

Un modello con tecniche *fuzzy* neuro-adattative per la classificazione dei laureati dell'Università di Bari secondo l'occupazione a un anno dalla laurea

Giuseppe Delvecchio, Francesco d'Ovidio¹

Università degli Studi di Bari

Riassunto: Nel presente lavoro si propone una nuova metodologia per migliorare la precisione dell'attribuzione dell'esito lavorativo ottenibile, con tecniche statistiche, a partire da un insieme di caratteristiche inerenti il percorso formativo universitario o relative alla ricerca di occupazione. Lo scopo di partenza è l'identificazione di alcune caratteristiche dei laureati (nello specifico, laureati dell'Università di Bari nel periodo 1995-2000) in grado di favorire l'inserimento lavorativo entro un anno dalla laurea. In quest'ottica viene sottoposto ad analisi di segmentazione un campione di laureati provenienti da detta popolazione, ottenendo risultati affetti da un consistente errore di classificazione. Si procede, quindi, ad implementare le regole risultanti dall'analisi di segmentazione in un sistema di reti neurali, utilizzando anche tecniche *fuzzy*, allo scopo di individuare possibili miglioramenti dell'impianto di regole e di definire l'importanza di una determinata configurazione di covariate e/o di dati. I risultati ottenuti verificano in modo più puntuale l'influenza, ai fini del *placement*, di caratteristiche personali e competenze acquisite.

Parole chiave: Inserimento lavorativo, Laureati, Università di Bari, Analisi di segmentazione, Tecniche *fuzzy* neuro-adattative, ANFIS.

1. Introduzione

Scopo del lavoro che qui si presenta è individuare una metodologia per migliorare la precisione della "previsione" dell'esito della ricerca lavorativa ottenibile, con tecniche statistiche, a partire da un insieme di caratteristiche inerenti il percorso formativo

¹ Il presente lavoro è stato realizzato nell'ambito del progetto "Transizioni Università-Lavoro e valorizzazione delle competenze professionali dei laureati: modelli e metodi di analisi multidimensionale delle determinanti", cofinanziato dal MIUR; coordinatore nazionale è Luigi Fabbris, coordinatore del gruppo di Bari è Francesco Delvecchio. Della nota in oggetto, opera congiunta dei due autori, va attribuita a F. d'Ovidio la redazione finale dei paragrafi 1 e 2, a G. Delvecchio quella dei paragrafi 3, 4 e 5.

universitario o relative alla ricerca di occupazione. La motivazione iniziale di detta scelta può essere fatta risalire alla constatazione che, sottoponendo ad analisi di segmentazione i dati rivenienti da un'indagine appositamente condotta su un campione di laureati dell'Università di Bari, con strategie simili a quelle utilizzate in un precedente lavoro (Crocetta e d'Ovidio, 2003), i risultati, per quanto interessanti, denunciavano un errore di classificazione elevato.

In una siffatta analisi, invero, l'errore di classificazione è in parte giustificabile, considerando che le variabili esplicative sono legate soprattutto ai percorsi formativi ed alla soddisfazione, mentre la variabile risposta (dicotomica) su cui si è investigato è l'aver trovato lavoro entro un anno di tempo dalla laurea (al netto del servizio militare eventualmente adempiuto dopo il conseguimento del titolo). Com'è ovvio, infatti, sulla variabile risposta hanno grande rilevanza anche fattori legati al mercato del lavoro non rilevati nell'indagine campionaria, molti dei quali comunque non rilevabili per propria natura.

2. Una prima analisi statistica delle determinanti dell'occupazione dei laureati presso l'Ateneo di Bari

Fra dicembre 2003 ed aprile 2004 è stata effettuata una rilevazione telefonica su un campione di laureati, allo scopo di verificare, a distanza di almeno tre anni dal conseguimento del titolo, la loro situazione dal punto di vista lavorativo².

Il questionario utilizzato per l'indagine telefonica, appositamente realizzato per essere completato in non oltre quindici minuti, è articolato in sezioni distinte destinate a raccogliere le caratteristiche socio-anagrafiche salienti dell'intervistato, informazioni sui soggetti non occupati, informazioni sui soggetti attualmente disoccupati ed informazioni sui soggetti attualmente occupati.

A queste informazioni sono state poi fatte corrispondere altre numerose informazioni, relative al curriculum formativo degli intervistati, disponibili presso gli archivi amministrativi dell'Università.

La popolazione di riferimento è composta da coloro che hanno conseguito la laurea presso l'Università di Bari dal 1995 al 2000, nelle seguenti 11 Facoltà³: Agra-

² Si ringraziano, in proposito, i dott. B. Amenduni, V. Ferrandes, L. Milone e C. Triggiani per la pazienza e la costanza con cui hanno effettuato la rilevazione ed il Centro Servizi Informativi dell'Università di Bari, nella persona del sig. G. Melchiorre, per l'affidabilità dimostrata nella fornitura di dati il più possibile corretti, necessari sia a reperire i laureati da intervistare, sia ad integrare le informazioni rilevate telefonicamente con quelle disponibili in archivio.

³ Fra le quali non è compresa la Facoltà di Scienze biotecnologiche, di recente istituzione, né, ovviamente, le Facoltà di Ingegneria ed Architettura, afferenti ad altro Ateneo (Politecnico di Bari).

ria, Economia, Farmacia, Giurisprudenza, Lettere e Filosofia, Lingue e letterature straniere, Medicina e Chirurgia, Medicina veterinaria, Scienze della formazione, Scienze politiche, Scienze matematiche, fisiche e naturali.

Dalla suddetta popolazione è stato estratto, con procedimento casuale, un campione stratificato proporzionale per corso di laurea e per genere. Il piano di campionamento prevedeva, per ottenere una buona rappresentatività ed un ridotto intervallo di confidenza delle stime, la rilevazione di oltre 3.000 interviste. La difficile reperibilità dei soggetti (anche utilizzando elenchi di riserva), unitamente alla estrema incompletezza di molte interviste, escluse quindi dall'indagine, ha portato il campione a 2.785 unità, comunque sufficientemente rappresentative della popolazione. Per motivi legati alla più difficile reperibilità e forse alla fretta degli intervistati di concludere l'intervista, ciò ha portato ad una minore quota di campionamento per i laureati delle Facoltà di Economia e di Giurisprudenza, per le quali si ha, quindi, una certa sottostima dei tassi di occupazione post-laurea (in quanto le interviste errate o non giunte a buon esito sono, presumibilmente, relative soprattutto a chi ha meno tempo e voglia di collaborare perché lavora oppure è in fase di tirocinio).

Nella Tab. 1 è riportata la composizione finale del campione, distinta per Facoltà e genere, a valle del procedimento di controllo.

Va qui sottolineata l'informazione relativa allo sbilanciamento fra i sessi (fenomeno che, nonostante l'errore campionario su descritto, si ripropone fedelmente nella popolazione): in particolar modo in Facoltà quali Lettere e filosofia, Lingue e letterature straniere e Scienze della formazione, ove le laureate ammontano all'85% del totale ed oltre (fino al 92% dell'ultima Facoltà citata), mentre uno sbilanciamento speculare si rileva soltanto fra i laureati della Facoltà di Agraria, quasi l'80% dei

Tabella 1. Distribuzione del campione di laureati presso l'Ateneo barese dal 1995 al 2000, per Facoltà e genere. Quota di campionamento per Facoltà.

Facoltà	Quota di campion.	Genere			% per genere		
		F	M	MF	F	M	MF
Agraria	14,6	11	42	53	20,8	79,2	100,0
Economia	9,1	239	195	434	55,1	44,9	100,0
Farmacia	11,6	53	25	78	67,9	32,1	100,0
Giurisprudenza	8,9	356	244	600	59,3	40,7	100,0
Lettere e filosofia	13,2	298	55	353	84,4	15,6	100,0
Lingue e letterature straniere	13,6	207	20	227	91,2	8,8	100,0
Medicina e chirurgia	13,2	108	110	218	49,5	50,5	100,0
Medicina veterinaria	14,2	12	13	25	48,0	52,0	100,0
Scienze della formazione	12,9	198	17	215	92,1	7,9	100,0
Scienze matematiche, fisiche e naturali	13,4	221	150	371	59,6	40,4	100,0
Scienze politiche	12,8	113	98	211	53,6	46,4	100,0
Università di Bari	10,6	1.816	969	2.785	65,2	34,8	100,0

quali è di genere maschile. Un tale sbilanciamento può avere qualche rilevanza nelle analisi successive per via di alcune particolarità del mercato del lavoro, in special modo nel Mezzogiorno⁴.

