



Classificazione firme statiche utilizzando i Hidden Markov Models

Alexandru PRIGOREANU

Università degli studi di Padova

Dipartimento di Matematica
Corso di laurea in Informatica

Relatore

Prof. Tullio VARDANEGA

Dicembre 13, 2013



Contenuti

1 Analisi

- Classificazione di firme statiche
- Processi generali
- Metodi di classificazione
- Casi d'uso

2 Progettazione

- Scelte effettuate
- Hidden Markov Models

3 Implementazione

4 Consuntivo

- Risultati
- Consuntivo

Classificazione di firme statiche

Obiettivo e difficoltà

- Obiettivo: Decidere se una firma è autentica/falsa



Classificazione di firme statiche

Obiettivo e difficoltà

- Obiettivo: Decidere se una firma è autentica/falsa
- Variazioni intrapersonali: Le firme personali possiedono grande variabilità, dovuta allo stato emotivo dei sottoscrittori, alla posizione di raccolta, ecc...

Alex Prigoreanu Alex Prigoreanu Alex Prigoreanu

Classificazione di firme statiche

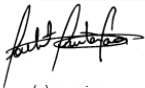
Obiettivo e difficoltà

- Obiettivo: Decidere se una firma è autentica/falsa
- Variazioni intrapersonali: Le firme personali possiedono grande variabilità, dovuta allo stato emotivo dei sottoscrittori, alla posizione di raccolta, ecc...
- Differenze interpersonali: Le firme di persone diverse possiedono caratteristiche elementari distinte



Terminologia

Tipi di falsificazione

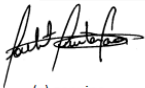


(a) genuine

Terminologia

Tipi di falsificazione

- Falsificazioni Casuali



(a) genuine

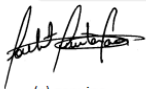


(b) random forgery

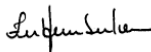
Terminologia

Tipi di falsificazione

- Falsificazioni Casuali
- Falsificazioni Semplici



(a) genuine



(b) random forgery

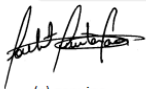


(c) simple forgery

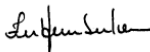
Terminologia

Tipi di falsificazione

- Falsificazioni Casuali
- Falsificazioni Semplici
- Falsificazioni Accurate



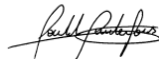
(a) genuine



(b) random forgery



(c) simple forgery

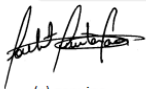


(d) skilled forgery

Terminologia

Tipi di falsificazione

- Falsificazioni Casuali
- Falsificazioni Semplici
- Falsificazioni Accurate



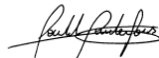
(a) genuine



(b) random forgery



(c) simple forgery



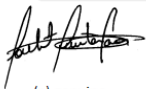
(d) skilled forgery

Valutazione della performance

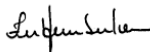
Terminologia

Tipi di falsificazione

- Falsificazioni Casuali
- Falsificazioni Semplici
- Falsificazioni Accurate



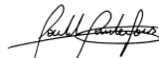
(a) genuine



(b) random forgery



(c) simple forgery



(d) skilled forgery

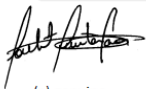
Valutazione della performance

- *False Acceptance Rate (FAR)*

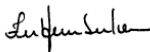
Terminologia

Tipi di falsificazione

- Falsificazioni Casuali
- Falsificazioni Semplici
- Falsificazioni Accurate



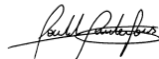
(a) genuine



(b) random forgery



(c) simple forgery



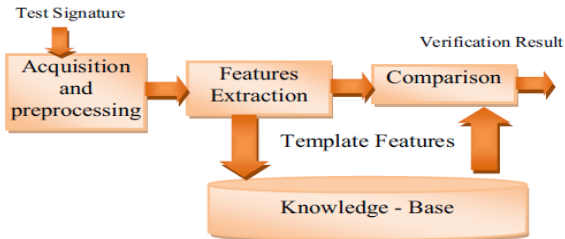
(d) skilled forgery

Valutazione della performance

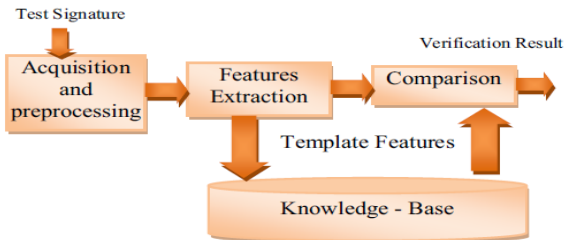
- *False Acceptance Rate (FAR)*
- *False Rejection Rate (FRR)*



