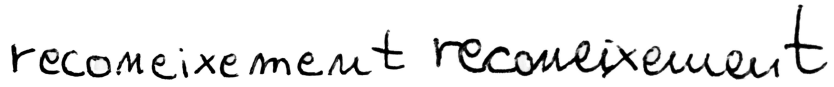
$$g(x) = \underset{\forall c \in C}{\operatorname{argmax}} \frac{p(c,x)}{p(x)} = \underset{\forall c \in C}{\operatorname{argmax}} \frac{p(x|c)p(c)}{p(x)} = \underset{\forall c \in C}{\operatorname{argmax}}\, p(x|c)p(c) \qquad (1.2)$$

Si es conegueren les distribucions exactes de $p(x|c)$ i $p(c)$, el problema de classificació quedaria resolt. Malauradament aquestes distribucions de probabilitat són desconegudes en les aplicacions reals i la major part del treball es concentra en trobar unes bones aproximacions d'aquestes a partir de les dades d'entrenament.

1.3 Reconeixement de text manuscrit

El reconeixement de text manuscrit (HTR, de l'anglès *Handwriting Text Recognition*) començà a desenvolupar-se a principis del segle XX amb l'aparició de dispositius que permetien classificar dígits o caràcters manuscrits sobre sensors d'una determinada forma [Gol14]. A mitjans del segle XX aparegueren els primers dispositius que permetien connectar-se en ordinadors i classificar dígits o caràcters aïllats [Dim57] i fins i tot sistemes que eren capaços de reconèixer paraules manuscrites aïllades [Har62]. Aquests primers dispositius estaven limitats per el desenvolupament tècnic d'aquell temps i solien necessitar-se dispositius especials sobre els que escriure el text. No era possible reconèixer el text d'un document manuscrit existent (com un llibre o un formulari). Fou més tard i gràcies a l'expansió de la informàtica quan els sistemes HTR començaren a adquirir la qualitat necessària per a aplicar-los en tasques de reconeixement real, com ara el reconeixement de dígits en xecs bancaris o el reconeixement de codis postals en el servei de correu.

Entre els sistemes HTR se'n diferencien dos tipus: el reconeixement *online* i l'*offline*. La diferència rau en el mètode en que el senyal d'entrada al sistema és adquirit. En els sistemes HTR *online* el reconeixement té informació temporal dels traços del text gràcies a l'ús de bolígrafs electrònics que capturen el moviment de l'eina d'escriptura, la velocitat d'escriptura, la pressió sobre la superfície, etc. El reconeixement *offline* es duu a terme una vegada el text ja ha sigut escrit sobre una superfície (típicament paper). Un escànner, càmera fotogràfica o de vídeo obté llavors una imatge de la superfície que conté el text a reconèixer i aquest és el senyal a utilitzar pel sistema de reconeixement. El cas del reconeixement *online* es considera un problema més senzill a resoldre



(a) Text amb símbols aïllats **(b)** Text amb símbols continus

Figura 1.2: Dos exemples d'una paraula escrita amb símbols aïllats i continus.

(és a dir, aconseguir sistemes amb un menor error de reconeixement) perquè es disposa d'informació temporal explícita que no està disponible en el reconeixement *offline*. Precisament els sistemes de reconeixement *offline* que més èxit tenen són aquells que intenten recuperar aquesta informació temporal que es troba implícita a les imatges, com per exemple els basats Models Ocults de Markov (HMM).

Una altra divisió en les tasques de reconeixement de text, és si els símbols que apareixen es troben aïllats o són continus. En el reconeixement de símbols aïllats, cadascun dels símbols a reconèixer es troba ben separat de la resta (figura 1.2a). No ocorre el mateix en el reconeixement de text continu, on pot haver diferents símbols units per un mateix traç (figura 1.2b). En el sistema d'escriptura occidental, el text continu és el més habitual en el text manuscrit. El cas continu es considera més difícil ja que la forma d'un símbol depèn fortament dels seus adjacents i l'unió de certs símbols pot ocasionar ambigüitat i requereix més context (per exemple, "rn" o "m", "vv" o "w", etc).

Entre els principals problemes als que s'enfronta un bon sistema de reconeixement del text manuscrit són:

- **Soroll en les imatges**: la imatge capturada pot contenir soroll que depèn de l'eina d'escriptura (un bolígraf o un llapis), la superfície (paper o cartró) o el tipus d'eina d'adquisició i la seva qualitat.

- **Diferents estils d'escriptura**: el mateix símbol és escrit de moltes formes, no sols depenent de l'autor, sinó també depenent del seu estat d'ànim, rapidesa en l'escriptura, context en el que es situa el símbol, etc.

L'etapa de preprocessament que es descrivia abans intenta per una banda eliminar o reduir el soroll en les imatges, i per altra, reduir les diferències en els

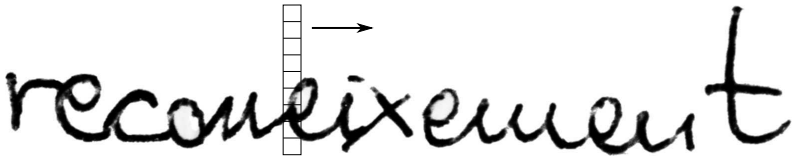


Figura 1.3: Finestra lliscant d'anàlisi en una imatge.

estils d'escriptura. La correcta segmentació del cos central del text és important per aplicar certes transformacions que tenen com a objectiu la reducció d'aquest segon problema.

Aquest projecte s'ha realitzat en un context de reconeixement de text manuscrit *offline* i continu.

