**Lezione 1a**

Per prima cosa indagheremo sulle tecniche, sviluppate negli ultimi trent'anni per il rilevamento del pitch basato solo sull'analisi della forma d'onda e dello spettro del suono. In una seconda fase indagheremo sull’ uso di tecniche di riconoscimento di pattern per recupero di informazioni di alto livello dal suono.

Un Pitch Detector (PD) è un algoritmo software o dispositivo hardware che accetta un segnale audio come proprio input e tenta di determinare la fondamentale periodo di intonazione di quel segnale (Roads, 1996) Il PD ha successo solo su un corpus limitato di suoni: non ha senso cercare di trovare il file altezza di un suono percussivo, impulsi o complessi masse sonore.

I rilevatori di intonazione hanno difficoltà a rilevare l'altezza in seguenti condizioni: - Transitori di attacco: analisi dettagliata dell'attacco di molti gli strumenti rivelano forme d'onda caotiche e instabili. - Low Frequencies: rilevatori di intonazione basati su analisi dello spettro hanno difficoltà con i toni bassi, che richiedono l'uso di PD nel dominio del tempo. Per determinare il periodo del pitch, a almeno tre cicli della forma d'onda dovrebbero essere campioni prima che l'analisi possa iniziare. Per un LA a 55 Hz, tre cicli richiedono 54 ms ... problemi per i sistemi in tempo reale. - Frequenze alte: all'aumentare della frequenza, un periodo di intonazione è rappresentato da un minor numero di campioni. La risoluzione con cui intonazione può essere determinato è, quindi, diminuito. -Tracciamento del tono miope: i PD iniziano ad analizzare il suono su intervalli di tempo da 20 a 50 ms. La percezione umana si basa su segmenti di tempo più lunghi. I PD possono traccia in modo miope i dettagli irrilevanti che sono stati prodotti involontariamente. • Ambiente acustico: l'ambiente acustico all'interno quale uno strumento o una voce si sente influisce sul accuratezza del rilevamento del passo.

La maggior parte degli algoritmi PD è nata da riconoscimento vocale e sintesi vocale ricerca Possiamo categorizzare approssimativamente i PD in cinque classi: –Algoritmi nel dominio del tempo –Tecniche basate sull'autocorrelazione –Tecniche di filtro adattivo –Dominio di frequenza –Modelli dell'orecchio umano

Idea: le tecniche nel dominio del tempo cercano di trovare schemi ripetitivi in la forma d'onda. Esempi: **–Zero-crossing PDs**: misurando l'intervallo tra i successivi zero-crossing della forma d'onda in frame consecutivi, si è in grado di rilevarla frequenza fondamentale. –**Filter-bank zero crossing:** per risolvere il problema di precisione di ZCPD, è preimpostata una preelaborazione da parte di un banco di filtri. L'algoritmo controlla l'ampiezza delle uscite del filtro ed esegue il rilevamento del passaggio per lo zero dei due filtri più bassi che hanno un'ampiezza significativa dopo il filtraggio [Kuhn, 1990] –**Elettroglottografo o larinografo:** un cantante indossa un archetto da collo che rileva gli impulsi emessi dalle corde vocali. L'idea è che sia più facile stimare il tono analizzando gli impulsi delle corde vocali.

I metodi di autocorrelazione confrontano un segnale con versione di se stessa ritardata da successive intervalli. L'obiettivo è trovare schemi ripetuti nel file segnale, indicatori di periodicità. Il PD rileva un passo nel periodo L, se l'autocorrelazione mostra i massimi ai ritardi L, 2L, 3L, ... e L è il divisore minimo al quale maxima può essere trovato.

Rilevatori di passo del filtro adattivi Due tecniche principali:

–Filtro adattivo passa banda: Il segnale di ingresso viene inviato a un filtro passa banda stretto sintonizzabile. La frequenza centrale del filtro passa-banda è sintonizzata al fine di ridurre al minimo la differenza tra x (n) e y (n)

- Filtro pettine ottimale: Questo metodo trova un filtro a pettine che riduce al minimo il segnale di uscita. Per ridurre al minimo il segnale, le tacche devono essere sintonizzate su frequenza dominante dell'ingresso.

