Progettazione di un sistema fuzzy per l'annotazione automatica di contenuti multimediali

ing. Maurizio Montagnuolo

ing. Alberto Messina
Rai - Centro Ricerche e Innovazione Tecnologica

L'articolo descrive parte dell'attività di ricerca svolta presso il Centro Ricerche e Innovazione Tecnologica della Rai nel campo della documentazione automatica degli archivi.
Questa ricerca è realizzata nell'ambito della collaborazione tra Centro Ricerche Rai e Università degli studi di Torino. La borsa di dottorato di ricerca dell'Ing. Montagnuolo è sponsorizzata da euriX s.r.l.

Sommario
Questo articolo presenta un framework per la caratterizzazione automatica di contenuti multimediali, basato su tecnologie di data mining, Semantic Web e logica fuzzy. L'approccio proposto è basato sul presupposto che i fruitori di contenuti multimediali utilizzino implicitamente un sistema di regole di inferenza fuzzy per la caratterizzazione semantica dei contenuti audiovisivi, e sull'asserzione che le regole utilizzate dai fruitori dei contenuti multimediali per inferire i concetti semantici di alto livello, nonché i concetti semantici stessi, non sono definite a priori su classificazioni statiche, bensì definite dinamicamente, sulla base degli schemi caratteristici derivati dalle caratteristiche spazio-temporali estratte dagli oggetti multimediali. L'uso delle ontologie [16] permette di esprimere i concetti semantici (oggetti, genere, eventi, azioni, ecc.) in modo formale, in maniera dipendente dal dominio e comprensibile dalle macchine, mentre l'uso della logica fuzzy permette di esprimere analiticamente il grado di incertezza insito nei modelli di classificazione.

1. Introduzione

Grazie al veloce sviluppo delle tecnologie informatiche, la disponibilità di contenuti digitali multimediali è in continua crescita[Nota 1]. In ambito professionale, i broadcaster televisivi pubblici europei rappresentano una miniera di contenuti di prima importanza in questo senso. In questo dominio applicativo in genere si rende necessario lo sviluppo di applicazioni che permettano l'organizzazione ed il recupero efficiente dell'informazione all'interno di database multimediali di ingenti dimensioni. Si consideri, ad esempio, il caso del reperimento di spezzoni audio/video soddisfacenti determinati criteri descrittivi, che rappresenta una delle operazioni più importanti, ma al contempo onerose, nell'ambito della produzione radiotelevisiva. Tipicamente, le operazioni di annotazione, indicizzazione e predisposizione alla ricerca degli archivi audiovisivi, costringono

Nota 1 - Si considerino ad esempio i motori di ricerca video YouTuBe (http://www.youtube.com/) e Google Video (http://video.google.it/)

i documentatori a diverse ore di lavoro manuale per l'annotazione degli archivi stessi, pertanto rendendo l'intero processo inefficiente e costoso. Dalla difficoltà di ottenere in tempi brevi una documentazione oggettiva e completa, derivano, infatti, gli alti costi dovuti alla produzione del materiale multimediale che gravano sull'industria dell'audiovisivo. Conseguentemente, si sta mettendo in opera, da parte dell'industria audiovisiva, un ingente sforzo nella ricerca di modelli produttivi che, sfruttando tutte le possibilità offerte dalle moderne tecnologie, consentano di ottenere notevoli tagli nei costi di produzione.

Piuttosto ovviamente, gli archivi multimediali sono una miniera di contenuti cui attingere per le nuove produzioni, e pertanto i metodi di reperimento ed accesso al materiale archiviato diventano sempre più importanti, nell'economia industriale e di gestione di un'impresa audiovisiva. Per la messa in opera di tali metodi, è necessario il coinvolgimento di ricercatori impegnati in diverse discipline dell'ingegneria, della fisica e dell'informatica, tra le quali possiamo annoverare: *Image Analysis, Pattern Recognition, Computer Vision, Database Multimediali, Reti Neurali e Software Engineering*. Nella comunità tecnico-scientifica internazionale, attualmente i maggiori sforzi sono concentrati nello studio di sistemi che, mediante una minima interazione con un utente umano, permettano di fornire una descrizione ed una classificazione di alto livello del contenuto di un oggetto multimediale. Infatti, considerando l'esempio degli archivi audiovisivi, la maggior parte degli attuali sistemi di *video retrieva*l basati sul contenuto (CBVRS – Content Based Video Retrieval Systems) utilizzano modelli matematici/statistici per la descrizione delle caratteristiche (*features*) di basso livello, quali il colore, la trama, la forma, l'audio ed il movimento. Questi modelli sono affetti da una grave problematica dovuta al fatto che un utente umano usa effettuare un'analisi del contenuto d'alto livello (semantico), mentre non ha in genere esplicita conoscenza della sua descrizione mediante modelli matematici. Di conseguenza, le metriche per valutare il grado di somiglianza tra oggetti multimediali sono spesso diverse se considerate dal punto di vista dell'utente o da quello della macchina. Per ovviare al suddetto problema è necessario prevedere degli strumenti automatici che permettano di determinare una correlazione tra le caratteristiche estratte di basso livello e le caratteristiche semantiche di alto livello. In quest'ambito si inserisce la nostra ricerca, avente lo scopo di analizzare lo stato dell'arte dei sistemi automatici di annotazione di oggetti multimediali, valutandone una possibile implementazione ed i relativi rischi, costi e benefici derivanti dal suo utilizzo nell'industria audiovisiva (per una prima analisi di queste problematiche vedi anche [15]).