A distanza di vari anni dalla laurea, la situazione lavorativa degli intervistati appare abbastanza definita, pur tenendo conto della cospicua quota di essi che si dichiara ancora in formazione⁵ (10,2%); al momento dell'intervista, infatti circa il 77% dei laureati dell'Ateneo barese aveva un'occupazione (il 51,1% a tempo indeterminato), mentre poco meno del 10% aveva perso un lavoro e non ne aveva ancora trovati altri; la piccola quota restante (7,1%) era ancora in cerca di prima occupazione. Sussistono consistenti differenze fra i sessi, soprattutto in termini di accesso a occupazioni a tempo indeterminato, e fra laureati di Facoltà diverse⁶. Tuttavia, questi dati costituiscono una fotografia, per quanto interessante, poco esplicativa della realtà, essendo riferiti a coorti di laureati abbastanza spaziate nel tempo (i laureati del 1995, infatti, sono nel mercato del lavoro da circa otto-nove anni, mentre quelli del 2000 da appena tre-quattro).

Ben più significativa è la situazione prospettata nella Tab. 2, che riporta le percentuali di laureati che hanno trovato lavoro entro un determinato tempo dalla laurea (tipicamente, 12, 24 e 36 mesi) e, complementariamente, quelle di chi non ha mai lavorato fra la laurea ed il tempo limite di tre anni. Si tenga conto che, per i soli laureati che hanno espletato il servizio militare obbligatorio dopo la laurea, allo scopo di rendere più comparabili i risultati, il calcolo è stato corretto sottraendo 12 mesi al tempo di inoccupazione dichiarato dagli intervistati.

Il 55,2% dei laureati dell'Università di Bari risulta essere occupato entro un anno dal conseguimento del titolo (comprendendo in tale quota, ovviamente, anche coloro che già lavoravano prima di laurearsi), mentre solo il 22,6% dopo tre anni è ancora in condizione non professionale: in cerca di lavoro, in formazione o, in pochi casi, inattivo. A causa della "flessibilità" lavorativa con cui da sempre i neo-laureati

⁴ Ove le donne spesso trovano o conservano un lavoro meno facilmente della controparte maschile, a causa di una maggiore discontinuità dovuta ad assenze per motivi familiari che la natura o gli usi associano alla figura femminile (gravidanze, malattie infantili, ecc.).

⁵ Si pone in evidenza che, per quanto riguarda i laureati in Medicina e Chirurgia che hanno ottenuto l'iscrizione ai corsi di Specializzazione previsti nel loro ordinamento didattico, si è presa la decisione di non inserirli nel novero delle persone "in formazione", bensì fra gli occupati a tempo determinato, poiché nel loro caso il rapporto con la struttura formativa assume a tutti gli effetti, anche contributivi, caratteristiche simili appunto a quelle dei contratti di lavoro a tempo determinato, essendo l'acquisizione di una "borsa di studio" condizione necessaria, precisamente regolamentata (cfr. decreto legislativo 8 agosto 1991, n. 257), per l'accesso alla specializzazione.

⁶ E, al loro interno, anche fra i corsi di laurea (anche molto differenti fra loro) che li compongono: si pensi, ad esempio, alla Facoltà di Scienze MM.FF.NN, ove si ritrovano lauree molto richieste dal mercato del lavoro, come Informatica, ed altre meno favorite. Purtroppo, a causa della numerosità campionaria abbastanza ridotta a cui l'indagine è stata costretta ad adeguarsi, il dettaglio per corso di laurea è talora riferito a numerosità esigue e, quindi, non significative dal punto di vista statistico.

Tabella 2. Distribuzione percentuale dei laureati secondo il tempo intercorso fra laurea e prima occupazione (al netto dell'eventuale servizio militare adempiuto dopo la laurea), per Facoltà e genere dell'intervistato.

Facoltà	Tempo netto fra laurea e primo lavoro				Totale
	fino a 12 mesi	13-24 mesi	24-36 mesi	Non occupati entro 36 mesi	
Agraria	90,6	3,8	-	5,7	100,0
Economia	74,4	13,4	2,3	9,9	100,0
Farmacia	78,2	6,4	2,6	12,8	100,0
Giurisprudenza	27,3	17,5	22,3	32,8	100,0
Lettere e filosofia	44,2	24,1	11,3	20,4	100,0
Lingue e letterature straniere	68,7	15,9	3,5	11,9	100,0
Medicina e chirurgia	18,8	2,3	1,8	77,1	100,0
Medicina veterinaria	60,0	12,0	-	28,0	100,0
Scienze della formazione	69,8	14,4	3,7	12,1	100,0
Scienze matem., fisiche e nat.	76,3	9,4	3,5	10,8	100,0
Scienze politiche	66,4	11,8	4,7	17,1	100,0
<i>Genere</i>					
Femmine	52,6	15,1	9,1	23,1	100,0
Maschi	60,1	11,9	6,5	21,6	100,0
Università di Bari	55,2	14,0	8,2	22,6	100,0

devono confrontarsi (anche se un tempo si chiamava più onestamente “precaricato”) una buona parte di costoro è poi entrata nel gruppo dei disoccupati.

Pur senza scendere nel dettaglio dell'analisi, si vuole qui far presente che oltre il 90% dei laureati in Agraria ha trovato lavoro entro un anno dalla laurea, pur se, presumibilmente, per buona parte di essi si trattava di lavoro a tempo determinato, in quanto al momento della rilevazione si è registrato un livello di disoccupazione superiore al 20%. L'occupazione entro l'anno è un traguardo che è stato raggiunto anche dal 78,2% dei laureati in Farmacia, dal 76,3% di quelli in Scienze MM.FF.NN. e dal 74,4% dei laureati in Economia, mentre le quote più esigue di laureati occupati entro il medesimo termine competono a Medicina Veterinaria (18,8%), per esigenze di formazione, ed a Giurisprudenza (27,3%), per i cui laureati, invece, assume molta importanza la necessità del tirocinio presso uno studio allo scopo di sostenere l'Esame di Stato, anche perché le competenze da essi acquisite prevedono la libera professione come sbocco preferenziale.

Fermando l'attenzione proprio sul risultato (peraltro abbastanza interessante) che oltre il 55% degli intervistati aveva un'occupazione di qualche tipo entro un anno dalla laurea, si è deciso di identificare, fra i potenziali elementi della formazione universitaria noti dall'indagine o dai dati amministrativi, i fattori che hanno potuto influenzare detto risultato, e la forza esplicativa di tali relazioni. Definendo, quindi, una variabile risposta dicotomica basata sull'evento “Lavoro entro un anno dalla lau-

Tabella 3. *Presumibili determinanti dell'occupazione dei laureati e relative scale di misura.*

Variabili	Scala di misura	Variabili	Scala di misura
Punteggio per la qualità degli insegnamenti specialistici	Discreta (0-100)	Livello delle conoscenze informatiche	Ordinale (1-4)
Punteggio per la qualità delle attività professionalizzanti	Discreta (0-100)	Livello di conoscenza dell'inglese	Ordinale (1-4)
Punteggio per la qualità delle attività pratiche	Discreta (0-100)	Costanza nella frequenza delle lezioni	Ordinale (1-4)
Punteggio per realizzazione del prestigio sociale	Discreta (0-100)	Corso di laurea di tipo applicativo	Categoriale dicotomica
Punteggio per realizzazione della sicurezza e stabilità lavoro	Discreta (0-100)	Avere conseguito abilitazione all'insegnamento	Categoriale dicotomica
Punteggio per realizzazione della vicinanza alla famiglia	Discreta (0-100)	Avere conseguito abilitazione all'esercizio della professione	Categoriale dicotomica
Punteggio per realizzazione della disponibilità di tempo libero	Discreta (0-100)	Avere svolto tirocinio post-laurea	Categoriale dicotomica
Numero di colloqui di lavoro	Discreta (0-99)	Avere svolto formazione post-laurea	Categoriale dicotomica
Voto di laurea	Discreta (80-110L)	Avere lavorato durante gli studi universitari	Categoriale dicotomica
Età alla laurea (in anni compiuti)	Discreta (23-60)	Genere (M / F)	Categoriale dicotomica
Numero di lingue parlate almeno discretamente	Discreta (0 - 5)	Laurea era un requisito per il lavoro	Categoriale dicotomica

rea”, sono state applicate, seguendo una procedura ormai consueta⁷, tecniche di analisi loglineare e logit per individuare le presumibili determinanti dell'occupazione a medio-breve periodo. Le variabili la cui influenza sulla risposta, in base all'analisi loglineare, è statisticamente significativa sono brevemente descritte nella Tab. 3; si sottolinea la presenza di elementi soggettivi di valutazione, espressi dagli intervistati attribuendo punteggi da 0 a 100 ad alcuni aspetti della qualità della formazione universitaria ai fini dell'attività lavorativa e ad altri aspetti specifici del lavoro svolto.

Sulla base di dette presumibili variabili esplicative, è stato elaborato un modello logit per la previsione del lavoro a breve-medio periodo, con procedura stepwise basata sul rapporto di massima verosimiglianza (con livello di significatività pari al 5% per l'inserimento di ogni esplicativa e del 10% per la sua rimozione). Il subcampione su cui è stata effettuata l'analisi è quello dei 2.414 intervistati che, dopo la

⁷ Detta procedura parte da un modello log-lineare saturato per tabelle di contingenza multidimensionali per poi eliminare, uno per volta, gli effetti non significativi mediante il test $G^2 = -2\ln\Lambda$; prendendo in considerazione solo le interazioni fra la variabile risposta qui considerata e le altre, si è poi costruito un appropriato modello di regressione logit.