Rilevatori di passo nel dominio della frequenza i rilevatori di passo FD cercano i picchi nello spettro che corrispondono a frequenze prominenti. Dopo aver trovato i picchi, il PD deve decidere quali le frequenze sono fondamentali e quali le frequenze sono armoniche o parziali. Problema tipico: l'orecchio umano divide le frequenze su scala logaritmica, mentre STFT lavora su a scala lineare. Questo fatto pone problemi di intonazione rilevamento a bassa frequenza, dove la risoluzione non poteva essere preciso come voleva.

Analisi del cepstrum Possono essere rappresentati molti suoni vocali e strumentali come risultato del filtraggio di un'eccitazione (l'originale vibrazione del suono, possibilmente contenente l'altezza) attraverso una struttura risonante (contenente l'effetto del tratto vocale o del corpo di uno strumento). Il tono è invece più enfatizzato nel segnale di eccitazione rispetto a quello osservato. Obiettivo dell'analisi del cepstrum: estrarre il segnale di eccitazione da quello osservato.

Idea: trasformare la moltiplicazione tra H (f) e X (f) in una somma col logaritmo. Qual è il vantaggio di questa trasformazione?

L'effetto di risonanza causa la modulazione a variazione lenta del segnale di eccitazione simile a un rumore bianco anche 🡪log H (f) sta lentamente variando • Il segnale di eccitazione è rappresentato da non modulato registro spettroà a variazione rapida Anche X (f) varia rapidamente Usando la rappresentazione logaritmica, log H (f) e log X (f) sono ora in relazione additiva, quindi possiamo filtrare il log H (f) dal log Y (f).

Per filtrare l'effetto del log H (f), applichiamo l'inverso Trasformata di Fourier per registrare Y (f) e selezionare solo i campioni che provocano una rapida variazione dell'ampiezza logaritmica dominio della frequenza. Questi sono i campioni con i ritardi temporali più elevati. Schema di calcolo:

PD basati sul modello dell'orecchio Negli ultimi decenni, una comprensione dettagliata del meccanismo dell'udito umano il sistema è stata raggiunta. I PD basati su questi studi modellano esplicitamente la struttura dell'orecchio umano mentre modellano l'effetto del suono prodotto dall'esterno e orecchio medio, coclea e nervo centrale sistema.

Un passo oltre i segnali: la classificazione del suono Oggigiorno è comune combinare un livello basso di strumenti di elaborazione del segnale con software e tecniche dalla ricerca in artificiale intelligenza. L'obiettivo è andare oltre il segnale grezzo per una più profonda comprensione del segnale. Due gruppi: –Sistemi che cercano di simulare le capacità di ascolto di ascoltatori umani formati –Nessun tentativo di emulare ascoltatori umani.

Classificazione basata su modelli: quadro generale Estrazione di descrittori audio sintetici (caratteristiche) dalla forma d'onda del segnale o dai suoi spettro Esempi di descrittori audio: –Energia della forma d'onda –Centroide dello spettro –... I descrittori audio sono raggruppati in un vettore f = [f1, f2, ..., fN] Ogni elemento nel vettore rappresenta un audio descrittore del segnale.

Classificazione basata su modelli: quadro generale Vengono quindi rappresentati i descrittori audio nello spazio [f1, f2, ..., fN] (spazio delle caratteristiche). Esempio con N = 2: vd. Il segnale è in finestra o segmentato (vedi più avanti per il differenza tra segmentazione e windowing). Per ciascuno segmento / finestra in cui si trova un punto lo spazio delle funzioni è generato.

Classificazione basata su modelli: immagine generale Il segnale contiene più eventi audio che devono essere classificati. Se le caratteristiche scelte sono rilevanti per il problema, i punti nello spazio delle caratteristiche cluster in modo che ogni cluster nella funzionalità lo spazio può essere assegnato a una classe audio.