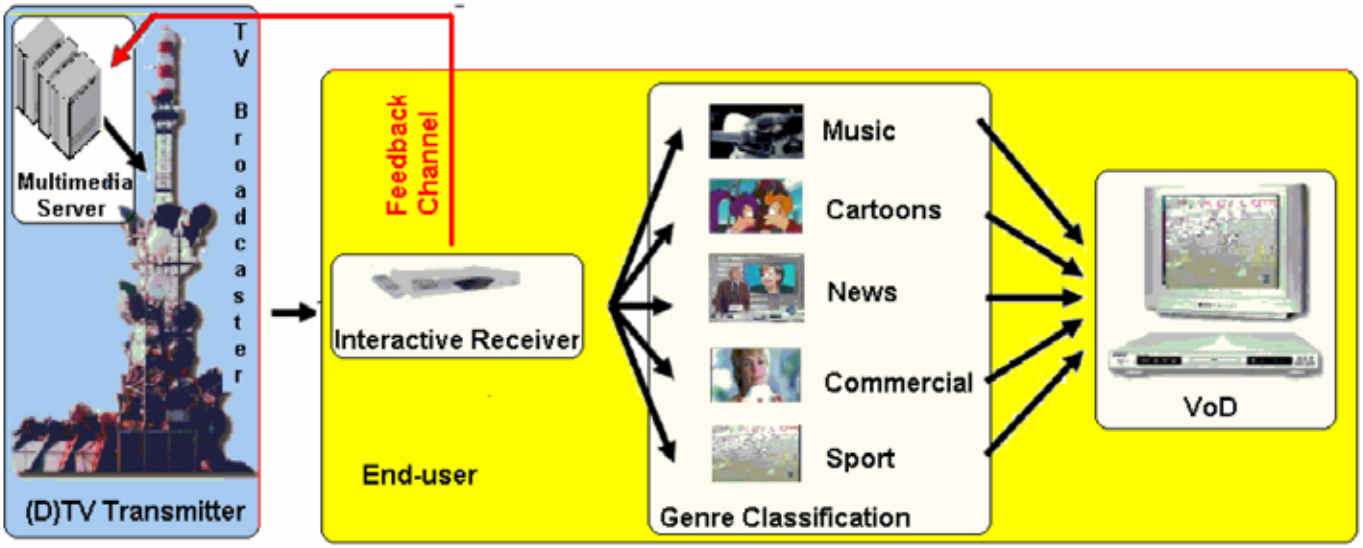
2. Semantica in Ambito Radiotelevisivo

Nel dominio radiotelevisivo, in cui operiamo, l'annotazione semantica dei programmi televisivi esprime informazioni semantiche associate allo scopo del programma stesso (ad esempio *'intrattenimento' (entertainment)*, *'informazione' (information)*, *'comunicazione' (communication)*, etc.) al suo genere (ad esempio *'talk-show'*, *'evento sportivo' (sport)*, *'telegiornale' (newcast)*, etc.), a concetti strutturali ed editoriali (ad esempio *'intervista' (interview)*, *'rapporto giornalistico' (report)*, *'conduzione' (anchorperson)*, etc. o ad eventi (ad esempio *'goal'*, *'esplosione'*, etc.).

Tra i diversi tipi di informazione sopraccitati, crediamo che la caratterizzazione automatica del genere rappresenti un punto di notevole interesse nella comprensione e classificazione dei contenuti multimediali. Tale tecnica, nota con il termine di *Teoria del genere*, permette di classificare i programmi televisivi in categorie rappresentative, caratterizzanti le intenzioni del regista, il ritmo e la struttura del programma, nonché ciò che uno spettatore si aspetta di osservare visionando il programma stesso.

A conferma dell'importanza che le tecniche di

Fig. 1 - Esempio di applicazione di Video-on-Demand (VoD) e televisione interattiva (iTV).
I contenuti multimediali sono classificati ed archiviati nei server del produttore televisivo.
L'utente può, per mezzo di un ricevitore interattivo, accedere ad uno specifico tipo di programma da lui desiderato e visionarlo sul proprio televisore.