Tabella 4. Effetti significativi del modello logit a risposta dicotomica "Occupazione entro un anno dalla laurea" relativo ai laureati dell'Università di Bari che lavorano o hanno lavorato dopo la laurea (1995-2000)

Effetti significativi	Stime dei parametri	Errori standard	p-value	Odds ratio
Intercetta	-2,034	0,566	<0,01	0,13
Frequenza saltuaria lezioni	-1,144	0,406	<0,01	0,32
Corso di laurea di tipo applicativo	-0,319	0,098	<0,01	0,73
Avere svolto formazione post-laurea	-0,283	0,160	0,08	0,75
Genere M \cap Età alla laurea	-0,082	0,034	0,01	0,92
Frequenza saltuaria \cap Numero colloqui di lavoro	-0,046	0,021	0,03	0,96
Laurea requisito lavoro \cap Punt. per disp. tempo libero	-0,012	0,005	0,01	0,99
Punt. realizzazione per disponibilità di tempo libero	-0,008	0,002	<0,01	0,99
Punteggio qualità degli insegnamenti specialistici	-0,007	0,003	0,02	0,99
Punteggio realizzazione per sicurezza/stabilità lavoro	0,009	0,002	<0,01	1,01
Punteggio realizzazione per prestigio sociale	0,011	0,004	<0,01	1,01
Frequenza saltuaria \cap Punteggio insegn. specialistici	0,012	0,006	0,06	1,01
Punteggio qualità delle attività professionalizzanti	0,013	0,002	<0,01	1,01
Genere M \cap Punteggio per prestigio sociale	0,014	0,007	0,03	1,01
Numero di colloqui di lavoro	0,017	0,011	0,10	1,02
Età alla laurea	0,067	0,017	<0,01	1,07
Numero di lingue parlate	0,228	0,068	<0,01	1,26
Frequenza saltuaria \cap Laurea requisito per il lavoro	0,642	0,252	0,01	1,90
Genere (M)	1,635	1,007	0,10	5,13
Laurea requisito per il lavoro	1,769	0,303	<0,01	5,87

laurea, hanno trovato occupazione, sia che l'abbiano poi cessata, sia che ancora lavorino, escludendo quindi coloro che hanno proseguito il proprio percorso formativo. Del modello di regressione logit identificato, si riportano qui solo i coefficienti significativi in ordine di influenza, da negativa a positiva (Tab. 4), senza entrare nello specifico delle relazioni trovate in quanto, nell'economia del presente lavoro, si ha interesse soprattutto a verificare quali variabili sono interessate.

Tuttavia, in detto modello va sottolineato l'incremento della probabilità nel trovar lavoro entro un anno legato all'orientarsi verso professioni che richiedono la laurea (pari a quasi sei volte quella relativa al caso opposto) o al fatto di essere maschio, e di non dover, quindi, combattere con la nota ritrosia dei datori di lavoro ad assumere donne, per la loro caratteristica (spesso negativa dal punto di vista del profitto d'impresa) di avere o di poter avere in futuro figli che sottraggono tempo ed attenzione agli impegni professionali. Sfavorevole alla probabilità di inserirsi in breve nel mondo del lavoro è, invece, l'aver conseguito una laurea di tipo applicativo (categoria intermedia fra lauree "scientifiche" ed "umanistiche", da noi qui definita per tener conto delle particolarità di Facoltà come Economia, Scienze Politiche o Giurisprudenza, comunemente considerate umanistiche), presumibilmente per la necessità,

per alcuni laureati, di effettuare un praticantato per poter poi conseguire l'abilitazione professionale. Ancor più negativo si rivela l'aver frequentato saltuariamente i corsi universitari. Per quanto riguarda l'influenza delle variabili di valutazione (che, essendo espresse su scala 0-100, presentano coefficienti prossimi a zero ma comunque significativi), va sottolineata l'influenza negativa della realizzazione lavorativa per disponibilità di tempo libero e della valutazione assegnata agli insegnamenti universitari specialistici (salvo per coloro che, presumibilmente perché studenti-lavoratori, frequentavano saltuariamente) e l'influenza positiva delle attività professionalizzanti svolte nel corso degli studi.

Al fine di identificare gli elementi del processo di formazione che forniscono i migliori *outcomes*, sono stati poi sperimentati vari metodi di segmentazione (C@rt, CHAID esaustivo, LAID-OUT⁸), riferendosi a diversi insiemi di variabili esplicative, nella cui costruzione si è tenuto conto anche dei risultati dell'analisi logit,

È opportuno qui ricordare brevemente il fondamento metodologico dell'analisi di segmentazione: essa, infatti, parte dal complesso dei dati del campione, che viene suddiviso in gruppi (*nodi*) via via più omogenei al proprio interno in termini di relazioni fra la variabile risposta nota (dipendente) e le variabili assunte come esplicative. Tale procedimento produce una "regola di classificazione". La migliore segmentazione, fra tutte quelle possibili, è quella che meglio risponde al criterio di omogeneità interna dei gruppi generati (*purity*): nella condizione ottimale, tutti i casi di ogni singolo *nodo finale* dovrebbero presentare una stessa modalità della variabile risposta. Ai fini esplorativi, però, è talvolta necessario sacrificare una struttura efficiente, ma di difficile interpretazione, per una più chiara pur se dotata di minore purezza.

Il procedimento di espansione dell'albero di classificazione si arresta, comunque, quando si verifica una delle regole di arresto predefinite, ossia quando:

1. tutti i casi di un nodo fanno rilevare per i predittori valori statisticamente identici;
2. tutti i casi del nodo hanno il medesimo valore della variabile risposta (*nodo puro*);
3. viene rilevata una dimensione minima, in genere fissata dal ricercatore, per il nodo "genitore" (da cui si diparte l'ulteriore classificazione) o il nodo "figlio", che può essere, a sua volta, genitore;
4. la profondità dell'albero ha raggiunto un valore massimo, anch'esso definito in base alle necessità della ricerca.

Per non espandere troppo l'albero di classificazione e mantenerne una certa interpretabilità, si è posto pari a 10 il massimo numero di livelli di segmentazione, fissando a 30 il numero minimo di casi per i nodi "genitore" e a 10 quello per i nodi "figlio". Si è fatto anche uso di procedure di sfoltoimento (*pruning*), ossia di eliminazione, a posteriori, di nodi superflui o ridondanti dal punto di vista della classificazione, pur se di qualche interesse dal punto di vista descrittivo.

⁸ Per i quali si rimanda a: Breiman *et al.*, 1991; Fabbris 1997; Fabbris e Martini, 2002; Kass 1980; Schievano 2002, 2003; Sonquist 1970.

Purtroppo, l'errore di classificazione risultante dalle procedure utilizzate, a seconda dell'algoritmo utilizzato, si è rivelato pari o superiore al 34% (in altri termini, l'attribuzione della condizione di "occupato entro un anno" o di quella opposta risulta corretta per non oltre il 66% degli intervistati).

Sembra logico, quindi, giungere alla conclusione (peraltro non sorprendente) che le variabili disponibili relative al processo formativo ed all'offerta di lavoro, sono insufficienti a descrivere il fenomeno del *placement* dei laureati in assenza di informazioni sul lato della domanda. Ciò, d'altra parte, non fa che confermare alcune conclusioni a cui si era pervenuti nel citato lavoro di Crocetta e d'Ovidio (2003), riguardante, nella fattispecie, i laureati dell'Università di Foggia.

Tuttavia, è sorto il dubbio che parte dell'errore di classificazione potesse dipendere dalla definizione stessa delle variabili di partenza, oppure dalla struttura dei vari sottogruppi da esse definiti: invero, cosa succederebbe se gli insiemi individuati dalla segmentazione non avessero i contorni netti, ma velati, sfuocati (*fuzzy*)?

Si è deciso, perciò, di tentare di migliorare la precisione della classificazione facendo uso di tecniche fuzzy neuro-adattative, utilizzando i risultati dell'analisi di segmentazione come *regole* iniziali a cui applicare la procedura di ottimizzazione.

3. Cenni sull'approccio *fuzzy* proposto

Come precedentemente esposto, l'idea di fondo è di sfruttare l'albero di segmentazione per ricavare le regole di un sistema fuzzy. In particolare, usando un insieme di dati input/output, il metodo "regola" i parametri delle funzioni membership tramite una rete neurale, in maniera tale da migliorare il sistema fuzzy stesso.