Processi generali



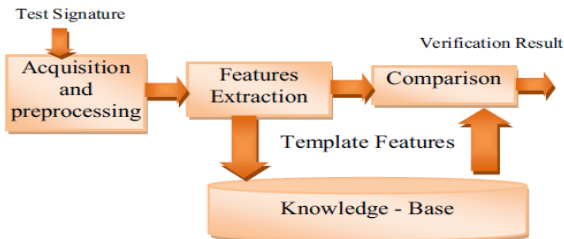
Processi generali



Preprocessings

- *Cropping*
- *Resizing*
- *Binarization*
- *Thinning*

Processi generali



Preprocessings

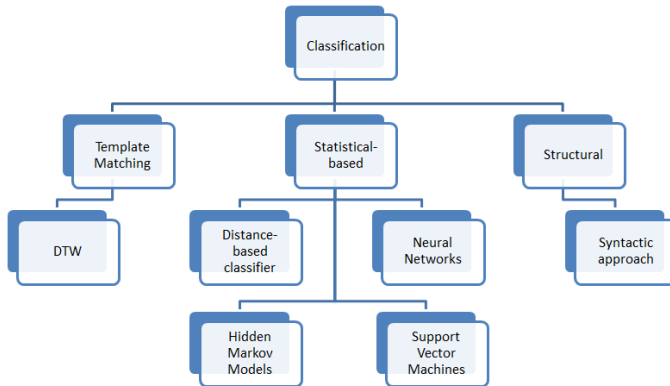
- *Cropping*
- *Resizing*
- *Binarization*
- *Thinning*