classificazione automatica del genere (*Automatic Genre Classification* - AGC) rivestono nell'ambito dell'annotazione di archivi radiotelevisivi, sia i grandi broadcaster televisivi (tra cui Rai), sia le industrie di elettronica di consumo stanno prendendo parte attiva nella sperimentazione di tali tecniche. Infatti, tali strumenti permettono ai broadcaster di gestire in modo più efficiente i propri archivi multimediali, diminuendo di conseguenza i costi di produzione; inoltre permettono ai singoli utenti di creare liste di programmi personalizzate quando si utilizzano servizi di Video-on-Demand (VoD) e TV interattiva (iTV). In questo scenario, mostrato in figura 1, dato un flusso video digitale in ingresso, un ricevitore interattivo ne permetterà la segmentazione, indicizzazione, classificazione e memorizzazione in un hard disk in esso integrato. Quindi, un utente potrà accedere ad uno specifico tipo di programma da lui desiderato semplicemente navigando, tramite una pratica interfaccia grafica (Graphical User Interface - GUI) visualizzata sul proprio apparecchio televisivo, le cartelle situate nella memoria del ricevitore stesso e organizzate secondo generi e sottogeneri familiari all'utente.

Ma come è possibile costruire concretamente un sistema che attui in maniera efficiente questo tipo di scenari? Si cercherà di fornire una risposta esauriente a questo quesito nei paragrafi successivi.

3. Considerazioni sulla Classificazione Automatica del Genere

Fischer et al. [1] sono considerati i pionieri dello studio di metodologie per la classificazione automatica del genere di sequenze video. La loro ricerca, che ebbe inizio nel 1995, consisteva nella classificazione di cinque generi televisivi

(*news*, *commercials*, *cartoons*, *tennis* e *formula 1*) utilizzando una combinazione di caratteristiche audiovisive direttamente estraibili dalle sequenze video. Sulla base di tale ricerca seguirono altri approcci, distinguibili per la scelta dei descrittori e dei classificatori utilizzati [2, 3, 4, 5, 6, 7, 8]. Sebbene, a causa della non uniformità nella scelta del database di test, dei descrittori e del tipo di classificatore non sia possibile confrontare direttamente le prestazioni dei diversi metodi proposti, è possibile notare alcuni punti deboli comuni tra essi.

Per prima cosa, tutti i metodi proposti sono mirati alla classificazione statica in poche categorie predefinite, utilizzando un approccio ispirato dal paradigma "*è questo o quel tipo di genere*" (ad esempio, *"questo video è sport"*). In altri termini, pochi autori hanno considerato il caso in cui un contenuto multimediale sia caratterizzato da generi differenti. Consideriamo ad esempio la finale dei mondiali di calcio disputata in Spagna nel 1982. Questo programma era a contenuto sportivo nel 1982, ma ad oggi potrebbe essere classificato sia come *pubblicità*, sia come *documentario*. Ciò introduce una seconda limitazione, riguardante il fatto che il linguaggio televisivo è un linguaggio vivo che si evolve nel tempo in accordo con l'evoluzione linguistica, sociale e culturale della comunità in cui e per cui i programmi sono prodotti. Pertanto, confrontare programmi prodotti in epoche diverse o anche trasmessi su canali differenti, potrebbe rivelarsi un compito arduo, dal quale potrebbero inoltre derivare risultati inattesi od errati nel processo di annotazione del materiale audiovisivo. In termini astratti, si potrebbe dire che le caratteristiche di contenuto, strutturali e cognitive di un oggetto multimediale sono associabili a generi e sottogeneri in stretta dipendenza da una informazione contestuale, la quale è nativamente estranea ad essi, ma fondamentale per la definizione del genere stesso. Questa informazione contestuale è difficilmente rappresentabile in maniera analitica, ma di certo le condizioni storiche, le caratteristiche editoriali del canale di pubblicazione sono ad esempio elementi cardine in questo senso. Anche per questi motivi, gli autori di questa ricerca ritengono opportuno parlare di *caratterizzazione* più che di *classificazione*, intendendo per classificazione il complesso processo secondario che associa una caratterizzazione ad un concetto semantico per via di un contesto di fruizione.

Un ultimo problema che vogliamo menzionare riguarda il fatto che gli approcci sino ad ora proposti non hanno tenuto in considerazione l'informazione derivante dall'analisi spazio-temporale delle relazioni esistenti tra un intero programma e le sottostorie in esso contenute (ad esempio, un talk-show ed ogni singolo tema in esso discusso). Infatti, i database di test sono stati solitamente costruiti *ad-hoc* montando in successione piccole sequenze video aventi come unica caratteristica comune il fatto di appartenere, *a giudizio dell'autore della ricerca*, allo stesso genere; di conseguenza, le relazioni gerarchiche tra un intero programma ed i capitoli in esso contenuti non sono state prese in considerazione nell'analisi dei modelli di classificazione.