A causa delle caratteristiche matematiche dei metodi fuzzy, per poter implementare una chiara attribuzione delle membership si è stabilito di delimitare nel modo seguente la tipologia di analisi di segmentazione da usare per la definizione delle regole:

- 1) variabile risposta dicotomica (non trasformata in logit);
- 2) alberi binari o al più ternari;
- 3) per evitare una complicazione delle regole eccessiva ed inutile (ai fini dell'ottimizzazione fuzzy), nel modello di classificazione sono inserite solo covariate al più ordinali con non oltre quattro modalità oppure continue (o anche ordinali assimilabili a continue).

Tali limitazioni hanno portato a creare un modello di segmentazione con metodo C@rt avente un errore di classificazione abbastanza rilevante (oltre il 35%, con 31 nodi finali), ma più suscettibile di miglioramenti rispetto ad altri.

Nei paragrafi che seguono si spiegherà brevemente il metodo proposto.

3.1 Cenni sull'ANFIS: Adaptive Neuro Fuzzy Inference System

La tecnica ANFIS (*adaptive neuro-fuzzy inference system*) permette ad un sistema fuzzy di apprendere le informazioni contenute in un insieme di dati.

Questo metodo è molto simile a quello utilizzato all'interno delle reti neurali, ed è stato descritto per la prima volta da Jang (1993): alle funzioni *membership* di un sistema fuzzy sono associati dei parametri, i quali sono "regolati" tramite un processo iterativo di apprendimento basato su un insieme di dati di tipo "input/output", in maniera tale da adattare la risposta del sistema fuzzy ai dati stessi (il metodo, in sostanza, minimizza la somma dei quadrati delle differenze fra gli output dell'insieme di dati, e gli output ottenuti applicando gli input dell'insieme di dati al sistema fuzzy). Accenniamo brevemente a questa tecnica.

Per semplicità, assumeremo che il sistema inferenziale fuzzy abbia due input, x ed y , ed un solo output, f . Supporremo, inoltre, che il sistema abbia due regole fuzzy del tipo di Takagi e Sugeno⁹ del primo ordine (Takagi and Sugeno, 1983):

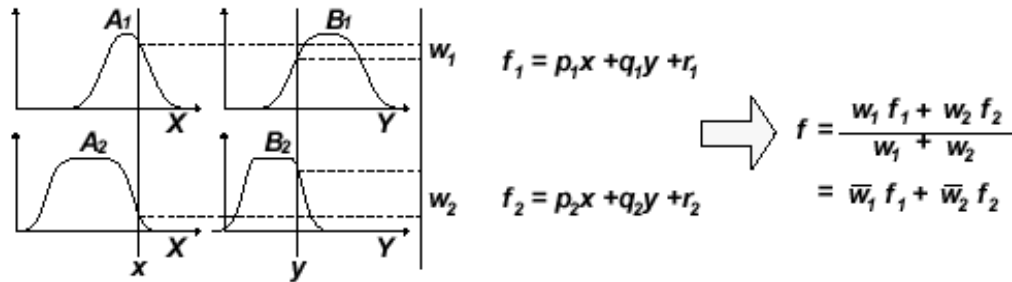
regola 1: se x è A_1 e y è B_1 , allora $f_1 = p_1 x + q_1 y + r_1$,

regola 2: se x è A_2 e y è B_2 , allora $f_2 = p_2 x + q_2 y + r_2$.

ove A_i e B_i sono i sottoinsiemi fuzzy associati a termini linguistici (ad es., piccolo, medio, grande, ecc.) attivati rispettivamente dagli input non fuzzy (numeri reali) x ed y ; p_i , q_i e r_i sono invece opportuni parametri.

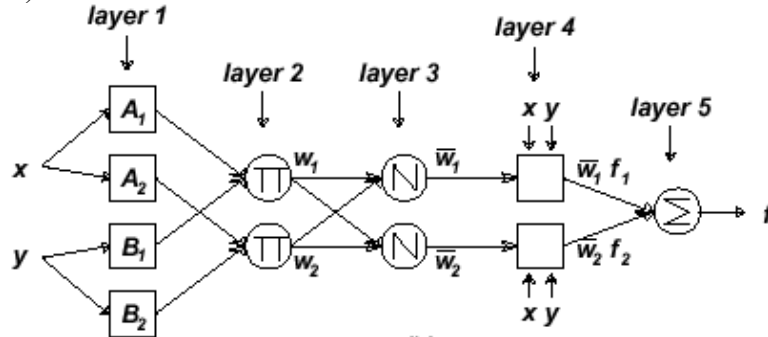
Il sistema fuzzy è mostrato in Fig. 1: i pesi w_i sono generalmente ottenuti adoperando due tecniche di attivazione dei sottoinsiemi fuzzy, ovvero la *correlation product encoding* e la *correlation minimum encoding* (Delvecchio, 2002).

Figura 1. Rappresentazione di un generico sistema inferenziale fuzzy con due input, x ed y , ed un solo output, z (modello di Takagi e Sugeno).



⁹ Le principali tecniche di inferenza *fuzzy* sono il "metodo di Mamdani" e il "metodo di Sugeno". Il primo (Mamdani and Assilian, 1975) rappresenta la metodologia più applicata: ha i vantaggi di essere intuitivo, diffusamente accettato, e ben adattabile agli input umani. Il secondo (Sugeno, 1985) può essere usato per modellare qualsiasi sistema di inferenza in cui le funzioni *membership* in output sono lineari o costanti: ha i vantaggi di essere computazionalmente efficiente, di lavorare bene con tecniche lineari e con tecniche di ottimizzazione e adattative, e di adattarsi bene all'analisi matematica (AA.VV., 1999, pp. 2-36, 2-37 e 2-91).

Figura 2. Architettura ANFIS del sistema inferenziale fuzzy di Figura 1: i nodi quadrati (o nodi adattativi) hanno parametri da “regolare”, mentre i nodi circolari (o nodi fissi) non ne hanno.



L'architettura ANFIS corrispondente al sistema fuzzy di Fig. 1 è mostrata in Fig. 2. In quest'ultima, sono evidenziati i cinque layer contenenti i nodi funzione (di forma quadrata o circolare), i quali applicano una particolare funzione ai loro ingressi o ai parametri in essi contenuti. In particolare, i nodi quadrati (o nodi adattativi) in Figura 2 hanno parametri da “regolare”, mentre i nodi circolari (o *nodi fissi*) non ne hanno.

Nel seguito descriveremo i cinque layer della rete neurale riportata in Fig. 2.

3.1.1 Layer 1

Ogni nodo in questo layer è un nodo quadrato con una funzione nodo del tipo:

$$O_i^{(1)} = \mu_{A_i}(x)$$

dove x è l'input del nodo i -esimo, e A_i è il termine linguistico (ad es., piccolo, medio, grande, ecc.) associato alla funzione di questo nodo. In altre parole, $O_i^{(1)}$ è la funzione membership di A_i (funzione indicata con $\mu_{A_i}(x)$), e perciò specifica il grado con cui un dato valore x appartiene ad A_i . Si noti che $\mu_{A_i}(x)$ è una funzione continua, differenziabile a pezzi, come ad esempio una funzione trapezoidale oppure triangolare, identificata da parametri da “regolare”.

3.1.2 Layer 2

Ogni nodo in questo layer è un nodo circolare, etichettato con Π in Fig. 2.

Applicando la regola del *correlation product encoding* (Delvecchio, 2002), tale nodo moltiplica fra loro i suoi ingressi e restituisce in uscita il loro prodotto. Nel nostro esempio,

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(x), \quad i = 1, 2.$$

Applicando, invece, la regola del *correlation minimum encoding*, tale nodo restituisce in uscita il minimo dei suoi ingressi, ovvero

$$w_i = \min(\mu_{A_i}(x), \mu_{B_i}(x)), \quad i = 1, 2.$$

3.1.3 Layer 3

Ogni nodo in questo layer è un nodo circolare, etichettato con N in Fig. 2. L'*i*-esimo nodo calcola:

$$\bar{w}_i = \frac{w_i}{\sum_i w_i} \quad i = 1, 2 \dots$$

Per convenzione, l'output di questo layer è chiamato *normalized firing strength*.

3.1.4 Layer 4

Ogni nodo in questo layer è un nodo quadrato con una funzione nodo del tipo:

$$O_i^{(4)} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad i = 1, 2,$$

dove \bar{w}_i è l'output del layer 3, e $\{p_i, q_i, r_i\}$ sono i parametri da "regolare" del nodo *i*-esimo, chiamati *consequent parameters*.

3.1.5 Layer 5

L'unico nodo in questo layer è un nodo circolare, etichettato con Σ in Fig. 2, che effettua la somma degli output del layer 4:

$$O_i^{(5)} = \sum_i O_i^{(4)} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad i = 1, 2.$$

3.1.6 L'addestramento della rete

Assumendo che un insieme di *P* dati sia usato per addestrare la rete, l'*errore di misurazione* (o *funzione energia*) per il generico dato *p*-esimo ($1 \leq p \leq P$) è dato da:

$$E_p = (T_p - O_{1,p}^{(5)})^2$$

dove T_p è il valore output del *p*-esimo elemento dell'insieme di dati di addestramento, e $O_{1,p}^{(5)}$ è invece il corrispondente valore in output fornito dalla rete (in output al layer 5). Perciò, l'errore di misura totale è:

$$E = \sum_{p=1}^P E_p.$$

Brevemente, ad ogni iterazione l'algoritmo¹⁰ calcola, per l'insieme di dati di addestramento, l'energia *E* e il valore dei parametri (da adoperare per l'iterazione successiva) che riducono tale energia.