Features

- *Calibre*
- *Spacing*
- *Distribution of pixels*
- *Slant*

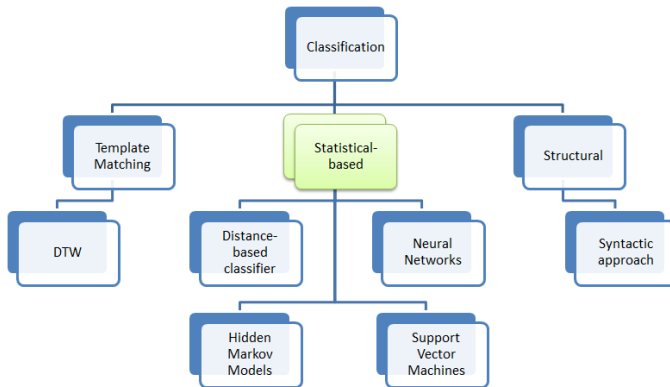


Metodi di classificazione



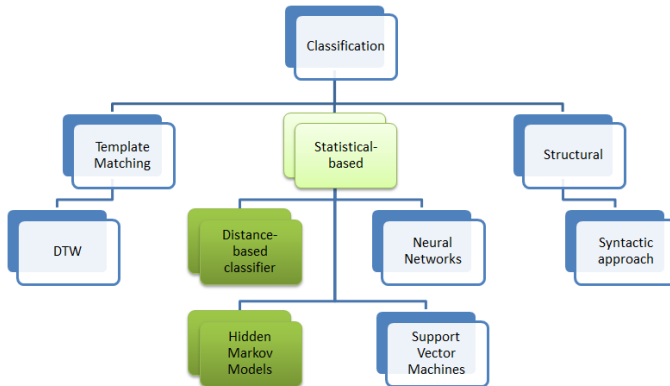


Metodi di classificazione

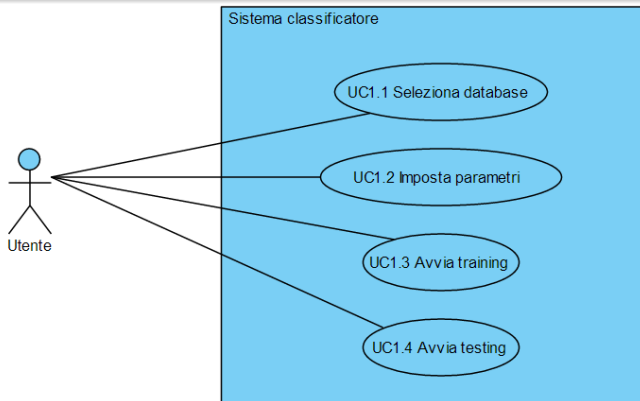




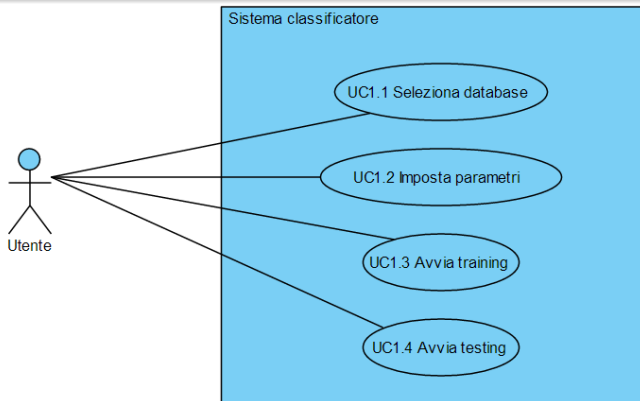
Metodi di classificazione



Casi d'uso



Casi d'uso



Requisito di qualità

- Garantire un'*accuracy* media del 80%



Scelte effettuate

Modello di ciclo di vita

- *Modello incrementale*



Scelte effettuate

Modello di ciclo di vita

- *Modello incrementale*

Strumenti

- *Java*
- *Eclipse*
- *Hidden Markov Models*



Scelte effettuate

Modello di ciclo di vita

- *Modello incrementale*

Librerie Java

- *ImageJ*
- *Jahmm*
- *JScience*
- *jhmm*

Strumenti

- *Java*
- *Eclipse*
- *Hidden Markov Models*



Scelte effettuate

Modello di ciclo di vita

- *Modello incrementale*

Librerie Java

- *ImageJ*
- *Jahmm*
- *JSscience*
- *jhmm*

Strumenti

- *Java*
- *Eclipse*
- *Hidden Markov Models*

Design Pattern

- *Model View Controller*
- *Composite*



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$
- una matrice $\Pi = \{\pi_i\} : \pi_i = P(q_1 = S_i)$



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$
- una matrice $\Pi = \{\pi_i\} : \pi_i = P(q_1 = S_i)$
- un insieme $V = \{V_1, V_2, \dots, V_M\}$ di simboli di osservazione



Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$
- una matrice $\Pi = \{\pi_i\} : \pi_i = P(q_1 = S_i)$
- un insieme $V = \{V_1, V_2, \dots, V_M\}$ di simboli di osservazione
- al t-esimo istante il processo emette uno fra i simboli a disposizione: $o_t \in \{V_1, V_2, \dots, V_M\}$

Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$
- una matrice $\Pi = \{\pi_i\} : \pi_i = P(q_1 = S_i)$
- un insieme $V = \{V_1, V_2, \dots, V_M\}$ di simboli di osservazione
- al t-esimo istante il processo emette uno fra i simboli a disposizione: $o_t \in \{V_1, V_2, \dots, V_M\}$
- una matrice $B = \{b_j(k)\} : b_j(k) = P(o_t = k \mid q_t = j)$

Hidden Markov Models

Definizione

$$\lambda = (A, B, \pi)$$

- un insieme $S = \{S_1, S_2, \dots, S_N\}$ di stati nascosti
- una matrice $A = \{a_{ij}\} : a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i)$
- una matrice $\Pi = \{\pi_i\} : \pi_i = P(q_1 = S_i)$
- un insieme $V = \{V_1, V_2, \dots, V_M\}$ di simboli di osservazione
- al t-esimo istante il processo emette uno fra i simboli a disposizione: $o_t \in \{V_1, V_2, \dots, V_M\}$
- una matrice $B = \{b_j(k)\} : b_j(k) = P(o_t = k \mid q_t = j)$
- vale la *proprietà di Markov*:

$$P(q_{t+1} = S_j \mid q_t = S_i, q_{t-1} = S_k, \dots, q_1 = S_1) = P(q_{t+1} = S_j \mid q_t = S_i)$$



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*
- 2 *Decoding problem*



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*
- 2 *Decoding problem*
- 3 *Learning problem*



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*
- 2 *Decoding problem*
- 3 *Learning problem*