L'obiettivo dei sistemi di classificazione del suono è quello di assegna un'istanza a una classe specifica. Per raggiungere questo obiettivo, una formazione adeguata fase è necessaria. –Un insieme di istanze viene inviato al sistema di addestramento al fine di stimare i cluster all'interno della funzionalità spazio. –La formazione può essere supervisionata o non supervisionata. Faremo riferimento qui ai sistemi supervisionati.

Alcune considerazioni Punti cruciali per una classificazione di successo: –Scelta di una serie di caratteristiche rilevanti: cluster di classi diverse devono essere distinte (cioè non sovrapposte) e i punti all'interno di ogni cluster devono essere vicino al baricentro del cluster. Questo fatto è fortemente influenzato dalla scelta delle caratteristiche. –Scelta di un sistema di classificazione adeguato che esegue il raggruppamento dello spazio delle caratteristiche. –Scelta di un set di formazione pertinente: una scelta di a set di formazione che non è rappresentativo del problema compromette gravemente la capacità di classificazione.

Architettura dei sistemi di classificazione del suono: **vd.**

Finestratura / segmentazione Il windowing si riferisce alla frammentazione del segnale in finestre regolarmente distanziate. Finestre multiple consecutive può contenere contenuti dello stesso evento. La segmentazione si riferisce alla frammentazione del segnale in finestre con spaziatura non regolare in modo che ogni segmento contiene un segnale che è uniforme dal punto semantico di Visualizza. Un algoritmo di segmentazione esegue un approssimativo suddivisione del flusso in eventi. La segmentazione è utile anche quando il costo computazionale è un problema: ci permette di evitare la classificazione di porzioni di il flusso che non è rilevante per il nostro problema (esempio: assenza di discorso in un sistema di riconoscimento vocale)

Estrazione delle caratteristiche Dall'origine dei sistemi di classificazione del suono [Moorer, 1975] a numero di classi di caratteristiche sono state ideate. Considereremo, qui, la categorizzazione delle funzionalità fornite da il consorzio MPEG-7 [Kim, 2004]: - Descrittori di base - Descrittori spettrali di base - Parametri di segnale di base - Descrittori timbrici temporali - Descrittori timbrici spettrali - Rappresentazioni di base spettrale Altri descrittori non standardizzati non possono essere classificati in nessuno di queste classi Nelle diapositive seguenti forniremo una descrizione di queste caratteristiche

Descrittori di base: forma d'onda audio minRange: il limite inferiore dell'ampiezza audio nel frame. maxRange: il limite superiore dell'ampiezza audio nel frame. Il descrittore della forma d'onda audio (AWF) è costituito da serie temporale risultante di questi (minRange, maxRange) coppie per ogni fotogramma.

Descrittori di base: Audio Power L'LLD Audio Power (AP) descrive il temporally potenza istantanea livellata del segnale audio. Il I coefficienti AP sono il quadrato medio della forma d'onda valori s (t) entro fotogrammi successivi non sovrapposti (Lw = hopSize) vd. Formula

Descrittori spettrali di base: inviluppo spettrale audio L'Audio Spectrum Envelope (ASE) è una frequenza logaritmica spettro che può essere utilizzato per generare uno spettro ridotto spettrogramma del segnale audio originale. Si ottiene sommando l'energia dello spettro di potenza originale all'interno di una serie di bande di frequenza. Le bande sono distribuito logaritmicamente (logaritmi in base 2) tra due fronti di frequenza loEdge e hiEdge La risoluzione spettrale r delle bande di frequenza all'interno l'intervallo [loEdge, hiEdge] può essere scelto tra otto valori possibili, compresi tra 1/16 di ottava e 8 ottave (vd 2 pagine dopo)

Descrittori spettrali di base: inviluppo spettrale audio L'Audio Spectrum Envelope (ASE) è una frequenza logaritmica spettro che può essere utilizzato per generare uno spettro ridotto spettrogramma del segnale audio originale. Si ottiene sommando lo spettro di potenza originale all'interno di una serie di bande di frequenza. Le bande sono distribuito logaritmicamente (logaritmi in base 2) tra due fronti di frequenza loEdge e hiEdge La risoluzione spettrale r delle bande di frequenza all'interno l'intervallo [loEdge, hiEdge] può essere scelto tra otto valori possibili, compresi tra 1/16 di ottava e 8 ottave