A tal proposito, adattando ciò che Jang ha dimostrato (1993), se indichiamo con α_i il parametro generico di una determinata rete neuro adattativa, risulta:

¹⁰ Esistono in realtà due tipi di algoritmo per l'ottimizzazione dei parametri del sistema fuzzy: *backpropagation* (basato sul metodo del gradiente, il quale sfrutta il gradiente per avvicinarsi alla soluzione ad ogni iterazione), e *metodo ibrido* (che combina il metodo del gradiente con il metodo dei minimi quadrati).

$$\Delta\alpha_i = -\eta \frac{\partial E}{\partial \alpha_i}$$

ove η denota il tasso di apprendimento (*learning rate*):

$$\eta = \frac{k}{\sum_i \left(\frac{\partial E}{\partial \alpha_i} \right)^2};$$

k è un opportuno parametro che influisce sulla velocità di convergenza dell'algoritmo

$$\frac{\partial E}{\partial \alpha_i} = \sum_{p=1}^P \frac{\partial E_p}{\partial \alpha_i}, \quad \frac{\partial E_p}{\partial \alpha_i} = \sum_{O^{(*)} \in S} \frac{\partial E_p}{\partial O^{(*)}} \frac{\partial O^{(*)}}{\partial \alpha_i}$$

ove S è l'insieme di nodi $O^{(*)}$ il cui output dipende da α_i .

Indicato con $\#(k)$ il numero di nodi del layer k -esimo e con $O_{i,p}^{(k)}$ il nodo funzione del layer k -esimo alla posizione i -esima in corrispondenza del dato p -esimo, si ha

$$\frac{\partial E_p}{\partial O_{i,p}^{(k)}} = \sum_{m=1}^{\#(k+1)} \frac{\partial E_p}{\partial O_{m,p}^{(k+1)}} \frac{\partial O_{m,p}^{(k+1)}}{\partial O_{i,p}^{(k)}};$$

si noti che il layer 5 ha un solo nodo, e pertanto il corrispondente nodo funzione viene indicato con $O_p^{(5)}$:

$$\frac{\partial E_p}{\partial O_{i,p}^{(4)}} = \frac{\partial E_p}{\partial O_p^{(5)}} \frac{\partial O_p^{(5)}}{\partial O_{i,p}^{(4)}}$$

L'algoritmo si arresta quando E raggiunge un minimo prefissato.

Tutto ciò premesso, è noto (Delvecchio, 2002) che per identificare il sistema *fuzzy* adoperato occorre ancora conoscere:

- il tipo di sistema e la modalità di attivazione dei sottoinsiemi *fuzzy* (cfr. par. 3.2);
- le *membership* associate alle grandezze in ingresso al sistema (cfr. par. 3.3).

Nel seguito del paragrafo si presentano e si commentano gli aspetti del sistema *fuzzy* che è stato adottato.

3.2 Il sistema fuzzy adoperato

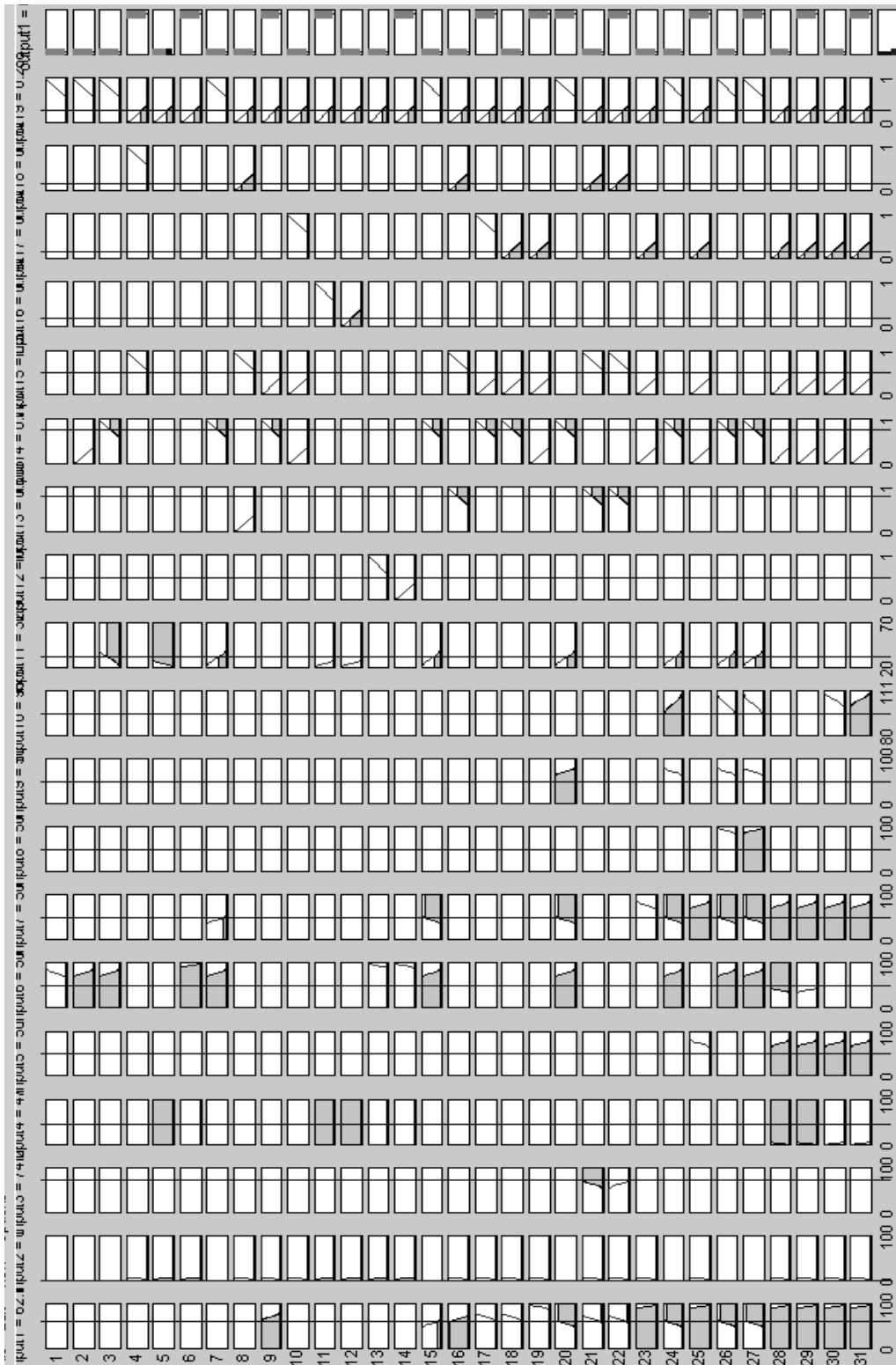
Il sistema *fuzzy* adoperato in questo lavoro si basa sulla tecnica di inferenza *fuzzy* conosciuta in letteratura come "metodo di Sugeno di ordine zero", ovvero "metodo di Takagi-Sugeno di ordine zero" (Takagi e Sugeno, 1983, Sugeno, 1985).

In particolare, una generica regola ha la forma (cfr. par. 3.1):

$$\text{se } x \text{ è } A_i \text{ e } y \text{ è } B_i, \text{ allora } f_i = r_i.$$

Nel nostro caso, inoltre, si è applicata la regola *Correlation minimum encoding* (cfr. par. 3.1.2).

Figura 3. Esempio di attivazione delle 31 regole (una ogni riga della figura) nello schema fuzzy proposto, così come visualizzate nel software implementato.



In Fig. 3 si riporta un esempio di attivazione delle 31 regole (una per ogni riga della figura), ricavate dall'albero della segmentazione, nello schema fuzzy proposto, così come visualizzate nel software implementato in Matlab. In ascissa vengono riportati gli *input* (genere, età alla laurea, voto di laurea, ecc.), mentre in basso a destra vi è il sottoinsieme in *output*, ottenuto “consolidando” i sottoinsiemi dell'ultima colonna, in output alle 31 regole (cfr., ad es., Crocetta e Delvecchio, 2003).

In particolare si noti che si sono ripetute come input le variabili che compaiono più volte in una stessa regola, altrimenti il Matlab non avrebbe permesso di inserirle.

3.3 Membership associate alle grandezze in ingresso al sistema

Distingueremo il caso di variabile ordinale da quella nominale.

3.3.1 Membership di variabile ordinale

Nella logica classica, la funzione di appartenenza dell'insieme “età alla laurea > 29” avrebbe valore 1 per una età maggiore di 29, e 0 altrimenti: pertanto la sua rappresentazione grafica avrebbe una tipica forma a “gradino”.