Podem aplicar l'equació 1.1 que s'havia presentat anteriorment per solucionar el problema del reconeixement de text *offline*. Suposem que disposem d'una imatge I que conté cert text, llavors l'objectiu és trobar la seqüència de símbols $\hat{s} = s_1, s_2, \ldots, s_N$ que maximitza la probabilitat de que aquests símbols siguen els representats a la imatge I.

$$\hat{s} = \underset{s \in \Sigma^*}{\operatorname{argmax}}\, p(s|I) \tag{1.3}$$

En el cas del reconeixement de text *offline* i continu, típicament s'extreu una seqüència de vectors de característiques $\boldsymbol{x} = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_T$ a partir de la imatge I. Aquests vectors s'extreuen a partir d'una finestra lliscant d'anàlisi que travessa tota la imatge. Per a cada posició on es fixa aquesta finestra, s'extreu un vector de característiques. La figura 1.3 mostra aquesta finestra. Per tant, el reconeixement es reduiria a trobar la seqüència $\hat{s}$ que maximitza la probabilitat en 1.1

$$\hat{s} = \underset{s \in \Sigma^*}{\operatorname{argmax}}\, p(s|\boldsymbol{x}) = \underset{s \in \Sigma^*}{\operatorname{argmax}}\, p(\boldsymbol{x}|s)p(s) \tag{1.4}$$

Habitualment, tant en entorns de reconeixement de text com de veu, la distribució a posteriori $p(\boldsymbol{x}|s)$ es modela utilitzant Models Ocults de Markov (HMM, de l'anglès *Hidden Markov Models*) i Models de Mixtures Gaussianes (GMM, de l'anglès *Gaussian Mixture Models*) [HAJ90, Jel98, Gha01] i $p(s)$

es modela utilitzant models de llenguatge de n-grames [Jel98, KS87, MB01]. Aquesta és l'aproximació que s'ha utilitzat en aquest projecte.

Capítol 2

Tècniques de segmentació del cos central

En aquest capítol s'introdueixen les dues tècniques per a la segmentació del cos central comparades en aquest projecte. La segmentació del cos central bàsicament tracta de trobar la frontera entre els ascendents, el cos central i els descendents en una imatge que conté text. Aquesta operació és de vital importància per a diferents tècniques del preprocessament, però sense dubte a la que més afecta és a la normalització del text, que al mateix temps, és una de les etapes del preprocessament que més efecte té sobre la qualitat del reconeixement.

Els ascendents i descendents són aquelles parts de la imatge on no resideix molta informació sobre el text representat, mentre que el cos central és aquella part on la major part de la informació resideix. Per exemple, pensem en el cas de les lletres "o", "p" i "q". El que diferència aquestes lletres, és bàsicament la línia vertical que es troba a l'esquerra del cercle, en el cas de la "p", i a la dreta, en el cas de la "q". Però la longitud d'aquesta línia vertical no aporta massa informació per a diferenciar entre aquestes 3 lletres. Sols és necessari saber si hi ha una línia vertical i on està situada aquesta. Si es traça una línia horitzontal per davall del cercle central de cadascuna de les lletres, aquella part de la imatge que resideix per baix, és el que s'anomenen *descendents*. El mateix passa amb els símbols "o", "b" i "d", però ara la línia vertical es dirigeix cap amunt. En aquest cas, tot el que es troba per damunt del cercle central, serien els *ascendents*. Finalment, en l'exemple anterior, anomenaríem *cos central* al

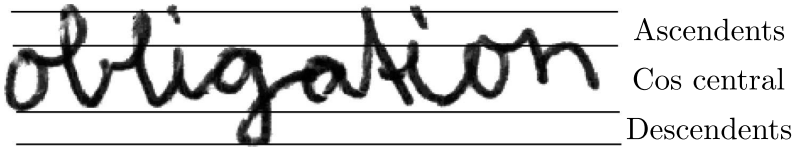


Figura 2.1: Zones d'ascendents, cos central i descendents d'una paraula. Imatge cedida per [Pas07]

cercle que queda en el centre dels símbols. En la figura 2.1 hi ha un exemple gràfic amb les zones diferenciades per a una paraula completa.

La línia que separa la zona d'ascendents del cos central s'anomena línia superior i la que separa el cos central dels descendents, inferior o línia base.

Una vegada s'ha segmentat el cos central de la imatge (s'han detectat les línies superiors i inferiors), el resultat es pot utilitzar per a normalitzar la grandària dels ascendents i descendents a una porció fixa del cos central, de manera que s'aconsegueix normalitzar l'altura de cadascun dels símbols representats a la imatge aconseguint que el cos central, on resideix la major part de la informació ocupe la major part de la imatge. Aquest procés de normalització és de vital importància per al reconeixement, ja que l'expressivitat dels models morfològics és limitada i resulta convenient que els vectors de característiques tinguen components el més significatives possibles. Per exemple, a la figura 1.3 pot observar-se que gran part de la finestra lliscant, de la qual s'obtindrà el vector de característiques, està ocupada per blanc, que no aporta cap informació a l'hora de classificar el símbol.

2.1 Aproximació heurística

L'algorisme que es descriu en aquesta secció per a detectar el cos central d'una imatge, fou presentat en [Rom05] i [Pas07] i consisteix en les següents operacions.

1. La línia es segmenta en diferents parts que estan completament separades per espais blancs. Cadascuna d'aquestes parts serà normalitzada per separat. És important tenir en compte que l'objectiu no és fer un

reconeixement del text de cadascuna d'aquestes parts, sinó segmentar el cos central del text de manera diferent en cadascuna d'aquestes per separat. La idea és que l'estil d'escriptura es conserva a cadascuna d'aquestes parts però pot variar d'una part a una altra.

2. Per a cada part del text, es detecten les línies base i superior de la imatge. Aquesta detecció es subdivideix en els següents passos:

   (a) Binarització de la imatge.

   (b) Suavitzat de la imatge original mitjançant l'algorisme Run-Length Smoothing Algorithm (RLSA).

   (c) Detecció de les vores superior i inferior de cada segment.

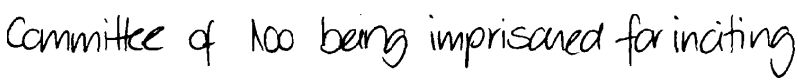
   (d) Càlcul de les rectes que millor s'ajusten a les fronteres superior i inferior, utilitzant la tècnica dels mínims quadrats.

La figura 2.2 mostra tots els passos de la normalització heurística per a una imatge que conté una línia de text de mostra.

Aquesta aproximació assumeix que les línies inferiors i superiors poden ser aproximades a una recta amb molt poc d'error. Aquesta assumpció pot no complir-se en la realitat i per tant, pot obtenir una mala aproximació a les línies inferiors i superiors. Un altre inconvenient d'aquest mètode és el llindar que s'utilitza per a fer l'esborronat en l'algorisme RLSA. En aquest algorisme s'uneixen dos píxels negres en una mateixa fila sempre que la separació siga menor que un cert llindar. Aquest llindar pot fer-se fixe o relatiu a la imatge. Siga com siga, si no s'ajusta bé aquest llindar pot ocasionar problemes com els de la figura 2.3, que alhora fan malbé les diferents etapes del preprocessament que depenen de la segmentació del cos central, com per exemple la normalització de la grandària.