4. Progettazione di un Framework per l'Annotazione Automatica di Contenuti Multimediali

4.1 Requisiti

E' nostra opinione che nella progettazione di un sistema innovativo per l'annotazione automatica di contenuti multimediali devono essere soddisfatti i seguenti requisiti [17]:

- Prevedere un modello concettuale dei dati in grado di caratterizzare il dominio applicativo e definire le relazioni esistenti tra i diversi livelli di informazione insiti in un oggetto multimediale;

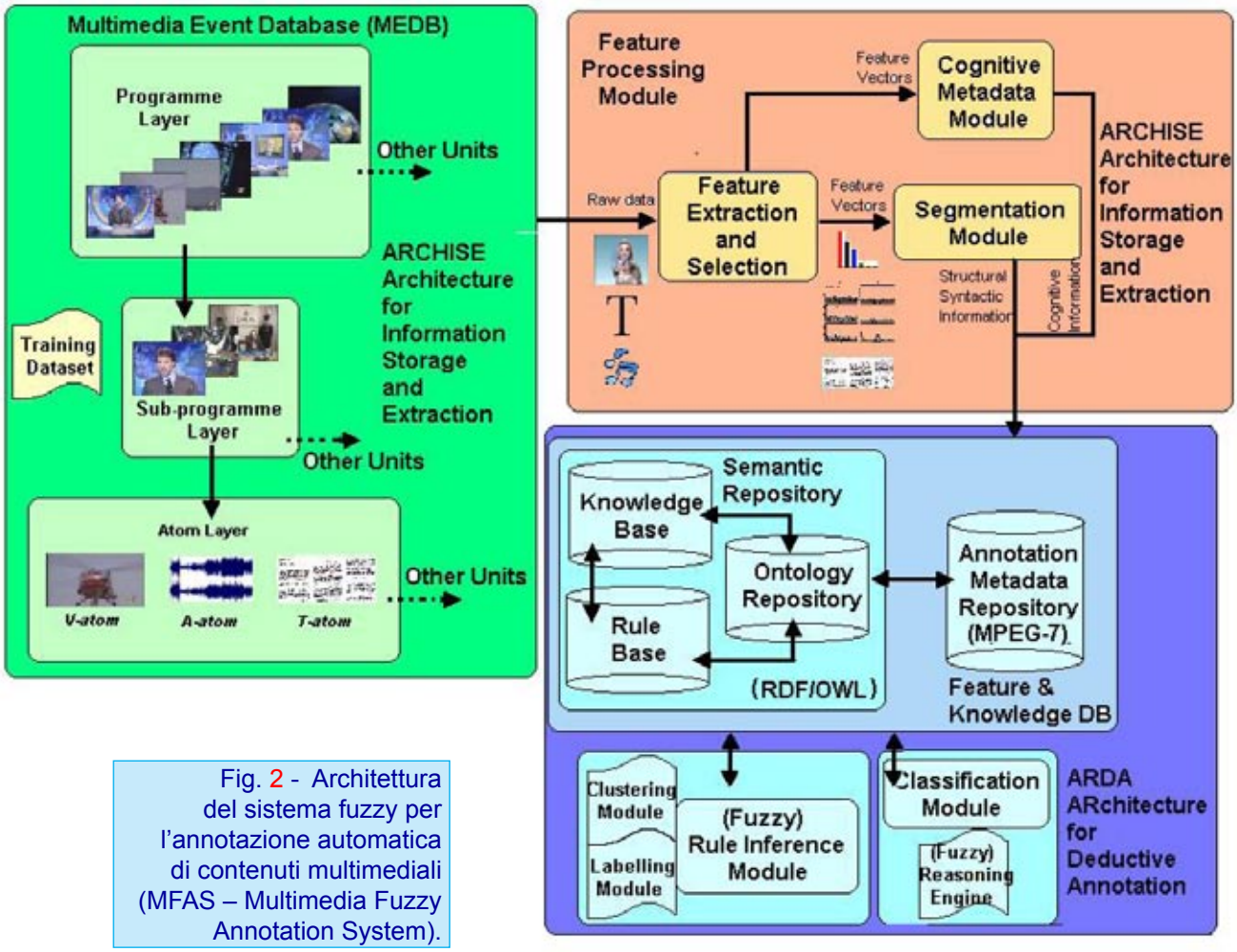


Fig. 2 - Architettura del sistema fuzzy per l'annotazione automatica di contenuti multimediali (MFAS – Multimedia Fuzzy Annotation System).

- Definire un modello di database per la rappresentazione efficiente del contenuto degli oggetti multimediali e delle relazioni esistenti tra oggetti differenti;
- Sviluppare un sistema che facendo uso delle tecnologie del Semantic Web, delle tecniche di Pattern Recognition e di regole di inferenza, permetta di associare le caratteristiche audiovisive di basso livello (automaticamente estraibili dagli oggetti multimediali) a concetti e classi semantiche definite secondo ontologie personalizzate in contesti di riferimento definiti.