Descrittori spettrali di base: diffusione dello spettro audio La diffusione dello spettro audio (ASS) è un altro semplice misura della forma spettrale Per un dato frame di segnale viene estratta la funzione ASS prendendo la deviazione dalla radice quadrata (RMS) di spettro dal suo baricentro ASC

Armonicità audio I suddetti LLD spettrali di base danno a rappresentazione levigata degli spettri di potenza. Essi non può riflettere la struttura dettagliata del periodico suoni a causa della mancanza di risoluzione in frequenza. Il descrittore di armonicità audio ne fornisce due misure delle proprietà armoniche di uno spettro: - il rapporto armonico: il rapporto tra la potenza armonica e il totale energia - il limite superiore di armonicità: la frequenza oltre quale lo spettro non può essere considerato armonico.

Armonicità audio - Rapporto armonico Il rapporto armonico (HR) è una misura di proporzione di componenti armoniche in spettro di potenza Per un dato indice di frame l, il file normalizzato la funzione di autocorrelazione del segnale è la prima stimato come:

Il ritardo massimo m corrisponde a periodo fondamentale massimo T0 (o equivalentemente il minimo fondamentale frequenza) che può essere stimata come Per ottenere l'HR, l'autocorrelazione è cercato il massimo, dopo averlo ignorato il picco zero lag:

Armonicità audio - Limite superiore di armonicità Il limite superiore dell'armonicità (ULH) è un file stima della frequenza oltre la quale il il segnale non può essere considerato armonico Procedura: –Il segnale viene filtrato a pettine per essere filtrato l'effetto armonico:

Il fattore G è il guadagno ottimale del pettine filtro, fornito da

Gli spettri di potenza dell'originale e del filtro a pettine segnali (P ’(k’) e P’c (k ’) rispettivamente) vengono calcolati per ogni frame l come descritto per Audio Spectrum Centroid

Per ciascuno degli spettri P ’(k’) e P’c (k ’), tutti i campioni di potenza che cadono oltre una data frequenza bin klim vengono sommati

I rapporti R (klim) sono calcolati sequenzialmente, decrementando klim da klim = kmax fino a prima frequenza bin kulh per la quale R (klim) è minore di una soglia di 0,5

La frequenza corrispondente fulh è data da f (Klow + kulh) Infine, il limite superiore dell'armonicità è dato da

Frequenza fondamentale audio Il descrittore della frequenza audio fondamentale (AFF) stima la frequenza fondamentale f0 di un segnale in segmenti in cui si presume che il segnale sia periodico. Per evitare problemi di pitch halving e pitch raddoppio (vedi [Kim, 2004]), lo spettro-temporale viene utilizzata l'autocorrelazione:

dove –ΓΤΑ è l'autocorrelazione temporale, –ΓSA è l'autocorrelazione spettrale –Β è un fattore di ponderazione

Il ritardo che massimizza lo spettro-temporale si trova l'autocorrelazione:

La frequenza fondamentale è infine data da

Altri descrittori non MPEG-7 Tasso di Zero-Crossing: lo Zero-Crossing Rate è calcolato contando il numero di volte che il file la forma d'onda audio attraversa l'asse zero

Frequenza di roll-off spettrale: il rolloff spettrale la frequenza è definita come la frequenza di seguito quale 85% dello spettro accumulato la magnitudine è concentrata

Flusso spettrale: il flusso spettrale è definito come variazione media del segnale ampiezza tra fotogrammi adiacenti. È calcolato come differenza quadratica media tra due spettrali successivi distribuzioni.

Descrittori timbrici I descrittori timbrici mirano a descrivere caratteristiche percettive dei suoni strumentali. Il timbro si riferisce alle caratteristiche che ne consentono uno per distinguere due suoni uguali in volume, tono e durata soggettiva.