Per tale motivo, nell'approccio fuzzy per le membership di variabili ordinali si sono adoperate funzioni tipo sigmoide (cfr. Fig. 4).

Figura 4. Rappresentazione della membership associata alla variabile ordinale “età alla laurea > 29”, così come visualizzata nel software implementato.

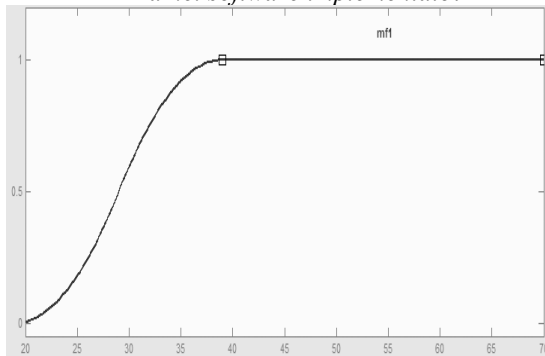
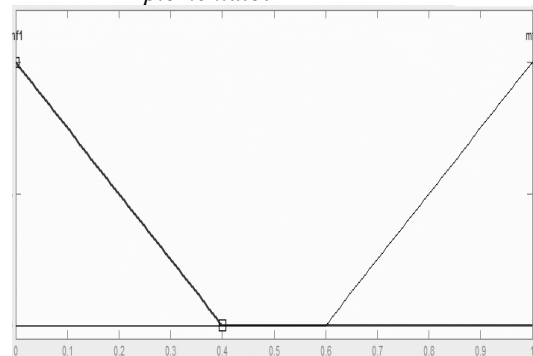


Figura 5. Rappresentazione delle membership associate alle variabili nominali, così come visualizzate nel software implementato.



3.3.2 Membership di variabile nominale

Le variabili nominali (genere, attività lavorativa durante il corso di laurea, conoscenze informatiche, ecc.) in realtà presentano valori delle membership pari a 0 oppure 1, e quindi non hanno quel grado di “sfumatura” tipico delle variabili *fuzzy*.

Ai fini dell'implementazione in Matlab, tuttavia, si sono dovute comunque adoperare funzioni membership, in particolare sono state scelte funzioni triangolari (cfr. Fig. 5) perché molto semplici ed usualmente adoperate (Kosko, 1995).

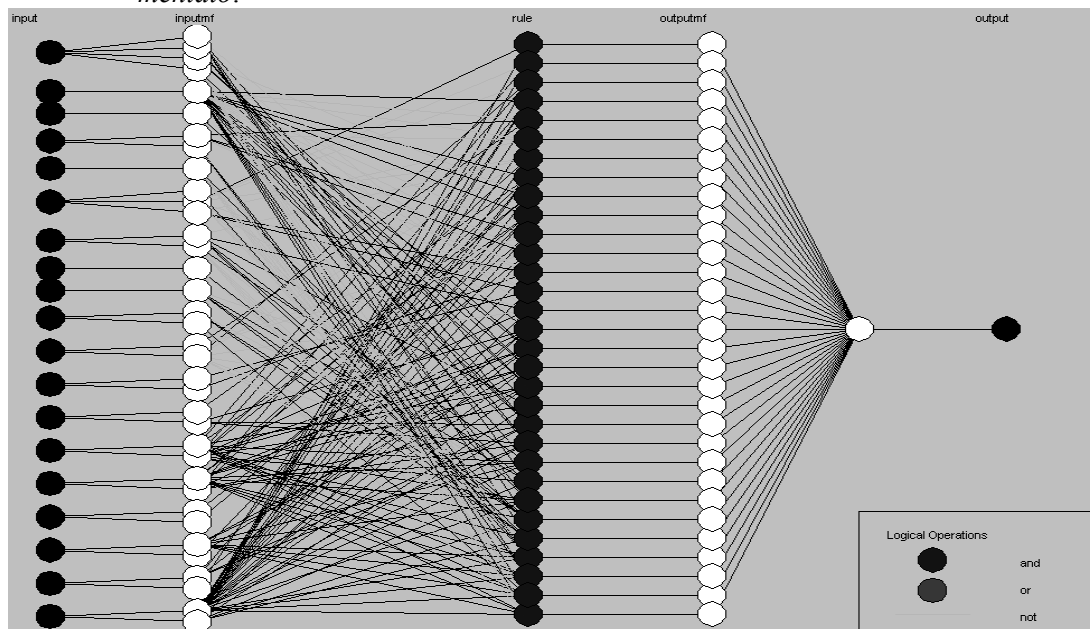
In particolare, si noti in figura che, poiché i valori in ingresso alle funzioni membership sono pari a 0 oppure 1, analogamente i valori assunti da tali funzioni saranno soltanto 0 ed 1.

3.4 *La rete neurale*

In Fig. 6 si riporta la rappresentazione della rete neurale ricavata dall'albero della segmentazione, così come visualizzate nel software implementato in *Matlab*.

Dal confronto di tale figura con la Fig. 2, in particolare si noti che: il secondo strato di neuroni di Fig. 6 rappresenta il layer 1 di Fig. 2, il terzo strato i layer 2 e 3, il quarto strato il layer 4, ed il quinto strato (neurone singolo) il layer 5.

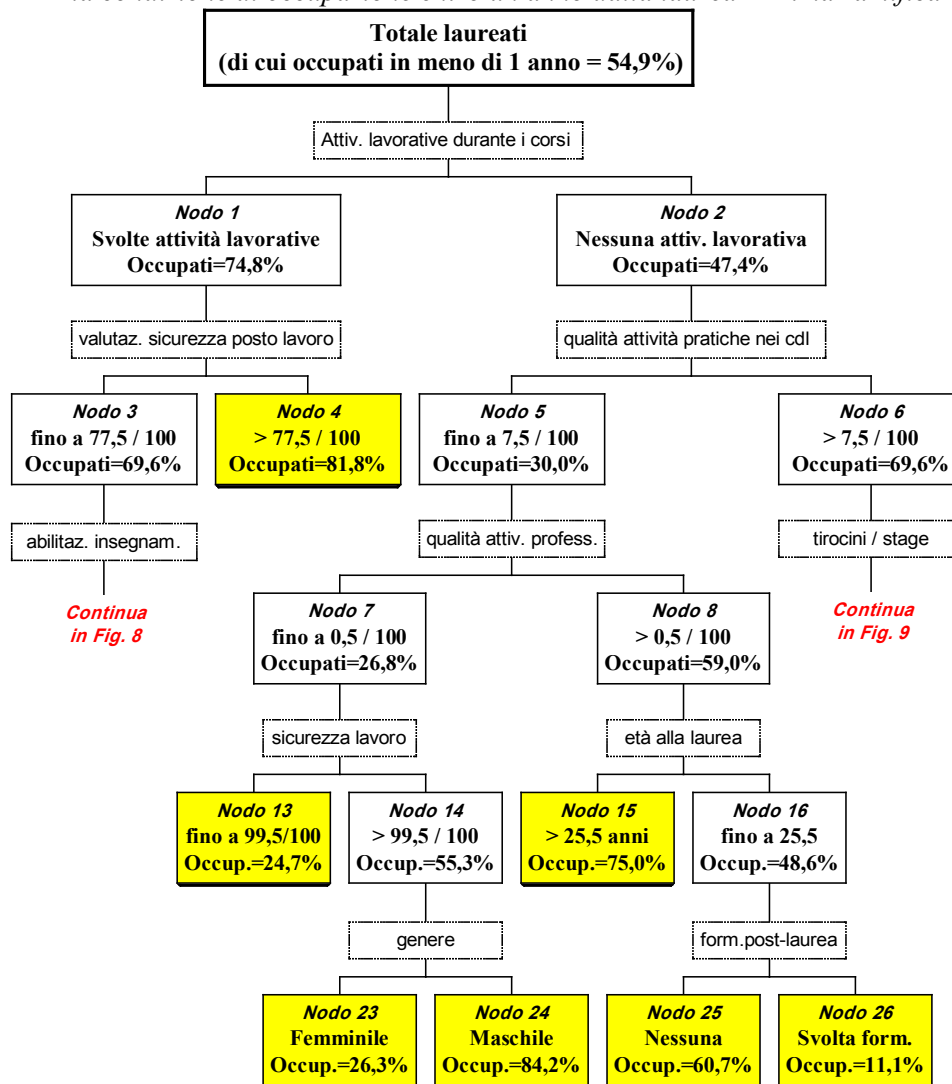
Figura 6. *Rappresentazione della rete neurale, così come visualizzate nel software implementato.*



4. Risultati ottenuti con l'applicazione della nuova metodologia proposta

Passiamo ora a descrivere l'albero di segmentazione ottenuto ed a considerare i nodi relativi alle diverse variabili che influenzano il tempo di inserimento lavorativo dei laureati dell'Università di Bari. Per motivi tipografici abbiamo diviso l'albero di segmentazione in più parti, riportate nelle successive Figure 7-9.