4.2 Architettura del Sistema

La figura 2 mostra lo schema a blocchi di un sistema fuzzy per l'annotazione semantica di contenuti multimediali (MFAS - Multimedia Fuzzy Annotation System). Un sistema MFAS deve prevedere diversi sottosistemi cooperanti tra loro, ognuno dei quali designato all'analisi di uno specifico aspetto caratterizzante l'oggetto multimediale analizzato. Inoltre, a differenza delle immagini fisse come le fotografie, l'analisi di una sequenza multimediale deve tenere in considerazione aspetti di natura diversa, ognuno dei quali descrivente una specifica peculiarità

della sequenza stessa. Occorre ricordare che una sequenza video non può essere considerata come una successione di immagini indipendenti, poiché tutti i fotogrammi della sequenza sono tra loro strettamente correlati. Le transizioni tra un fotogramma ed il successivo, e le variazioni delle caratteristiche di colore, trama, e forma osservate analizzando gruppi di fotogrammi possono essere più importanti, ai nostri fini, dell'informazione contenuta in un singolo fotogramma. Inoltre, ogni oggetto ha caratteristiche peculiari che lo distinguono dagli altri, (velocità d'azione e di ripresa, linguaggio visivo, audio, ecc.) secondo la classe cui appartiene.

L'architettura proposta fa uso della logica fuzzy e tecniche di rule mining al fine di simulare il ragionamento umano nel compito di aggregare oggetti multimediali "simili tra loro". Riteniamo che questo compito elementare sia la base del processo di classificazione dei contenuti multimediali che avviene durante la fruizione dei contenuti stessi. L'uso di metodi di tipo fuzzy è preferibile ai metodi di classificazione classici (utilizzanti ad esempio reti neurali, reti bayesiane, ecc), poiché permette di tenere in considerazione la natura polivalente dei dati da analizzare. Infatti, poiché ogni singolo oggetto multimediale può appartenere ad una, nessuna, o più classi semantiche contemporaneamente, è più corretto fornire, per ogni oggetto, il suo grado di appartenenza ad ogni classe elementare definita, piuttosto che fornirne una classificazione netta del tipo "*questo oggetto appartiene alla classe A, piuttosto che alla classe B*". Con questa tecnica, l'obiettivo di costruire una classificazione semantica finale, passa attraverso un'analisi polivalente delle caratteristiche di contenuto.

Il sistema è composto da due moduli principali, brevemente descritti nei paragrafi seguenti:

i. Il modulo ARCHISE (**ARCH**itecture for **I**nformation **S**torage and **E**xtraction), utilizzato per estrarre l'informazione di basso-medio livello sul contenuto di un oggetto multimediale e per conservare l'insieme di oggetti di riferimento per la costruzione dei modelli di caratterizzazione (MEDB);

ii. Il modulo ARDA (**AR**chitecture for **D**eductive **A**nnotation), che è un sistema esperto che, simulando il ragionamento umano, permette di effettuare l'annotazione semantica degli oggetti multimediali.

4.2.1 Il Multimedia Event Database (MEDB)

Nel nostro scenario sperimentale, il Multimedia Event Database (MEDB) contiene circa 100 ore di programmi televisivi tratti dalle tre reti RAI e da altre emittenti locali. Ogni programma archiviato costituisce una *Programme Unit*, ovverosia un'entità multimediale fruibile singolarmente o appartenente ad un'altra unità. Al fine di mantenere la struttura coerente con le notazioni presenti in letteratura si è deciso di suddividere il materiale archiviato in sette classi semantiche, ciascuna di esse rappresentante uno specifico genere televisivo. Le classi individuate sono: *'Newscast'*, *'Commercials'*, *'Cartoons'*, *'Football'*, *'Musics'*, *'Weather Forecasts'* e *'Talk Shows'*. Si noti che questa classificazione è usata per motivi meramente pratici di organizzazione del database, e non rappresenta la finalità della ricerca, che è invece, ricordiamo, quella di caratterizzare i contenuti multimediali attraverso un insieme di informazioni che servano ad associare in maniera dinamica, attraverso un processo dipendente dal contesto, gli oggetti multimediali stessi a classi semantiche significative dal punto di vista del fruitore.

4.2.2 Il Modulo di Processamento del Contenuto Audiovisivo

Il modulo di feature processing (FPM) è costituito da diversi sottosistemi atti all'estrazione ed al processamento del contenuto audiovisivo delle Programme Unit.

Il modulo di estrazione e selezione delle feature (FESM – Feature Extraction and Selection) estrae i descrittori audiovisivi di basso livello da ogni Programme Unit, e li rappresenta in conformità allo standard MPEG-7 [9].

Il modulo di segmentazione (SM) opera una segmentazione spazio-temporale della Programme Unit. Ogni sequenza è dapprima suddivisa in scene aventi caratteristiche visive omogenee utilizzando un algoritmo di similarità tra fotogrammi. Successivamente, un algoritmo di Image Clustering permette di raggruppare tra loro tutte le scene aventi grado di somiglianza superiore ad una certa soglia. Per ogni gruppo di scene simili (cluster) è fornito un identificatore univoco, il numero di scene ad esso appartenenti e i fotogrammi inziale, mediano e finale di ogni scena, nonché la posizione temporale di ciascun elemento facente parte del cluster.

Il modulo relativo all'estrazione dell'informazione cognitiva è costituito da motori di riconoscimento vocale (Speech-to-text) e facciale (Face recognition), provvedendo quindi ad un'estrazione di un tipo di informazione audiovisiva di medio livello semantico.