Modello ADSR I descrittori timbrici temporali vengono estratti dall'inviluppo del segnale nel dominio del tempo. L'inviluppo del segnale descrive il cambiamento di energia del segnale ed è generalmente equivalente al cosiddetto ADSR (Attack, Decay, Sustain, Release) di un suono musicale.

Registra il tempo di attacco Il tempo di attacco del registro (LAT) è definito come il tempo del suono l'inviluppo impiega per raggiungere l'ampiezza massima di a segnale da un tempo di soglia minimo. La sua motivazione principale è la descrizione delle insorgenze di singoli campioni sonori da diversi strumenti musicali. LAT è definito come il logaritmo (base decimale) di durata dal tempo T inizia quando il segnale inizia al tempo Tstop quando raggiunge il suo valore massimo (per a suono percussivo) o la sua parte sostenuta (per un suono sostenuto suono, cioè senza fase di decadimento).

Centroide temporale Il TC è definito come la media temporale su l'inviluppo energetico del segnale. Il centroide basato sul tempo risultante è dato da

Coefficienti di cepstrum di frequenza mel I coefficienti di cepstrum di frequenza mel (MFCC) può essere utilizzato come una caratteristica eccellente vettore per rappresentare la voce umana e segnali musicali. In particolare, la parametrizzazione MFCC di la parola si è dimostrata utile per la parola riconoscimento. Viene utilizzata una scala di frequenza mel non lineare, che approssima il comportamento del sistema uditivo.

Per convertire una frequenza f in hertz nel suo equivalente in mel, viene utilizzata la seguente formula:

Gli MFCC si basano su un filtro di sottobanda per mezzo di una serie di filtri triangolari le cui frequenze centrali sono equidistanti secondo la scala mel.

Viene misurata l'energia logaritmica dello spettro all'interno della banda passante di ciascun filtro, risultando in una rappresentazione ridotta dello spettro Infine si ottengono i coefficienti cefalici attraverso una trasformata discreta del coseno (DCT) dello spettro logico ridotto

La stima di derivata e accelerazione di le caratteristiche dell'MFCC sono solitamente aggiunto all'iniziale vettore al fine di prendere in conto il temporale cambiamenti negli spettri VD. SCHEMI

Classificazione del suono: formazione e test Una volta estratte le funzionalità audio dai segnali, vengono inviati al classificatori. Se le caratteristiche scelte sono rilevanti per il problema a portata di mano, si raggrupperanno nel file caratterizzano lo spazio in classi diverse. Ogni cluster fa riferimento a un singolo evento acustico (classe).

Formazione: dato un set di dati "etichettato" (ogni l'elemento nel set di dati è assegnato in passare a un evento acustico), stimare il parametri di una o più funzioni che suddividere lo spazio delle funzioni in porzioni, ciascuna appartenente a uno specifico evento audio. Test: dato un dato senza etichetta e il file parametri della funzione di cui sopra, assegnargli un classe.

Metodologie di classificazione del suono Esistono diverse strategie di classificazione, secondo la specifica funzione parametrica utilizzato per raggruppare lo spazio delle funzioni. Analizzeremo tre metodi di classificazione: –K Nearest Neighbors (kNN): decide la classe di un punto sconosciuto nello spazio delle caratteristiche basato su la classe dei k Vicini più vicini –Gaussian Mixture Models (GMM): la funzione è una miscela di funzioni di densità di probabilità gaussiane (pdf). –Support Vector Machines (SVM): la funzione lo spazio è suddiviso da iperpiani

K Nearest Neighbors Consideriamo il training set formato da le coppie (f1, Y1),…, (fN, YN) dove fi è il punto nello spazio delle caratteristiche e Yi è il etichetta associata. Gli viene ora assegnato un punto x la cui etichetta è sconosciuto. • Training: fase di training dell'algoritmo consiste solo nella memorizzazione della caratteristica vettori ed etichette di classe della formazione campioni. • Nella fase di classificazione, k è definito dall'utente costante e il vettore x è classificato assegnando l'etichetta che è più frequente tra gli allenamenti k campioni più vicini a x.