Figura 7. Albero di segmentazione ottimizzato dei laureati dell'Università di Bari secondo la condizione di occupazione entro un anno dalla laurea - Prima ramificazione.



La Fig. 7 evidenzia le variabili più influenti per l'analisi effettuata. Come si vede, nel complesso la percentuale dei laureati occupati entro un anno dalla laurea (54,9%) risulta solo di poco superiore rispetto a quella dei non occupati. Il nostro obiettivo è verificare come la presenza di certe caratteristiche possa modificare tale equilibrio fornendo più o meno informazioni sulle possibilità di ingresso nel mondo del lavoro a particolari categorie di laureati.

Tale segmentazione pone al primo livello, come variabile discriminante, l'aver o no lavorato durante il corso degli studi: evento che, ovviamente, accomuna sia giovani che hanno compiuto esperienze professionalizzanti, sia lavoratori più anziani

che si sono laureati per scopi di progressione di carriera (gruppo numericamente valutabile intorno al 10% del campione).

Più interessanti sono le osservazioni che si traggono dal secondo livello dell'albero, ove sono poste le valutazioni fornite dagli intervistati a due aspetti diversissimi fra loro: per chi ha lavorato durante i corsi, infatti, ad un'alta realizzazione della stabilità lavorativa (punteggio superiore a 77,5/100)¹¹ corrispondono le quote più elevate di occupati; fra chi non ha lavorato da studente, invece, la quota maggiore di occupati entro un anno si legge per coloro che valutano in modo comunque superiore allo zero (> 7,5/100) le attività pratiche apprese durante i corsi. Fra coloro che hanno dato alla qualità delle attività pratiche un voto prossimo a zero e che hanno valutato "zero" anche la qualità delle attività professionalizzanti, peraltro, la quota di occupati in tempi brevi è molto bassa (26,8%).

Altre variabili discriminanti da valutare con attenzione sono l'età alla laurea (i più giovani trovano infatti lavoro più facilmente), il genere (essendo come sempre favoriti i maschi) e la formazione post-laurea, che ovviamente influisce negativamente sull'occupazione in tempi brevi: infatti, chi deve svolgere tirocini, master o altro ha meno tempo per dedicarsi alla ricerca di occupazione.

Nel secondo e nel terzo ramo di segmentazione (Figure 8 e 9) vi sono altre relazioni interessanti, che solo per motivi di spazio non è possibile qui approfondire. Si sottolinea in questa sede soltanto la minore occupazione in tempi brevi fatta rilevare, per gli stessi motivi su esposti, da chi ha sostenuto gli esami per l'abilitazione all'insegnamento, mentre chi non aveva bisogno di tale titolo ha messo a frutto la laurea in un lavoro più o meno adeguato.

Proprio l'adeguatezza del lavoro trovato è un altro fattore discriminante evidenziato in Fig. 8, a ridosso dell'età alla laurea: fra i laureati intervistati, infatti, è ben più elevata rispetto alla media la quota di occupati che valutano in modo abbastanza negativo (con un punteggio non superiore a 45/100) la coerenza del lavoro svolto con il proprio titolo: tale quota è infatti pari all'84,6%, mentre fra i laureati che hanno trovato un lavoro coerente con il proprio percorso formativo gli occupati entro un anno ammontano a meno del 50%. La flessibilità e la capacità di far fronte a compiti per cui l'Università non ha fornito formazione è quindi un fattore vincente.

La Fig. 9, oltre al consueto ed intrinseco ritardo dovuto a corsi di abilitazione e formazione post-laurea, pone in evidenza (già al quarto livello nello schema generale, ma in seconda linea nella figura) la minore possibilità occupazionale legata al possesso di una laurea umanistica: 29,1% contro il 61,7% dei laureati d'altro orientamento.

¹¹ Come si è accennato in precedenza, agli intervistati è stato chiesto di assegnare un punteggio, da 0 a 100, alla propria soddisfazione per quanto riguarda aspetti dell'attività lavorativa, fra cui la stabilità dell'occupazione, ed aspetti riguardanti la propria formazione universitaria. Al presente, per quanto riguarda la stabilità lavorativa, ai fini della *purezza* del nodo (cfr. paragrafo 2) il punto discriminante, che corrisponde alla maggiore discretizzazione possibile dei risultati dell'ottimizzazione ottenuta con le tecniche fuzzy, è appunto il punteggio di 77,5 su 100.

Figura 9. Albero di segmentazione ottimizzato dei laureati dell'Università di Bari secondo la condizione di occupazione entro un anno dalla laurea - Terza ramificazione.

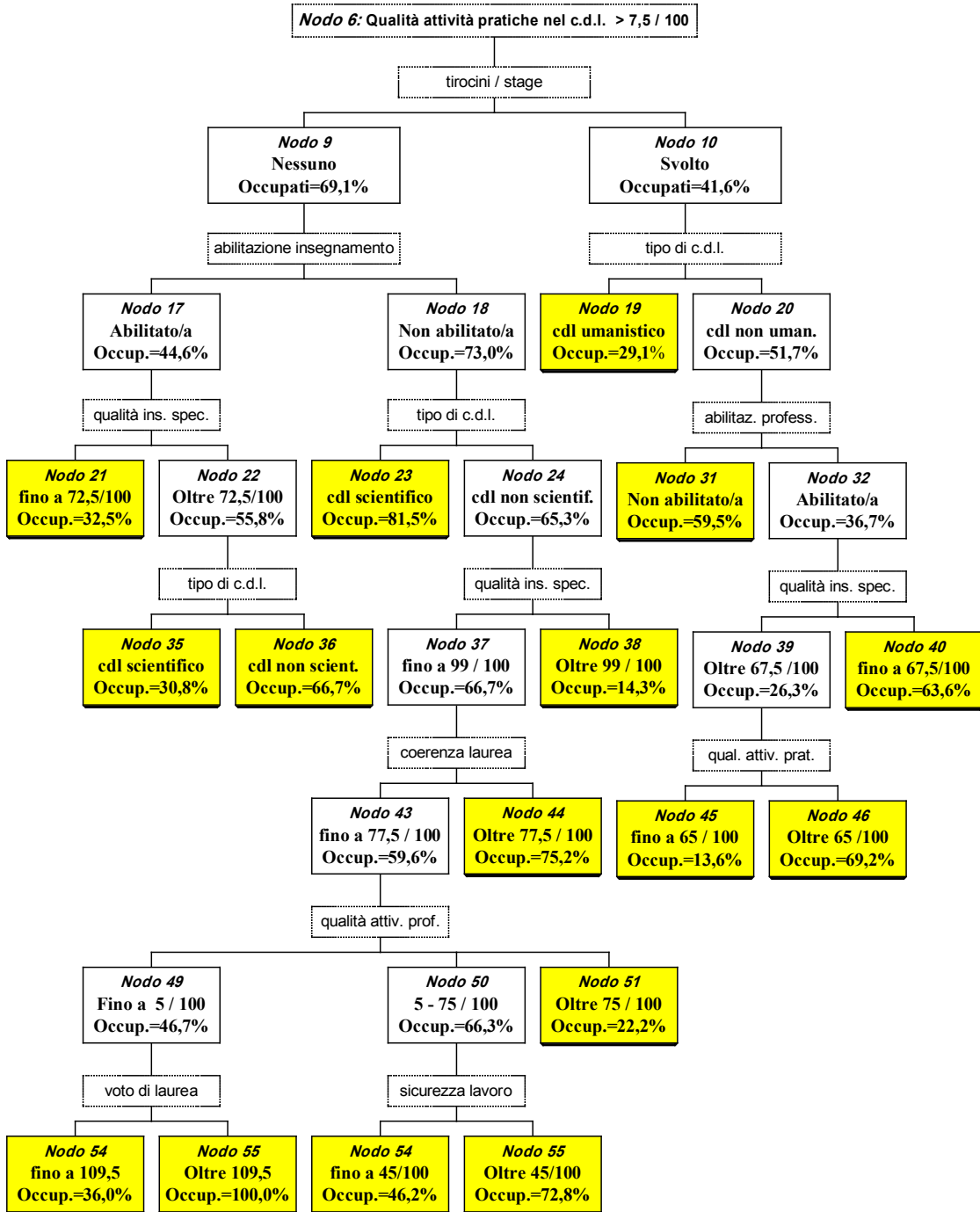


Tabella 8. *Tempi di occupazione osservati e previsti per i laureati che hanno lavorato dopo la laurea (verifica delle regole di segmentazione ottimizzate).*

<i>Tempi di occupazione previsti</i>	<i>Tempi di occupazione osservati</i>		Totale
	Fino a un anno	Oltre un anno	
Fino a 1 anno	1060	335	1395
Oltre 1 anno	272	747	1019
Totale	1332	1082	2414

Tabella 9. *Analisi dell'errore di classificazione.*

<i>Esito della previsione tramite l'analisi di segmentazione</i>	Occupati entro un anno	Non occupati entro un anno	Totale
% classificazione esatta	79,6	69,0	74,9
% classificazione errata	20,4	31,0	25,1

In definitiva, riportando al database di partenza le regole ottenute tramite la presente ottimizzazione ed attribuendo così ai laureati la qualifica “prevista” di occupati o non occupati entro 12 mesi, si può calcolare, per confronto con il corrispondente ammontare di effettivi occupati/non occupati (Tab. 8), si ottiene un errore di classificazione poco superiore al 25%, con un miglioramento di oltre 10 punti percentuali rispetto a quelli dell'albero di segmentazione utilizzato per la definizione delle regole (Tab. 9). Il risultato appare quindi interessante, soprattutto tenendo conto dei limiti tecnici che si è dovuto affrontare, ma soprattutto è suscettibile di sviluppi sia metodologici che interpretativi.