Tre soluzioni (utilizzando la Programmazione Dinamica)

- 1 *Forward algorithm*



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*
- 2 *Decoding problem*
- 3 *Learning problem*

Tre soluzioni (utilizzando la Programmazione Dinamica)

- 1 *Forward algorithm*
- 2 *Viterbi's algorithm*



Utilizzo dei HMM

Tre problemi (in genere intrattabili)

- 1 *Evaluation problem*
- 2 *Decoding problem*
- 3 *Learning problem*

Tre soluzioni (utilizzando la Programmazione Dinamica)

- 1 *Forward algorithm*
- 2 *Viterbi's algorithm*
- 3 *Baum-Welch algorithm*



Vettori di osservazione

Preprocessings

- Cropping
- Binarization
- Skeletonization
- Segmentation

Vettori di osservazione

Preprocessings

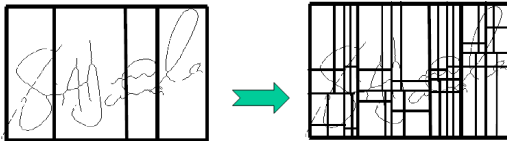
- Cropping
- Binarization
- Skeletonization
- Segmentation



Vettori di osservazione

Preprocessings

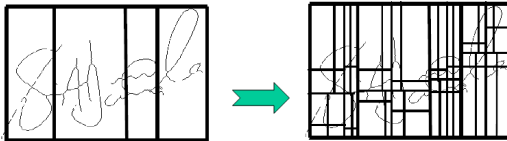
- Cropping
- Binarization
- Skeletonization
- Segmentation



Vettori di osservazione

Preprocessings

- Cropping
- Binarization
- Skeletonization
- Segmentation



Features

- Slant
- DCT

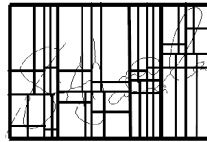
Vettori di osservazione

Preprocessings

- Cropping
- Binarization
- Skeletonization
- Segmentation

Features

- Slant
- DCT



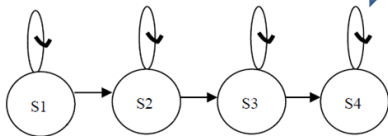
v01 v02	v03 v04	v05 v06	v07 v08
v09 v10	v11 v12	v13 v14	v15 v16
v17 v18	v19 v20	v21 v22	v23 v24
v25 v26	v27 v28	v29 v30	v31 v32
v33 v34	v35 v36	v37 v38	v39 v40
v41 v42	v43 v44	v45 v46	v47 v48
v49 v50	v51 v52	v53 v54	v55 v56
v57 v58	v59 v60	v61 v62	v63 v64
Ov1	Ov2	Ov3	Ov4

Inizializzazione HMM

Feature Extraction Output

v01	v02	v03	v04	v05	v06	v07	v08
v09	v10	v11	v12	v13	v14	v15	v16
v17	v18	v19	v20	v21	v22	v23	v24
v25	v26	v27	v28	v29	v30	v31	v32
v33	v34	v35	v36	v37	v38	v39	v40
v41	v42	v43	v44	v45	v46	v47	v48
v49	v50	v51	v52	v53	v54	v55	v56
v57	v58	v59	v60	v61	v62	v63	v64
Ov1		Ov2		Ov3		Ov4	

Model topology



* nr. firme **training set**

BW training set

validation set

KMeans Clustering

$A, \pi \mid B$

Modello inizializzato



Metrica Software	HmmGui	HmmTraining	HmmTesting
Complessità ciclomatica	2.50	5.50	2.33



Metrica Software	HmmGui	HmmTraining	HmmTesting
Complessità ciclomatica	2.50	5.50	2.33
Media parametri per metodo	0.77	0.60	0.65



Metrica Software	HmmGui	HmmTraining	HmmTesting
Complessità ciclomatica	2.50	5.50	2.33
Media parametri per metodo	0.77	0.60	0.65
Linee di codice (<i>LOC</i>) per metodo	13.69	26.83	12.43



Metrica Software	HmmGui	HmmTraining	HmmTesting
Complessità ciclomatica	2.50	5.50	2.33
Media parametri per metodo	0.77	0.60	0.65
Linee di codice (<i>LOC</i>) per metodo	13.69	26.83	12.43
Linee di commento per <i>LOC</i>	15.60	29.40	17.50