• La scelta della funzione distanza è fondamentale per il classificazione

• La distanza euclidea può essere utilizzata per le caratteristiche continue. Altre scelte:

• Distanza da Manhattan:

• Distanza Minkowski:

• Quando sono in uso funzioni non continue, Hamming la distanza potrebbe essere utilizzata.

Esempi: • La distanza tra 1011101 e 1001001 è 2 • La distanza tra 2173896 e 2233796 è 3

• Un problema sorge quando la classe la distribuzione nel set di allenamento non è uniforme. In questo caso, la classe più rappresentata sarà il più selezionato. • Per superare questo problema, le distanze può essere ponderato proporzionalmente a l'inverso del numero di elementi in quella classe. • Un'altra soluzione è partizionare il file spazio nella formazione fase. Basato sulla formazione distribuzione, ogni punto nello spazio è assegnato una classe di appartenenza. Il La figura mostra un esempio per tre classi (rosso, verde e blu). La fase del test è quindi realizzato trovando il file regione in cui cade il punto

Classificazione bayesiana • Classificazione bayesiana e processo decisionale si basa sulla teoria della probabilità e il principio di scelta del più probabile o l'opzione con il rischio più basso (costo previsto). • Idea: concepire l'attività di classificazione come probabilistico o, in altre parole, trova la classe con la più alta probabilità di essere il generatore del vettore di caratteristiche osservate, quindi che il rischio di errore sia ridotto al minimo.

Esigenze • Considera di tenere le classi ω1,…, ωK • Miriamo a trovare la classe più adatta probabilmente ha generato il vettore di feature x sotto osservazione. • Per fare questo dobbiamo dedurre la probabilità funzione di densità p (ωk | x), k = 1,… K. • Questi PDF possono essere conosciuti a priori o stimato dal set di formazione. • Una volta che i PDF sono stati stimati, il file la classificazione trova la classe k che massimizza p (ωk | x)

Classificazione bayesiana Problema: non è noto a priori e non può essere stimato direttamente. Soluzione alternativa: teorema di Bayes

–P (x) è la probabilità totale di x

–P (ωk) è la probabilità a priori della classe k-esima

Bisogni e requisiti: stima degli a-priori probabilità Nota: la stima di p (x) non è necessaria, come

I requisiti che si presentano sono quindi: 1. Stimare le probabilità a priori In molte applicazioni la probabilità a priori della classe Non è disponibile. Soluzioni tipiche: –Assumere una distribuzione uniforme –Stimazioni basate su corpora (importante per applicazioni come il riconoscimento vocale)

Modelli di miscele gaussiane

2. Stimare p (x | ωk) L'approccio GMM presuppone che la densità di un il processo osservato può essere modellato come una somma ponderata delle densità dei componenti date da:

• vettore casuale x d-dimensionale

• bm (x) densità gaussiana, parametrizzato dalla sua media μm e la matrice di covarianza Σm

• Numero M di componenti

• peso in cm per la gaussiana m-esima componente Nota: il simbolo λ nell'equazione precedente non deve essere confuso con le classi. Qui λ denota l'insieme di addestramento

: matrice di covarianza della classe pdf

: vettore medio della classe pdf

Il modello di cui sopra può ospitare funzionalità di valore complesso utilizzando il coniugare i valori invece della trasposizione

Formazione di GMM La formazione di GMM è svolta da trovare i parametri che massimizzano il probabilità del set di dati. Un algoritmo utile in questo senso è il file Massimizzazione delle aspettative (EM).

Massimizzazione delle aspettative (EM) EM è un metodo iterativo che si alterna tra l'esecuzione –Un passo di aspettativa (E), che calcola un aspettativa della probabilità della caratteristica vettori xi rispetto alla stima corrente del PDF delle variabili latenti –E un passo di massimizzazione (M), che calcola i parametri che massimizzano il log atteso probabilità trovata nel passaggio E. VD ESEMPIO GRAFICO

+ VD SLIDE 77

Discussione EM si ferma quando viene raggiunto un numero fisso di passi o quando alcuni criteri di convergenza sono soddisfatti. • criterio di verosimiglianza logaritmica: termina quando la verosimiglianza logaritmica non cambia da un'iterazione alla seguente

• EM utilizza un numero fisso di classi • Altre tecniche (Figeuiredo-Jain, vedi [Paalanen 2004]) stimare anche il numero ideale di classi da utilizzare. • Viene tipicamente utilizzata la verosimiglianza (logaritmo del pdf) invece di pdf • Testare i GMM: applicazione del teorema di Bayes • Per evitare il problema del mal condizionamento del le caratteristiche delle matrici di covarianza sono tipicamente normalizzate tra un valore minimo e uno massimo.