2.2 Aproximació utilitzant aprenentatge supervisat

Aquesta aproximació a la segmentació del cos central del text que fa ús d'aprenentatge supervisat fou presentada en [GMBZMB08] i fa ús d'una xarxa neuronal multicapa per a classificar certs punts de la imatge en cinc classes: ascendent, línia superior, línia inferior, descendent i altres.

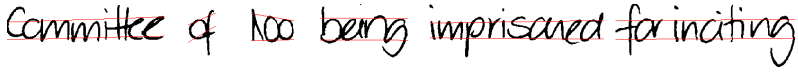


(a) Imatge original.

(b) Resultat de l'algorisme RLSA.

(c) Resultat de la detecció de la vora superior.

(d) Resultat de la detecció de la vora inferior.



(e) Resultat final de la segmentació.

Figura 2.2: Resultat de cada pas de la segmentació del cos central utilitzant l'algorisme heurístic.

Per tal d'utilitzar aquesta tècnica s'han d'extreure un conjunt de punts extrems locals a partir del contorn de la imatge. Aquests punts s'obtenen de la següent manera.

1. Per a cada columna de la imatge, es busquen els punts de la frontera entre un píxel de fons (típicament blancs) i un píxel d'un símbol (típicament negres). D'aquesta manera s'obté el contorn de la imatge.

2. Es desplaça una finestra d'anàlisi sobre el contorn de la imatge i es seleccionen els píxels del contorn màxims i mínims locals. Un punt màxim és aquell en el qual la resta de píxels del contorn veïns es situen per davall. Un mínim, en el que es situen per sobre.

Una vegada s'han obtingut els extrems locals, es situa una finestra de $W_w \times H_w$ píxels centrada en cada punt. A aquesta finestra se li aplica una distorsió d'ull de peix, de manera que els punts a prop del lloc d'interés queden

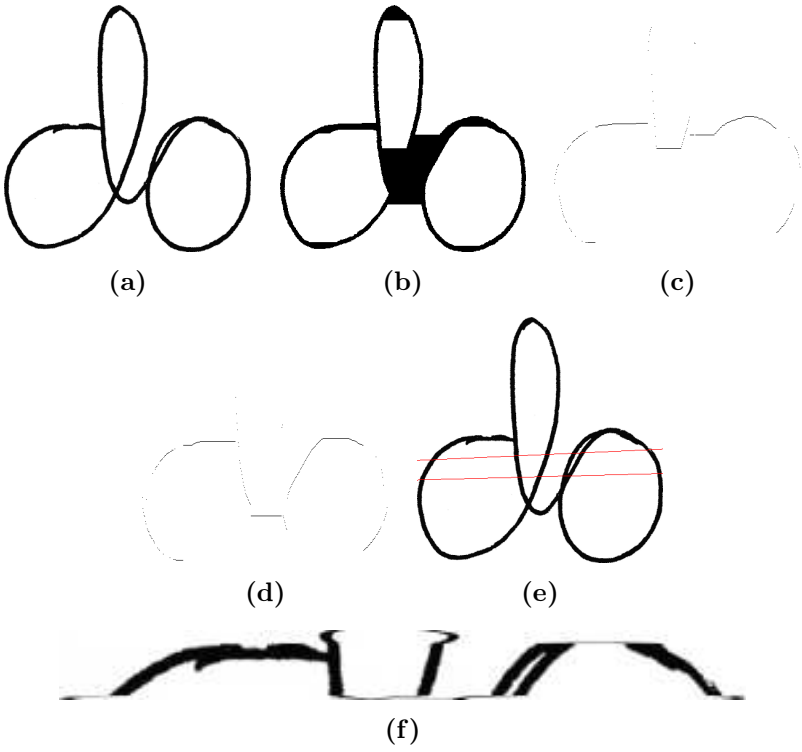


Figura 2.3: Exemple d'una segmentació dolenta i els seus efectes. La figura 2.3a mostra la imatge original, la figura 2.3b el resultat de l'algorisme RLSA, les figures 2.3d i 2.3c les vores inferior i superior detectades, la 2.3e mostra les línies obtingudes a partir de les vores i finalment la figura 2.3f el resultat de la normalització.

ressaltats sobre aquells més llunyans i finalment es redueix a una grandària de $W_f \times H_f$ píxels. Aquesta serà l'entrada a la xarxa neuronal.

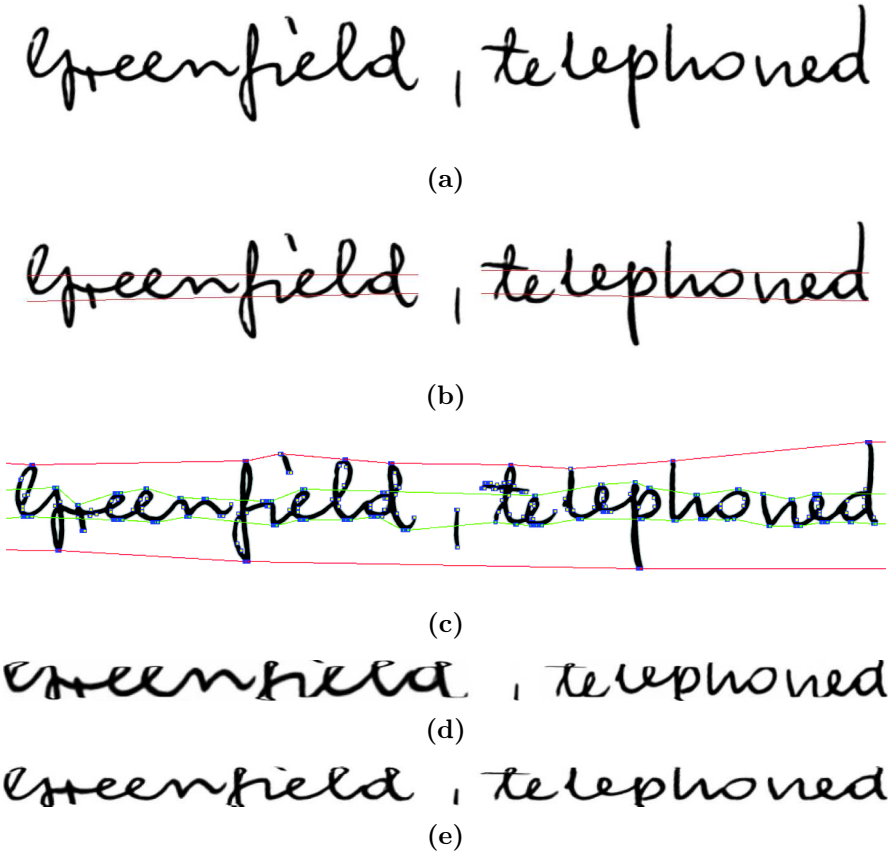
Per a entrenar la xarxa neuronal, s'utilitzen imatges amb els seus extrems locals classificats manualment, de manera que la xarxa neuronal aprèn a classificar un extrem local en una de les cinc classes mencionades anteriorment. Alternativament, per a facilitar l'entrenament del sistema, un humà pot corregir els punts classificats automàticament per un altre sistema més rudimentari per a la detecció de les línies de referència [GMBZMB08]. Els detalls sobre