4.2.3 Il Modulo di Annotazione Semantica

ARDA (Architecture for Deductive Annotation) è un sistema intelligente mediante il quale è possibile inferire ed annotare automaticamente i concetti semantici intrinseci negli oggetti multimediali.

Tale procedura è realizzata utilizzando:

i. Un archivio semantico, contenente le ontologie descriventi i concetti da estrarre e rappresentare, la base di conoscenza del dominio applicativo e le regole d'inferenza;

ii. Un modulo fuzzy per il raggruppamento delle Programme Unit in accordo con le caratteristiche audiovisive, cognitive e strutturali da esse estratte;

iii. Un modulo di classificazione che è usato per assistere l'operatore nelle operazioni di classificazione e archiviazione del materiale audiovisivo.

4.2.3.1 L'uso delle Ontologie

Le ontologie [16] sono uno strumento utile nelle applicazioni di annotazione di contenuti multimediali poiché permettono di esprimere i concetti semantici (oggetti, genere, eventi, azioni, ecc.) in modo formale, dipendente dal dominio e comprensibile dalle macchine. Inoltre, è facile sviluppare ontologie personalizzate per la descrizione sia delle proprietà percettive e strutturali degli oggetti multimediali, sia delle relazioni esistenti tra tali proprietà. Infine, le ontologie sono strumenti scalabili e flessibili, che permettono dunque una facile integrazione e riuso delle stesse. In termini molto semplicistici un'ontologia è un dizionario di termini organizzato, nel quale è possibile definire relazioni gerarchiche tra termini, relazioni funzionali tra termini e condizioni che esprimono le proprietà delle relazioni stesse (p.es. transitività, relazione inversa).

Ovviamente, questa tecnologia presenta anche degli svantaggi. Il problema più grande consiste nel fatto che non esistono ancora ontologie standardizzate, e questo introduce di fatto delle ambiguità dovute alle diverse modalità con cui lo stesso concetto può essere rappresentato (si consideri ad esempio la lingua utilizzata per esprimere l'ontologia). Un secondo problema riguarda la dimensione delle ontologie stesse. Infatti, più un'ontologia è grande, più problematica risulta la sua gestione e riutilizzabilità. Per questo motivo sarebbe preferibile utilizzare un insieme di ontologie più piccole, ognuna di esse modellante un singolo aspetto del dominio applicativo o dell'informazione analizzata, piuttosto che un'unica grande ontologia. Ovviamente in questo secondo caso occorrerà arricchire la

base di conoscenza del sistema con le relazioni esistenti tra le diverse ontologie.

L'ultimo problema che vogliamo menzionare riguarda la scelta del tipo di linguaggio da utilizzare nella generazione dell'ontologia. Crediamo che la scelta migliore sia rappresentata dall'uso del linguaggio OWL (Web Ontology Language) [10]. Tale scelta è dettata dal fatto che l'OWL è un linguaggio standard raccomandato dal W3C. Inoltre, per la sua struttura e definizione, esso è più flessibile, intuitivo ed espressivo di altri linguaggi, quali ad esempio MPEG-7 XML, RDF e RDFS [11]. Infine, a tutt'oggi sono facilmente reperibili parser e motori di ragionamento automatico basati su tale linguaggio.

Nella nostra sperimentazione la tecnologia delle ontologie sarà primariamente adottata per la rappresentazione delle caratteristiche strutturali dei contenuti audiovisivi (p. es. numerosità dei cluster di immagini, loro mutua relazione) e per la derivazione automatica di proprietà secondarie godute dagli elementi strutturali e di conseguenza dalle Programme Unit che li contengono. Queste proprietà formeranno una base essenziale di informazioni per la fase di aggregazione e caratterizzazione delle Programme Unit.

4.2.3.2 L'uso della Logica Fuzzy

La logica comunemente utilizzata nei settori informatici e tecnologici è quella booleana, seguendo la quale ogni oggetto può essere membro di un unico insieme. Tale assioma è però in contrapposizione a quanto accade nel mondo reale, in cui le entità ad esso appartenenti non possono essere categoricamente riconosciute come appartenenti ad una o ad un'altra categoria in maniera esclusiva, ma possono appartenere ad entrambe con un certo grado di verità. Anche gli oggetti multimediali presentano tale proprietà, potendo appartenere, come già accennato precedentemente, ad una, nessuna, o più classi semantiche contemporaneamente. Consideriamo ad esempio un servizio sportivo all'interno del telegiornale; esso deve essere considerato nella categoria *'newscast'* o nella categoria *'sports'*? A nostro avviso può essere rappresentato da entrambe le categorie.

La logica fuzzy (dall'inglese "sfumato", "sfuocato") [12] è stata introdotta per formalizzare concetti del mondo reale che non possono essere categoricamente riconosciuti come veri o falsi, ma che possono avere un certo grado di verità. Dall'esempio precedente, appare evidente come, per le sue proprietà, la logica fuzzy, sia particolarmente efficace nelle applicazioni d'estrazione e interpretazione dell'informazione. Storicamente, la teoria sugli insiemi fuzzy, che pone le premesse della logica fuzzy, fu introdotta da Lotfi Zadeh[Nota 2] nel 1965, come strumento matematico per rappresentare i diversi livelli d'incertezza in campo linguistico. Inizialmente nacque come estensione della classica teoria degli insiemi.