Support Vector Machines (SVM) SVM è un classificatore binario che apprende il confine tra articoli appartenenti a due classi differenti. VD. ESEMPIO

Definizioni: –Feature vettori che definiscono il confine tra classi: vettori di supporto –Metà distanza tra i confini: margine - SVM (non) separabili: set di dati per i quali a iperpiano che divide tutti i vettori delle caratteristiche classi (impossibile) possono essere trovate - SVM (non) lineare: il confine che divide le due classi sono (non) un iperpiano ma un funzione più complessa

SVM lineare

Hyperplane:

–W: direzione

–W0: posizione

–Distanza di un punto x da un iperpiano:

Normalizzazione: ridimensionare w in modo che i punti appartenenti alle classi ω1 e ω2 più vicine alla hyperplane sono caratterizzati da g (x) = + / - 1.

Formazione di SVM: trova l'iperpiano per quale

e il margine è massimizzato. Nota: il margine è

L'attività di cui sopra può essere riformulata come file processo di minimizzazione vincolato: abbiamo per calcolare i parametri w, w0 in modo che

–J (w) = 1/2 || w || 2 è ridotto al minimo

-soggetto a

dove yi è un indicatore della classe (+1 per ω1 e -1 per ω2)

Soluzione: l'iperpiano ottimale è un lineare combinazione di alcuni dei vettori di caratteristiche:

dove pesa i punti nel set di dati in la definizione di iperpiano. Punti su il confine ha un valore diverso da zero, i punti non sul confine hanno a peso zero.

Dato un generico punto x, vogliamo inferire il classe a cui appartiene. VD. FUNZIONE DI DECISIONE

SVM non lineare Problema: un iperpiano che separa il file Impossibile trovare due classi. Strategia: operare una trasformazione su spazio delle caratteristiche, che trasforma il non lineare confine in uno lineare

Esempio: SLIDE 87• In questo caso alcuni punti del primo class rientra nell'area decisionale del seconda classe e viceversa. • Per includere questo caso, una penalità viene introdotto il fattore, che crea un file margine morbido tra le due classi

SVM di classe N SVM è un classificatore binario in natura. Alcune estensioni per un numero maggiore di esistono classi in letteratura.

Selezione delle caratteristiche Quando è presente un gran numero di funzioni, possono sorgere alcuni problemi: –Overfitting: molte funzionalità forniscono relative informazione –Costo computazionale: come il numero di caratteristiche aumenta, il costo computazionale aumenta pure Obiettivo: trova un sottoinsieme di funzionalità che possono garantire i migliori risultati di classificazione

Molte funzionalità sono già state estratte dai dati, in modo da essere utilizzato come input per un file sistema di classificazione. La selezione della funzione è posizionata tra estrazione e classificazione delle caratteristiche processi.

In letteratura (vedere [Dash97] per una completa sondaggio), sono stati due approcci principali proposto per la selezione delle caratteristiche: –Metodi filtro: l'algoritmo è indipendente da classificatore utilizzato e seleziona le caratteristiche sulla base di metriche di valutazione delle prestazioni calcolate direttamente dai dati –Metodi wrapper: l'algoritmo seleziona le caratteristiche la base di alcune metriche di rendimento ricevute come feedback dal classificatore I metodi wrapper offrono prestazioni superiori rispetto a metodi di filtraggio, ma al costo di un molto più alto carico computazionale. Nelle prossime diapositive considereremo i metodi di filtro.