La forma delle membership (le quali esprimono l'importanza, per una qualsiasi unità rilevata, di appartenere ad una fra due o più classi adiacenti dell'albero di segmentazione individuato) con i parametri ottimizzati potrebbe, inoltre, fornire ulteriori informazioni all'indagine: una pendenza molto ripida della sigmoide esemplificata in Fig. 4, tanto da farla assomigliare maggiormente ad un “gradino”, potrebbe essere dovuta a qualche evento particolare verificatosi nel periodo considerato (ad esempio, una “manovra finanziaria” contemplante il blocco delle assunzioni nel settore pubblico).

L'applicazione del metodo, tuttavia, ha trovato non poche difficoltà in quanto il toolbox *Matlab* adoperato non permette di escludere dall'analisi i parametri delle membership delle variabili nominali (cfr. 3.3.2). Ciò ha comportato, in fase di addestramento della rete neurale, frequenti arresti dell'algoritmo in minimi relativi e non assoluti dell'energia (cfr. 3.1.6). Il metodo potrebbe, pertanto, essere migliorato implementando un programma ad hoc.

Gli Autori ritengono, inoltre, di poter ridurre ulteriormente l'errore di classificazione con un modello di Sugeno del primo ordine (cfr. 3.1).

5. Conclusioni

I risultati della metodologia proposta dagli Autori nel presente lavoro, pur se non ottimali, si presentano come una buona base di partenza per migliorare la precisione dell'attribuzione dell'esito lavorativo a partire da informazioni pregresse, e più in generale per risolvere problemi di previsione nella classificazione.

La variabile risposta (dicotomica) su cui si è investigato è l'aver trovato lavoro entro un anno di tempo dalla laurea. Sono stati sperimentati vari metodi di segmentazione, che però hanno fatto rilevare errori di classificazione di circa il 35%; tale cospicuo livello di errore è presumibilmente dovuto al fatto che la metodologia di segmentazione, anche quando contempra tecniche di *look-ahead* (Fabbris 1997), fa tuttora uso di metodi matematici di ricerca di minimi funzionali di tipo *hillclimb*, efficienti con funzioni unimodali ma non altrettanto con quelle polimodali (cfr. Delvecchio 2004).

Gli Autori hanno tentato, perciò, di migliorare la precisione della classificazione facendo uso di tecniche fuzzy neuro adattative. In particolare, l'albero di segmentazione ricavato con algoritmo C@rt è stato utilizzato per ricavare le regole di un sistema fuzzy. Successivamente, tramite una rete neurale, sono stati aggiustati i parametri delle funzioni membership, in maniera tale da migliorare il sistema fuzzy stesso. Si è ottenuto, in tal modo, un miglioramento dell'errore di classificazione di oltre il 10% rispetto a quanto ottenuto con gli usuali algoritmi di segmentazione.

L'albero di segmentazione risultante fornisce informazioni di interesse immediato, anche se a volte prevedibili: ad esempio, a posteriori appare logico che chi ha scelto di proseguire la propria formazione professionale con tirocini o stage, dopo aver fatto proprie le nozioni eminentemente teoriche impartite all'Università, ha avuto meno tempo e occasioni di inserirsi nel mondo del lavoro.

Ciò che può essere importante, e che ci si ripromette di approfondire in altra occasione, è definire quali possano essere le sottopopolazioni di laureati che con queste regole sono classificate meglio, indagando quindi i motivi del residuo errore di classificazione in modo da accrescere ancora la capacità previsiva del metodo.

Bibliografia

- AA.VV. (1999), *Fuzzy Logic Toolbox for Use with MATLAB*, User's Guide Version 2, MathWorks, Inc..
- BREIMAN L., FRIEDMAN J.H. OLSHEN R.A., STONE C.J. (1984) *Classification and Regression Trees*, Wadsworth Inc., Belmont California.
- CROCETTA C., D'OVIDIO F. (2003) La valutazione dell'inserimento lavorativo dei laureati all'Università di Foggia attraverso un'analisi di segmentazione, in: M. CIVARDI (a cura di) *Transizione Università-Lavoro: la definizione delle competenze*, CLEUP, Padova: 111-132.
- CROCETTA C., DELVECCHIO G. (2003) Una misura fuzzy della soddisfazione della formazione universitaria per l'ingresso nel mondo del lavoro, in: M. CIVARDI (a cura di) *Transizione Università-Lavoro: la definizione delle competenze*, CLEUP, Padova: 148-169.
- DELVECCHIO G. (2002) Un approccio fuzzy per la valutazione del rischio da mobbing, in: G. PUGGIONI (a cura di) *Modelli e metodi per l'analisi di rischi sociali e sanitari*, vol. 2, CLEUP, Padova: 248-266.
- DELVECCHIO G. (2004) Gli algoritmi genetici per la determinazione dei massimi e minimi vincolati nello studio della quantificazione delle mutabili ordinali, in: E. AURELI CUTILLO (a cura di), *Strategie metodologiche per lo studio della transizione Università-lavoro*, CLEUP, Padova: 177-198.
- FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.
- FABBRIS L., MARTINI M. C. (2002) Analisi di segmentazione binaria con una variabile dipendente trasformata in *logit*, in: G. PUGGIONI (a cura di) *Modelli e metodi per l'analisi di rischi sociali e sanitari*, CLEUP, Padova: 21-36.
- JANG J. S. R. (1993) ANFIS: Adaptive Network Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23 (maggio 1993), **3**: 665-685.
- KASS G. (1980) An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, **29.2**: 119-127.
- KOSKO B. (1992) *Neural Networks and Fuzzy Systems: a Dynamical System Approach to Machine Intelligence*, Prentice-Hall, Englewood Cliffs.
- KOSKO B. (1995) *Il fuzzy-pensiero, teoria e applicazioni della logica fuzzy*, Baldini & Castoldi, Milano.
- MAMDANI E. H., ASSILIAN S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies*, Vol. 7, n. **1**: 1-13.
- SCHIEVANO C. (2002) *LAID-OUT.1*: un programma per l'analisi di segmentazione binaria con riferimento ad una variabile dicotomica trasformata in *logit*, in: G.

- PUGGIONI (a cura di) *Modelli e metodi per l'analisi di rischi sociali e sanitari*, CLEUP, Padova: 21-36.
- SCHIEVANO C. (2003) Determinazione della numerosità minima dei gruppi nell'analisi di segmentazione con una variabile dipendente trasformata in *logit*, in: L. FABBRIS (a cura di) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*, CLEUP, Padova: 395-400.
- SONQUIST J. A. (1970) *Multivariate Model Building. The Validation of a Search Strategy*, Institute for Social Research, The University of Michigan, Ann Arbor (Mich.).
- SUGENO M. (1985), *Industrial Applications of Fuzzy Control*, Elsevier Science Publications Co..
- TAKAGI T., SUGENO M. (1983) Derivation of fuzzy control rules from human operator's control actions, *Proceedings of the IFAC Symposium On Fuzzy Information, Knowledge Representation and Decision Analysis* (luglio 1983): 55-60.

Fuzzy neural-adaptive methodologies to classify the graduates of the University of Bari by employment one year after graduation

Summary: *This study proposes a new methodology to improve the accuracy of a 'positive' evaluation of job placement possibilities, using statistical methods, on the basis of a set of characteristics inherent to the university degree programme or related to the search for employment. The primary aim is to identify some graduate characteristics (specifically, graduates of the University of Bari in 1995-2000) capable of favouring job placement within one year from graduation. For this purpose, a segmentation analysis is carried out on a sample of graduates from the above-mentioned population, the results of which are affected by a considerable classification error. Subsequently the rules resulting from the segmentation analysis are implemented into a neural network system, also using fuzzy methodologies, in order to identify possible improvements in the system of rules and to define the importance of a given configuration of covariates and/or data. The results obtained verify, with greater precision, the impact of personal characteristics and acquired competencies in terms of job placement.*

Keywords: *Job placement, Graduates, University of Bari, Segmentation analysis, Fuzzy neural-adaptive methodologies, ANFIS.*