In termini matematici tradizionali, un elemento dell'universo può appartenere o non appartenere ad un insieme. Cioè, l'appartenenza di un elemento è netta, rigida, (ovvero crisp), come espresso dall'equazione

$$\mu_A(x) = \begin{cases} 1 & se \quad x \in A \\ 0 & se \quad x \notin A \end{cases}$$

Un insieme fuzzy, viceversa, rappresenta una generalizzazione dell'insieme ordinario in cui non vengono solamente indicati gli elementi ad esso appartenenti, ma anche il grado di appartenenza di ciascuno di essi all'insieme stesso. Si perviene quindi alla seguente definizione di un insieme fuzzy:

data una collezione di oggetti x l'insieme A in X è l'insieme di coppie ordinate

$$A = \{(x, \mu_A(x)) : x \in X \}$$

Nota 2 - http://en.wikipedia.org/wiki/Lotfi_Zadeh

in cui $\mu_A(x) \in [0,1]$ indica il grado di appartenenza dell'oggetto x all'insieme A.

Da quanto enunciato, il lettore potrebbe confondere la logica fuzzy con la teoria della probabilità. Questa confusione può essere chiarita, se pensiamo che la logica fuzzy tratta eventi deterministici (ovvero fatti tangibili, misurabili), mentre la teoria della probabilità riguarda la verosimiglianza di eventi non deterministici (ovvero stocastici). Pertanto, la logica fuzzy esprime l'incertezza riscontrabile nella definizione di un concetto, fornendo il grado di similitudine di oggetti nei confronti di particolari proprietà intrinseche dell'oggetto, mentre la teoria della probabilità esprime l'incertezza dell'occorrenza di fenomeni, fornendo la probabilità del verificarsi di un evento sulla base delle sue occorrenze.

Nel nostro sistema la logica fuzzy è utilizzata per effettuare un raggruppamento di Programme Unit aventi caratteristiche strutturali, cognitive o audiovisive similari; ciò viene effettuato applicando l'algoritmo Fuzzy C-Means (FCM) [13] monodimensionale su ogni vettore n-dimensionale contenente i valori numerici dei descrittori estratti da ogni Programme Unit. L'algoritmo FCM è una procedura iterativa in cui ogni cluster c_i (*i = 1, 2, …, C*) è considerato come un insieme fuzzy. Ogni elemento dello spazio considerato x_j (*j = 1, 2, …, P*) può pertanto appartenere simultaneamente a più cluster con differenti indici di appartenenza u_{ij} (*i = 1, 2, …,* C; *j = 1, 2, .., P*), variabili ad ogni iterazione dell'algoritmo. L'agoritmo opera come segue:

1. *INIZIALIZZAZIONE:*

a. *Scegliere il numero di cluster C per il numero P di elementi da raggruppare;*

b. *Scegliere il valore del livello di ambiguità (dall'inglese fuzzyness) F (tipicamente F = 2). Tale parametro è utilizzato per controllare il livello di ambiguità tra i diversi cluster. Per F = 0 i cluster sono di tipo classico, in cui ogni elemento può appartenere ad un solo cluster (algoritmo k-means), mentre per F >> 0 il livello di ambiguità tra i cluster aumenta;*

c. *Scegliere la funzione di distanza, utilizzata per determinare il grado di appartenenza di un elemento ai cluster;*

d. Scegliere un criterio di distanza tra le matrici dei gradi di appartenenza, stimate per iterazioni successive $\|U(t+1)-U(t)\|$;

e. Scegliere la soglia δ che determina la condizione di termine dell'algoritmo;

f. Inizializzare la matrice U(t=0)

2. *AGGIORNAMENTO DEL BARICENTRO DI OGNI CLUSTER:*

$$\forall i = 1,2,\ldots,C \quad b_i(t+1) = \frac{\sum_{j=1}^{P} [u_{ij}(t)]^F X^j}{\sum_{j=1}^{P} [u_{ij}(t)]^F}$$

3. *AGGIORNAMENTO DEI GRADI DI APPARTENENZA:*

$$\forall i = 1,2,\ldots,C \ \ e \ \ \forall j = 1,2,\ldots,P$$
$$u_{ij}(t+1) = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|X^j - b_i(t+1)\|}{\|X^j - b_k(t+1)\|} \right)^{\frac{2}{F-1}}}$$

4. *CONDIZIONE DI TERMINE:*

if $\|U(t+1)-U(t)\| \le \delta$

termina algoritmo

else

t = t+1

goto 2

L'algoritmo FCM ha il vantaggio di ottenere un buon compromesso tra l'efficacia della aggregazione (clustering) ed il costo computazionale. Dall'altra parte, la sua efficienza dipende fortemente dalla scelta iniziale della matrice U(t=0).