l'entrenament de xarxes neuronals poden consultar-se en diverses fonts sobre aprenentatge automàtic [DH73, Bis06, Mur12].

Per a segmentar el cos central d'una imatge de *test*, es classifiquen tots els punts utilitzant la xarxa neuronal entrenada anteriorment. Com que la classificació feta per la xarxa neuronal pot ser sorollosa i pot classificar com a punts de les línies de referència píxels que en realitat no ho són, s'aplica un filtre a l'eixida de la xarxa neuronal de manera que tots els punts que són classificats com a "ascendents", "descendents", "línia superior" o "línia base" i no superen un cert llindar en la probabilitat emesa per la xarxa neuronal, són finalment classificats com a "altres". Llavors tots els punts pertanyents a una mateixa classe són utilitzats per a construir una polilínia per a cadascuna de les quatre línies de referència (els píxels classificats com "altres" no s'utilitzen).

La figura 2.4c mostra el resultat d'aplicar l'anterior algorisme a una imatge que conté text manuscrit. Els punts marcats són els extrems locals classificats, les línies delimiten les zones d'ascendents i descendents.

Les figures 2.4b i 2.4c mostren les diferències en les assumpcions que fan cadascun dels dos mètodes explicats. La major informació que aporta la detecció de les línies de referència utilitzant el mètode supervisat pot oferir clarament una millora significativa a l'hora d'aplicar el procés de normalització a la imatge. En els següents capítols s'explicarà com s'han quantificat aquestes diferències per demostrar que, efectivament, l'ús del mètode supervisat millora considerablement el reconeixement de text manuscrit.



(a)

(b)

(c)

(d)

(e)

Figura 2.4: Imatge original (2.4a), segmentació del cos central segons l'aproximació heurística i basada en aprenentatge supervisat (2.4b, 2.4c) i resultat de la normalització segons cada aproximació (2.4d, 2.4e). Figures 2.4a, 2.4c i 2.4e cedides per [GMBZMB08].

Capítol 3

Corpus

En aquest capítol es detallen les dades utilitzades per a realitzar els experiments que quantifiquen les diferències entre els dos mètodes descrits en el capítol 2.

3.1 IAMDB

El corpus IAMDB[1] fou recopilat pel grup d'investigació *Computer Vision and Artificial Intelligence* (FKI) dins del *Institute of Computer Science an Applied Mathematics* (IAM), a l'Universitat de Berna. El corpus és d'accés gratuit per a propòsits de recerca i és un dels més utilitzats per al reconeixement de text manuscrit. La primera versió es presentà en la ICDAR (International Conference of Document Analysis and Recognition) el 1999 [MB99]. El corpus és una transcripció manual del corpus Lancaster-Oslo-Bergen (LOB), descrit a la secció 3.3. Diferents paràgrafs del LOB es repartiren a un grup de persones que reescriviren el text manualment sense cap tipus de restricció en quant al tipus, estil o eina d'escriptura. En 2002, el text fou segmentat per línies i per paraules aïllades [ZB02] i presentada l'última versió en la revista IJDAR [MB02]. Els experiments realitzats en aquest treball utilitzaren la versió del corpus segmentada per línies.

La taula 3.1 conté les estadístiques de la versió de 2002 del corpus IAMDB, que ha sigut utilitzada. El corpus original compta amb dos conjunts de validació. Per a aquest projecte, a l'igual que han fet altres autors [BB08a, GLF+09,

[1] `http://www.iam.unibe.ch/fki/databases/iam-handwriting-database`

Conjunt	Línies	Escriptors
Train	6161	283
Validation 1	900	46
Validation 2	940	43
Test	1861	128
Total	9862	500

Taula 3.1: Estadístiques del corpus IAMDB original.

Conjunt	Línies	Escriptors
Train	6161	283
Validation	920	56
Test	2781	161
Total	9862	500

Taula 3.2: Estadístiques de la partició feta a partir del corpus IAMDB original.

EBCBGMZM11], s'ha utilitzat el segon conjunt de validació per a ampliar els conjunts de test i el primer de validació, de manera que la distribució final utilitzada per a l'entrenament, validació i test dels sistemes ha sigut l'expresada en la taula 3.2. Aquests conjunts són totalment disjunts i on cada escriptor ha participat únicament en un dels conjunts. El corpus IAMDB fou utilitzat per a entrenar els models morfològics del reconeixedor de text.

La figura 3.1 conté algunes de les imatges que formen part de la versió corpus IAMDB utilitzada.

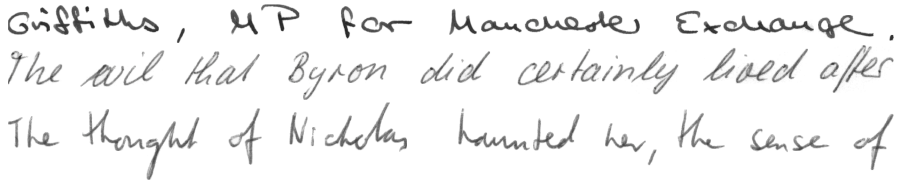


Figura 3.1: Exemples de línies de text extretes del corpus IAMDB.