Verifica e Validazione

- Analisi
- Progettazione
- Implementazione
- Documentazione



Risultati

Numero di firme analizzate: 770

Numero di firme classificate: 770

True positives: 229

True negatives: 303

False positives: 82

False negatives: 156

Accuracy: 69,09%

FAR: 0.212987

FRR: 0.405195

Firme da controllare manualmente: 0
(0,00% delle firme totali)

OK



Risultati

=====

Numero di firme analizzate: 770

=====

Numero di firme classificate: 770

True positives: 229
True negatives: 303
False positives: 82
False negatives: 156

Accuracy: 69,09%

FAR: 0.212987
FRR: 0.405195

=====

Firme da controllare manualmente: 0
(0,00% delle firme totali)

=====

OK

Cause

- Scelte progettuali
- Overfitting
- Tempo



Risultati

Numero di firme analizzate: 770

Numero di firme classificate: 770

True positives: 229
True negatives: 303
False positives: 82
False negatives: 156

Accuracy: 69,09%

FAR: 0.212987
FRR: 0.405195

Firme da controllare manualmente: 0
(0,00% delle firme totali)

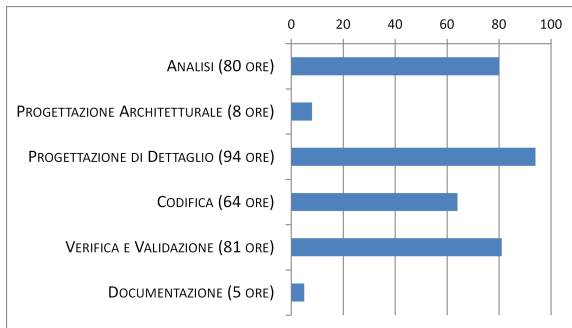
OK

Cause

- Scelte progettuali
- Overfitting
- Tempo

Possibili sviluppi

- Features
- Continuous HMM
- HMM scoring
- Pseudocounts
- Regularization
- Weighting
- Genetic Algorithms



Altre informazioni quantitative





Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

Innanzitutto ringrazio il professor *Vardanega* per avermi seguito durante lo *stage* e per avermi fornito utili consigli per la stesura della relazione.



Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

Innanzitutto ringrazio il professor *Vardanega* per avermi seguito durante lo *stage* e per avermi fornito utili consigli per la stesura della relazione.



Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

In secondo luogo ringrazio i professori del *Corso di Laurea in Informatica* dai quali ho avuto l'opportunità di ampliare le mie (modeste) conoscenze. Ringrazio i seguenti professori e professoresse (in ordine di apparizione): *Alessandro Sperduti, Federico Menegazzo, Gilberto Filè, Antonio Grioli, C.E. Palazzi, Carla De Francesco, Lorenzo Finesso, Francesca Rossi, Livio Colussi, Massimo Marchiori, Francesco Tapparo, Paolo Baldan, Silvia Crafa, Luigi De Giovanni, Ombretta Gaggi.*



Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

Ringrazio l'azienda *Corvallis* e in particolare il *tutor* interno *Alberto Pietrogrande* per l'esperienza di stage vissuta.



Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

Ringrazio i compagni di corso (ormai amici), coi quali ho svolto alcuni dei progetti didattici, per avermi accettato e sopportato pazientemente (soprattutto quando ero in errore). In particolare ringrazio il gruppo *SevenTech*, formatosi per il progetto di *Ingegneria del Software*, composto da Alessio, Enrico Bo., Enrico Br., Giulio, Nicola, Pietro. Ringrazio il gruppo *IMangiaLibri*, formatosi per il progetto di *Tecnologie Web*, composto da Alberto e Dario. Ringrazio il gruppo *GeoNav*, formatosi per il progetto di *Gestione di Imprese Informatiche*, composto da Fabio e Fabrizio.



Ringraziamenti

È stato un privilegio poter trascorrere questi anni nelle aule e nei laboratori del *Dipartimento di Matematica* dell'*Università di Padova* a studiare materie che mi appassionano. Ora che sono giunto quasi alla fine di questo percorso di laurea triennale in *Informatica*, voglio ringraziare le persone il cui contributo non deve essere dimenticato.

Come ultimo ma non meno importante, ringrazio la mia famiglia per aver reso possibile il proseguimento dei miei studi.