Dopo aver applicato l'algoritmo FCM ogni Programme Unit sarà identificata, per ogni data caratteristica da un vettore contenente i gradi di appartenenza ad ogni cluster; mediante un'operazione di fuzzificazione potrebbe essere pertanto possibile associare i centroidi di ogni cluster ad un'etichetta linguistica definita nell'ontologia, producendo una mappatura tra i descrittori di basso livello ed i concetti semantici di livello superiore.

4.2.3.3 L'uso di Tecniche di Data Mining

Il modulo di inferenza delle regole (RIM) è utilizzato per identificare le relazioni più frequenti esistenti tra i descrittori e di concetti. Tali relazioni sono espresse sotto forma di regole di associazione $R = \{r_1, r_2, ..., r_m\}$, memorizzate in un apposito database (RB – Rule Base). Ogni regola associa $r : F^{m,n} \rightarrow C$ una sequenza di descrittori di basso livello f appartenente all'insieme dei F dei descrittori, alle classi semantiche di alto livello, definite nell'insieme dei concetti C ed espressi da apposite ontologie. Le regole sono generate applicando tecniche di data mining [14] ad ogni transazione $T = \{t_1, t_2, ..., t_m\}$. Formalmente, $t_i = \bar{f} \cup \bar{c}$, in cui f è una possibile combinazione dei vettori delle feature e c una possibile combinazione dei concetti semantici rappresentati nell'ontologia. Queste tecniche saranno applicate a valle dell'applicazione di FCM sugli insiemi fuzzy generati.

5. Conclusioni

L'articolo ha esaminato e descritto a grandi linee l'architettura di un sistema per la caratterizzazione e l'annotazione automatica di contenuti multimediali, facente uso della logica fuzzy, tecnologie del semantic web e tecniche di data mining. Crediamo che, utilizzando tecniche di analisi del contenuto audiovisivo multimodali, in unione con le tecnologie del Web semantico e sistemi intelligenti sia possibile ottenere un sistema flessibile ed affidabile, in grado di avanzare lo stato nell'arte nella minimizzazione del gap semantico esistente tra la rappresentazione delle caratteristiche audiovisive di basso livello e la deduzione di concetti semantici di livello più alto. Questo sistema è attualmente oggetto di studio e sperimentazione presso il Centro Ricerche e Innovazione Tecnologica della RAI, attraverso la collaborazione di alcuni ricercatori RAI e di personale dell'Università degli studi di Torino che fruisce della sponsorizzazione della borsa di dottorato di ricerca da parte di euriX s.r.l.

Bibliografia

1. Fischer S., Lienhart R. and Effelsberg W. (1995), Automatic recognition of film genres, in ACM Multimedia 1995, San Francisco, USA, pp. 295-304.

2. Truong B.A. and Dorai C. (2000), Automatic genre identification for content-based video categorization, Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain, Vol. 4, pp. 230-233.

3. Xu L.Q. and Li Y. (2003), Video classification using spatial-temporal features and PCA, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME2003), Baltimore, MD, USA.

4. Glasberg R., Elazouzi K. and Sikora T. (2005), Cartoon-Recognition using Visual-Descriptors and a Multilayer Perceptron, WIAMIS, Montreux, May 28-31.

5. Dimitrova N, Agnihotri L. and Wei G. (2000), Video classification based on HMM using text and faces, In European Signal Processing Conference, Tampere, Finland.

6. Liu Z., Huang J. and Wang Y. (1998), Classification of TV programs based on audio information using hidden Markov Model, Proc. of the IEEE Signal Processing Society Workshop on Multimedia Signal Processing.

7. Roach M.J., Mason J.S.D. and Pawlewski M. (2001),Video genre classification using dynamics, In ICASSP'2001.

8. Dinh P.Q., Dorai C. and Venkatesh S. (2002), Video genre categorization using audio wavelet coefficients, In ACCV 2002.

9. ISO/IEC 15398 Multimedia Content Description Interface, 2001.

10. WWW Consortium (W3C) (2004). Web Ontology Language (OWL), http://www.w3.org/2004/OWL/

11. WWW Consortium (W3C) (1999). Resource Description Framework (RDF), http://www.w3.org/RDF/

12. Timothy J. Ross (2004), Fuzzy Logic with Engineering Applications, 2nd Edition, Wiley

13. Frank Höppner, Frank Klawonn, Rudolf Kruse, Thomas Runkler (1999), Fuzzy Cluster Analysis, Wiley

14. Ian H. Witten, Eibe Frank (2005), Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann.

15. A. Messina, D. Airola Gnota, "Automatic Archive Documentation based on Content Analysis", IBC 2005, Amsterdam 11 September 2005.

16. WWW Consortium (W3C) (2001), Semantic Web, http://www.w3.org/2001/sw/

17. M. Montagnuolo and A. Messina, "Multimedia Knowledge Representation for Automatic Annotation of Broadcast TV Archives", Proceedings of the 4th Special Workshop on Multimedia Semantics (WMS06), Chania, Crete, Greece, June 19-21, 2006, pp. 80-94.