3.2 Brown

El Corpus Estàndard d'Anglès Americà del Present, o simplement corpus Brown, va ser presentat originalment el 1961[FK79] per la Universitat de Brown i conté 500 textos d'aproximadament 2000 paraules cadascun, escrites en anglès americà actual i un total de 1.014.312 paraules entre tots els textos. El corpus conté informació sobre les categories de cada paraula en els textos. La taula 3.3 conté un resum dels textos continguts en el corpus Brown. Aquest corpus fou utilitzat per a construir el model del llenguatge utilitzat en els experiments.

3.3 Lancaster-Oslo-Bergen

El corpus Lancaster-Oslo-Bergen (LOB) fou recopilat i presentat el 1986[JAG86] per investigadors de la Universitat de Lancaster, la Universitat d'Oslo i el *Norwegian Computing Centre for the Humanities*, en Bergen. El corpus fou recopilat com una alternativa en anglès britànic al corpus de la Universitat de Brown, descrit a la secció 3.2, que fou desenvolupat a partir de textos en anglès americà. El LOB conté 500 textos d'unes 2000 paraules aproximadament i al voltant d'un milió de paraules en tot el corpus. Cada paraula del corpus fou anotada posteriorment en categories. La taula 3.3 conté un resum dels textos continguts en el corpus LOB. Part d'aquest corpus fou utilitzat per a construir el model del llenguatge dels experiments. S'exclogueren aquells texts que s'havien utilitzat per transcriure les línies de text de *test* del corpus IAMDB.

3.4 Wellington

El corpus Wellington d'Anglès Neozelandès escrit, o corpus Wellington, fou presentat el 1993[Bau93] per la Universitat Victoria de Wellington, a Nova Zelanda, i fou desenvolupat a partir de textos escrits en l'anglès utilitzat a Nova Zelanda. El corpus es desenvolupà per a fer-lo comparable als corpus de LOB i Brown descrits anteriorment (seccions 3.2 i 3.3). Conté també 500 textos de diferents categories, d'unes 2000 paraules cada text i al voltant d'un milió de paraules en tot el corpus. La taula 3.3 conté un resum dels tres corpus utilitzats. Aquest corpus fou utilitzat per a construir el model del llenguatge utilitzat en els experiments.

Categoria	Descripció	Brown	LOB	Wgton
A	Premsa: reportatges	44	44	44
B	Premsa: editorial	27	27	27
C	Premsa: ressenyes	17	17	17
D	Religió	17	17	17
E	Habilitats, oficis i aficions	36	38	38
F	Tradició popular	48	44	44
G	Belles lletres, biografia, assajos	75	77	77
H	Miscel·lània	30	30	30
J	Escrits científics	80	80	80
K	Ficció general	29	29	29
L	Ficció de misteri i policiaca	24	24	24
M	Ciència Ficció	6	6	6
N	Ficció d'aventures i de l'oest	29	29	29
P	Històries d'amor i romàntiques	29	29	29
R	Humor	9	9	9
Total		500	500	500

Taula 3.3: Resum dels textos continguts en els corpus Brown, LOB i Wellington.

Capítol 4

Experimentació

La principal tasca d'aquest projecte fou la de dissenyar els experiments de manera que els resultats foren el més significatius possible i que l'única diferència que hi hagués entre el reconeixement emprant els dos mètodes fou la de la segmentació del cos central del text en les imatges. D'aquesta manera es pot conèixer exactament quines són les diferències quantitatives entre els dos mètodes, calculant la diferència en l'error de reconeixement obtingut a l'emprar les dues alternatives.

A banda de la publicació original on s'explicava l'aproximació que utilitza aprenentatge supervisat per a la segmentació del text [GMBZMB08], els mateixos autors publicaren a la revista *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), el 2011, una comparativa entre dos estratègies per a la modelització dels models morfològics. Una basada únicament en HMM i l'altra un híbrid entre HMM i ANN [EBCBGMZM11]. L'interessant d'aquesta publicació per a l'objectiu d'aquest projecte és que utilitzava la segmentació del cos central supervisada (secció 2.2) en ambdós casos i donava més detalls sobre l'entrenament complet del reconeixedor (model de llenguatge, models morfològics, etc) que no pas l'article original on s'explica la segmentació del cos central utilitzant aprenentatge supervisat.

El disseny dels experiments ací descrits busca reproduir els resultats publicats en [EBCBGMZM11] del reconeixedor basat únicament en HMM utilitzant la segmentació del cos supervisada i comparar-los en un altre reconeixedor basat també en HMM i utilitzant la segmentació heurística. Malauradament,

molts dels detalls de l'entrenament que dugué als resultats publicats eren desconeguts, així com els detalls d'implementació exactes de les ferramentes utilitzades.

Per solucionar aquests contratemps i intentar reproduir de la manera més exacta possible els resultats, els autors de la publicació ens proveïren de les característiques extretes a partir de les línies del corpus IAMDB, així com del model de llenguatge, construït a partir dels corpus de Brown, LOB i Wellington.

4.1 Mesura d'avaluació

Per a quantificar el comportament dels dos sistemes comparats en la tasca de reconeixement en el corpus d'IAMDB s'utilitza el *Word Error Rate* (WER). Aquesta mesura d'error compara els errors en les paraules en la transcripció del sistema amb la transcripció de referència.

$$\text{WER} = \frac{\text{Insercions} + \text{Substitucions} + \text{Esborrats}}{\text{Nombre de paraules en la referència}} \cdot 100 \tag{4.1}$$

Un WER igual a zero sols s'aconsegueix si la transcripció produïda pel sistema és igual a la transcripció de referència. En la fórmula 4.1, "Insercions", "Substitucions" i "Esborrats" fa referència al nombre mínim de paraules que han de sofrir una d'aquests transformacions en la transcripció obtinguda pel sistema per tal d'obtenir la transcripció de referència. Aquesta mesura d'error és molt semblant al *Character Error Rate* (CER), on les operacions es compten a nivell de símbol en lloc de a nivell de paraula. Cal observar que és possible obtenir un WER major al 100% si la transcripció té més errors que paraules hi han en la transcripció de referència. Això és degut a que el sistema pot produir una transcripció amb més paraules que el text de referència.

4.2 Preprocessament de les imatges

En els dos reconeixedors entrenats, s'aplicà el següent preprocessament a les línies de text d'IAMDB.

1. Neteja de la imatge. L'objectiu d'aquesta etapa és la de netejar el soroll en les imatges utilitzades. En el cas de la segmentació basada en

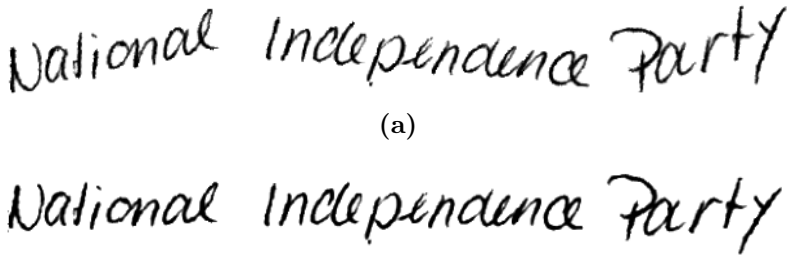
aprenentatge supervisat, prèviament s'eliminava el soroll utilitzant una xarxa neuronal multicapa que donat una graella de la imatge centrada en un píxel, obtenia el valor del píxel central restaurat. Aquesta xarxa neuronal s'entrenava a partir d'imatges amb soroll i la seva imatge equivalent sense soroll. Més detalls sobre aquest mètode poden trobar-se en [GMBZMB08, EBCBGMZM11]. En el cas de la segmentació heurística, el mètode utilitzat per a la neteja de la imatge és l'explicat en [Pas07]. En primer lloc, la imatge en blanc i negre es normalitza a l'escala $[0, 255]$, siguent 0 el píxel més obscur de la imatge original i 255 el més clar. D'aquesta manera s'augmenta el contrast a la imatge original. Després s'aplica un filtre mediana, amb el qual el valor de cada píxel és determina com la mediana dels valors d'una finestra centrada en aquest píxel.

2. Correcció del *slope*. L'objectiu és corregir l'angle que forma la línia de text amb la línia horizontal de la imatge (veure figura 4.1). En el cas del reconeixedor que feia ús de la segmentació heurística, la correcció es fa seguint la tècnica descrita en [Pas07]. Aquesta tècnica consisteix en girar la imatge provant diferents angles de rotació, trobar la seva projecció horitzontal i escollir aquell angle que produïsca una columna amb el valor acumulat dels píxels major. En el cas del reconeixedor que feia ús de la segmentació basada en aprenentatge automàtic, la tècnica ve descrita en l'article [GMBZMB08]. Aquesta tècnica consisteix en trobar la línia de referència inferior, amb un mètode semblant al descrit en 2.2, calcular l'angle entre aquesta línia i la base de la imatge i corregir aquest angle rotant la imatge.

3. Correcció del *slant*. L'objectiu d'aquesta etapa és la de corregir l'angle que forma cada lletra amb la línia vertical de la imatge (veure figura 4.2). En ambdós casos la tècnica utilitzada és la descrita en [PTV04, Rom05, Pas07]. Aquesta tècnica consisteix en transformar la imatge aplicant una transformació *shear* aplicant diferents angles, calcular la projecció vertical de la imatge i escollir aquell angle amb el qual s'obté una major desviació típica en la projecció.

4. Segmentació del cos central. En un dels reconeixedors entrenats s'utilitza la segmentació heurística descrita a la secció 2.1 i en el segon la segmentació supervisada descrita a la secció 2.2.

5. Normalització de l'altura d'ascendents i descendents. Aquesta etapa té



(a)

(b)

Figura 4.1: Exemple de la correcció del *slope* [Pas07].

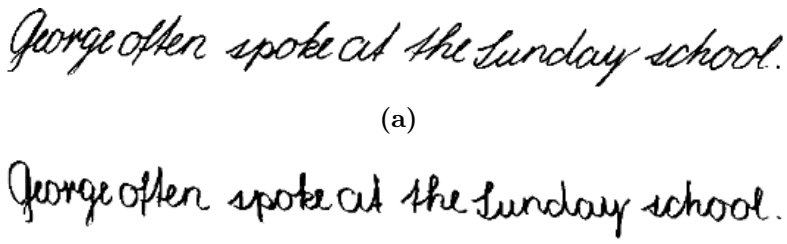
com a objectiu normalitzar l'altura dels símbols. En ambdós reconeixedors, la tècnica és la mateixa. Una vegada segmentat el cos central, per a cada columna de píxels, la zona d'ascendents s'escala a un 20% de la imatge i la de descendents a un 10%.

6. Extracció de característiques. En ambdós reconeixedors, s'utilitza la tècnica descrita en [TJG+04, Pas07]. Es recorre la imatge utilitzant una finestra lliscant amb 20 graelles i s'extreu per a cada graella el nivell mitjà de gris normalitzat, la derivada horitzontal del nivell de gris i la derivada vertical del nivell de gris. De manera que de cada finestra d'anàlisi s'obtenen 60 característiques.

Tot i que les etapes de neteja de la imatge i correcció del *slope* també varien, proves preliminars dutes a terme pels autors que desenvoluparen els mètodes basats en xarxes neuronals, mostren que les diferències entre els mètodes utilitzats en aquests casos no són significatives. D'aquesta manera, l'única variable amb un efecte significatiu sobre el resultat del reconeixement que es veu alterada és la segmentació del cos central.

4.3 Model del llenguatge

El model del llenguatge utilitzat en ambdós sistemes fou un proveït per els autors de l'article on es descriu el mètode de segmentació basat en xarxes neuronals i que és semblant al que ells utilitzaren en la publicació [EBCBGMZM11]. De nou, l'idea d'utilitzar un model del llenguatge el més semblant possible era



(a)

(b)

Figura 4.2: Exemple de la correcció del *slant* [Pas07].

poder reproduir els resultats obtinguts en la publicació.

El model del llenguatge està basat en n-grames, un modelatge molt popular en la literatura a l'hora de construir models de llenguatge [MS99]. En aquest cas, el model de llenguatge estava format per bigrames entrenats a partir de textos de tres corpus distints: el corpus Brown (secció 3.2); el corpus LOB, excloent aquells textos que contenien línies incloses en el conjunt de test del corpus IAMDB (secció 3.3) i el corpus Wellington (secció 3.4). La ferramenta utilitzada per a la generació del model del llenguatge fou la popular SRI Language Modeling Toolkit [S$^+$02], utilitzant el suavitzat de Kneser-Ney modificat [CG99].

S'utilitzaren 51560 frases del corpus LOB, 51763 del corpus Brown i 20592 del Wellington. Per tal de modelar el fet que el text en les imatges de IAMDB està fragmentat en línies, les frases anteriors foren fragmentades aleatòriament per tal de simular les línies i s'obtingueren 400000 d'aquests fragments. Finalment, s'afegiren també les 6161 línies del conjunt d'entrenament del corpus IAMDB.

De totes les línies anteriors, sols s'utilitzaren les 20000 paraules més comunes per tal de construir el model del llenguatge i fer un sistema amb un diccionari obert, de manera que es simula un entorn sense restriccions en el vocabulari. Un altre factor important d'aquest model del llenguatge és el fet que totes les lletres foren convertides a minúscules.

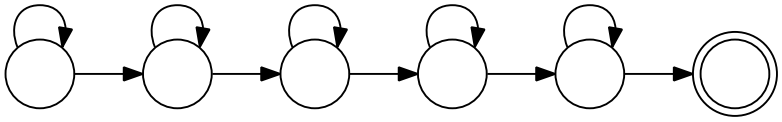


Figura 4.3: Topologia esquerra-dreta estricta d'un Model de Markov.

4.4 Models morfològics

L'habitual combinació de Models Ocults de Markov i Models de Mixtures Gaussianes fou emprada per tal de modelar el model morfològic de cada símbol. En els dos sistemes comparats, s'utilitzà la ferramenta Hidden Markov Model Toolkit (HTK) [YWB93] per construir aquests models. Les distribucions de Gauss en cada mixtura tenien una matriu de covariàncies diagonal i els Models Ocults de Markov tenien una topologia esquerra-dreta estricta (veure figura 4.3). En total s'entrenaren 78 models morfològics, un per cada símbol ocorregut en les dades d'entrenament d'IAMDB.

Els models morfològics utilitzats en el cas de la segmentació utilitzant xarxes neuronals foren entrenats amb els mateixos paràmetres que els utilitzats en la publicació anterior. De nou, l'objectiu era que els resultats obtinguts foren el més semblants possibles als publicats. Aquest eren models de 8 estats i una mixtura de 64 gaussianes en cada estat. Aquesta configuració fou estimada a partir de provar diferents combinacions de paràmetres i escollint aquella amb un millor comportament en el corpus de validació.

Pel que fa al sistema que utilitzava la segmentació heurística, també es provaren diferents combinacions del nombre de gaussianes en els GMM i el nombre d'estats en els HMM. La taula 4.1 mostra els resultats d'aquesta estimació en un subconjunt de les línies de validació. En els dos sistemes, el paràmetre *Grammar Scale Factor* (GSF) es fixà a 40 i el *Word Insertion Penalty* (WIP) es fixà a 0. En [EBCBGMZM11] es comprova que amb un valor GSF d'entre 40 i 50 s'obtenien els millors resultats en el reconeixement, amb diferències no significatives estadísticament entre ells. Per a l'execució d'aquest experiment en el qual es buscava el nombre òptim d'estats en el HMM i del nombre de components en la Mixtura de Gaussianes, es van extreure 300 línies escollides

		Components de la mixtura		
		16	32	64
Nombre d'estats	7	54.91	46.75	42.84
	8	54.24	46.08	**42.21**
	9	54.41	47.73	44.53

Taula 4.1: WER en un subconjunt de validació utilitzant diferents paràmetres dels HMM en el sistema amb segmentació heurística. *Beam* igual a 500.

aleatòriament de les 920 del conjunt de validació i es va escollir un llindar de poda *beam* igual a 500 per a que el temps de reconeixement no fóra massa elevat.

S'observa en la taula 4.1 que la millor configuració de les provades fou aquella amb 8 estats per cada HMM i un total de 64 components gaussianes en els GMM. Aquesta fou la configuració utilitzada per conduir els experiments finals.

4.5 Postprocessament al reconeixement

Una vegada reconegut el text se li aplica un filtre que intenta normalitzar les transcripcions. Aquest filtre va ser també utilitzat en la publicació [EBCBGMZM11] i fou proveït pels autors de la mateixa per tal d'obtenir uns resultats el més similar possibles amb els originals de la publicació.

Aquest filtre realitza les següents operacions sobre la transcripció.

- Els signes dobles de cita (“, ”) i les barres verticals (|) són substituïdes per signes simples de cita (‘, ’).
- Les contraccions de l'anglès (p. ex: *'s*, *'d*, *'nt*, etc.) són precedides amb un espai en blanc.
- Els signes de puntuació (punts, comes, dos punts, punts i coma, etc.) són precedits amb un espai en blanc.
- Els signes de cita i els parèntesis són seguits d'un espai en blanc.

	Validació		Test	
	No Filtre	Filtre	No Filtre	Filtre
Heurística	40.93	38.74	46.85	45.54
Apr. Supervisat	**35.18**	**32.99**	**41.15**	**40.08**
Diferència	-5.75	-5.75	-5.70	-5.46

Taula 4.2: WER de les dues alternatives comparades. *Beam* igual a 800.

- Els espais en blanc seguits són substituïts per un únic espai en blanc.

4.6 Resultats finals

La taula 4.2 mostra el WER obtingut amb les dues alternatives estudiades en aquest projecte i la seva diferència. Per a aquestes proves els paràmetres escollits foren els detallats anteriorment i amb un llindar de poda *beam* igual a 800, el mateix que el que utilitzaren els autors de la publicació on es detalla l'aproximació basada en aprenentatge supervisat. Les columnes etiquetades com a "No Filtre" corresponen a les mesures d'error obtingudes abans d'aplicar el filtre descrit a la secció 4.5, les columnes etiquetades com a "Filtre" corresponen a les mesures d'error obtingudes una vegada el filtre s'havia aplicat a les transcripcions.

Pot observar-se que la diferència en tots els casos està al voltant de -5.67, sent aquesta per al conjunt de *test* de -5.46 punts WER. Aquesta és la diferència entre el WER obtingut amb el sistema que utilitzava una segmentació basada en aprenentatge supervisat i el sistema basat en una aproximació heurística. Al ser aquesta diferència negativa en tots els casos, indica que el sistema utilitzant l'enfocament d'aprenentatge supervisat millora sempre a l'enfocament heurístic, per als experiments duts a terme sobre el corpus IAMDB. Aquest decrement de -5.46 punts absoluts en el WER, significa una millora (decrement) relativa del 11.99% respecte al WER del sistema heurístic.

Capítol 5

Conclusions

La taula 5.1 situa els resultats obtinguts en aquest treball en comparació amb resultats obtinguts prèviament en la mateixa tasca de reconeixement.

En aquesta taula s'observa que els resultats obtinguts a partir del sistema que empra una segmentació del cos central basada en aprenentatge supervisat no correspon als resultats obtinguts en la publicació. Això és degut a que, a pesar d'intentar replicar els experiments amb el major detall possible i comptar amb la col·laboració dels autors de l'article, hi ha nombrosos detalls sobre l'experimentació publicada que eren desconeguts i/o no van poder ser replicats. D'ací aquesta petita diferència en les mesures obtingudes.

Més important encara, s'observa que l'utilització d'una segmentació del cos central del text manuscrit utilitzant aprenentatge supervisat redueix un

	Validació	Test
Bertolami et al. [BB08b]	**30.98**	**35.52**
España Boquera et al. [EBCBGMZM11]	32.80	38.80
Seg. Supervisada	32.99	40.08
Seg. Heurística	38.74	45.54

Taula 5.1: Comparació del WER de les alternatives estudiades amb altres publicacions.

11.99% el WER respecte a l'alternativa tradicional basada en un enfocament heurístic, en la tasca de reconeixement del corpus IAMDB, un important corpus utilitzat a l'hora de mesurar els errors en les transcripcions dels sistemes de reconeixement automàtic de text manuscrit.

Queda per tant comprovada la hipòtesi que es presentava al descriure les diferències entre els dos enfocaments per a la segmentació del cos central del text manuscrit: l'ús d'una tècnica basada en aprenentatge supervisat millora significativament la segmentació heurística. Aquesta millora ha sigut quantificada satisfactoriament a pesar de les dificultats a l'hora de reproduir els detalls de les experimentacions publicades amb anterioritat i sols queda per tant fer front a la qüestió de si aquesta millora compensa el cost econòmic i temporal d'utilitzar aquesta tècnica basada en aprenentatge supervisat per a sistemes reals.

Bibliografia

[Bau93] L. Bauer. *Manual of information to accompany the Wellington corpus of written New Zealand English*. Department of Linguistics, Victoria University of Wellington, 1993.

[BB08a] R. Bertolami and H. Bunke. Ensemble methods to improve the performance of an English handwritten text line recognizer. *Arabic and Chinese Handwriting Recognition*, pages 265–277, 2008.

[BB08b] R. Bertolami and H. Bunke. Hidden Markov model-based ensemble methods for offline handwritten text line recognition. *Pattern Recognition*, 41(11):3452–3460, 2008.

[Bis06] C.M. Bishop. *Pattern Recognition and Machine Learning*, volume 4. Springer New York, 2006.

[BP63] Mr. Bayes and Mr. Price. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, 53:370–418, 1763.

[CG99] S.F. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4):359–393, 1999.

[DH73] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. John Wiley, New York, 1973.

[Dim57] T. L. Dimond. Devices for Reading Handwritten Characters. *Managing Requirements Knowledge, International Workshop on*, 0:232, 1957.

[EBCBGMZM11] S. España Boquera, M.J. Castro Bleda, J. Gorbe Moya, and F. Zamora Martínez. Improving offline handwritten text recognition with hybrid HMM/ANN models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):767–779, 2011.

[FK79] W.N. Francis and H. Kucera. Brown corpus manual. *Letters to the Editor*, 5(2):7, 1979.

[Gha01] Z. Ghahramani. An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(01):9–42, 2001.

[GLF+09] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(5):855–868, 2009.

[GMBZMB08] J. Gorbe-Moya, S. España Boquera, F. Zamora-Martínez, and M. J. Castro Bleda. Handwritten Text Normalization by using Local Extrema Classification. In Alfons Juan-Císcar and Gemma Sánchez-Albaladejo, editors, *PRIS*, pages 164–172. INSTICC PRESS, 2008.

[Gol14] Hyman Eli Goldberg. Controller. US Patent 1,117,184, 1914.

[HAJ90] XD Huang, Y. Ariki, and MA Jack. Hidden Markov Models for Speech Recognition, 1990.

[Har62] LD Harmon. Handwriting reader recognizes whole words. *Electronics (August 1962)*, 1962.

[JAG86] S. Johansson, E. Atwell, and G. Garside, R.and Leech. The tagged LOB Corpus: User's Manual. 1986.

[Jel98] F. Jelinek. *Statistical methods for speech recognition*. MIT press, 1998.

[KS87] M. Katz Slava. Estimation of probabilities from sparse data for the language model component of a speech recognition. *IEEE Transactions on Acoustics, Speech and Signal processing, ASSP-35*, pages 400–401, 1987.

[MB99] U.V. Marti and H. Bunke. A full English sentence database for off-line handwriting recognition. In *Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on*, pages 705–708. IEEE, 1999.

[MB01] U.V. Marti and H. Bunke. Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(01):65–90, 2001.

[MB02] U.V. Marti and H. Bunke. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[MS99] Christopher D. Manning and Hinrich Schuetze. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1 edition, June 1999.

[Mur12] K.P. Murphy. Machine Learning: a Probabilistic Perspective. 2012.

[Pas07] M. Pastor. *Aportaciones al Reconocimiento Automático de Texto Manuscrito*. Tesis doctoral en informàtica, Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, 2007.

[PTV04] M. Pastor, A. Toselli, and E. Vidal. Projection profile based algorithm for slant removal. *Image Analysis and Recognition*, pages 183–190, 2004.

[Rom05] V. Romero. *Mejora de la normalización de tamaño de texto manuscrito off-line*. Projecte final de carrera en informàtica, Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, 2005.

[S+02] A. Stolcke et al. SRILM - An extensible language modeling toolkit. In *Proceedings of the international conference on spoken language processing*, volume 2, pages 901–904, 2002.

[TJG+04] AH Toselli, A. Juan, J. González, I. Salvador, E. Vidal, F. Casacuberta, D. Keysers, and H. Ney. Integrated handwriting recognition and interpretation using finite-state models. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(04):519–539, 2004.

[YWB93] S. Young, PC Woodland, and WJ Byrne. HTK: Hidden Markov Model Toolkit V1.5, 1993.

[ZB02] M. Zimmermann and H. Bunke. Automatic segmentation of the IAM off-line database for handwritten English text. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, pages 35–39. IEEE, 2002.

Índex de figures

Índex de taules