Selezione delle caratteristiche scalari Valuta la separabilità delle classi di ciascuna funzionalità presi individualmente. Crea un vettore di feature in modo iterativo, incluso in ogni passaggio la caratteristica che massimizza la funzione

C è un criterio di separabilità di una classe (vedere più avanti), ρ è la correlazione tra la funzionalità testata e ciascuna delle caratteristiche già selezionate

Selezione delle caratteristiche vettoriali Valuta la separabilità di classe di un vettore di d Caratteristiche La complessità computazionale cresce in modo esponenziale quando la dimensione d aumenta Algoritmo di ricerca mobile: combina sequenziale avanti (aumenta il numero di funzioni utilizzate) e indietro (diminuisce il numero di funzioni utilizzate) passaggi di selezione.

Criteri di separabilità delle classi: misure scalari Viene presa in considerazione una singola caratteristica. L'obiettivo è per valutare il potere di discriminazione di questo caratteristica tra le classi Per semplicità, si supponga di avere solo due classi i e j Kullback J Divergence: d = Dij + Dji dove Dij e Dji sono le divergenze tra Kullback e Leibler le distribuzioni delle due classi

Criteri di separabilità delle classi: misure scalari Rapporto discriminante di Fisher:

dove μi, j sono i mezzi delle caratteristiche per le due classi e σ2 i, j sono le varianze. FDR vantaggi caratteristiche che massimizzano il distanza dei centroidi delle classi e ridurre al minimo la varianza della classe PDF.

Matrice di dispersione all'interno di una classe: informazioni sull'intradizione varianza

dove Si e Pi sono, rispettivamente, la matrice di covarianza e la probabilità a priori per la classe ωi:

Si = E [(x - μi) (x - μi) T]

Matrice di dispersione tra classi: informazioni sull'interclasse varianza dove μ0 è il vettore medio globale

Matrice di dispersione della miscela Sm = Sw + Sb Criteri di separabilità delle classi correlati: - assume valori elevati quando i campioni in lo spazio l-dimensionale sono ben raggruppati intorno alla loro media, all'interno di ogni classe e il cluster di classi diverse sono ben separati. - generalizza il Fisher- Rapporto discriminante rispetto al caso multidimensionale

Valutare le capacità discriminatorie di un file vettore di l caratteristiche Matrici a dispersione: fornire informazioni su la dispersione delle classi nel ldimensionale spazio, senza alcun presupposto (es. normalità) sulla distribuzione delle caratteristiche

Un suggerimento sui modelli Markov nascosti I modelli di Markov nascosti (HMM) sono abituati modellare i processi che variano nel tempo secondo uno schema specifico. HMM deriva dal noto Markov Catene.

Elementi di una catena di Markov: –Un insieme di stati Ns –Una matrice NsxNs che definisce la transizione probabilità da uno stato all'altro Esempio tipico: il processo di lancio della moneta

Dalle catene di Markov ai modelli di Markov nascosti Considera lo scenario in cui non possiamo vedere lo stato della sorgente, ma tutto quello che possiamo fare è avere un'osservazione di una quantità che è statisticamente correlato allo stato in cui sono.

Un HMM è definito come un insieme di stati Ns (nascosto, non visibile) Un insieme di probabilità di transizione di stato {aij} dove aij è il probabilità di transizione dallo stato Si allo stato Sj Un insieme di funzioni di densità di probabilità d-dimensionale {bj (x)} di la quantità x osservata per ogni stato Sj. –Gli HMM continui generalmente impostano bj (x) su un multivariato Distribuzione gaussiana con media μf e matrice di covarianza Σj, la funzione di densità di probabilità per lo stato j è quindi completamente caratterizzato da bj = {μj, j}. Un insieme di probabilità di stato iniziale {Πi} per ogni stato Si. –Definisce la probabilità dell'HMM di trovarsi allo stato Si al inizio del processo.

Decodifica di HMM Obiettivo: trova il percorso tra gli stati con il più alto probabilità

Se sono disponibili misurazioni M e N gli stati sono presenti nel modello, i percorsi NM lo sono possibile La decodifica di HMM viene eseguita utilizzando il noto algoritmo di Viterbi [Viterbi, 1967]

La formazione di HMM corrisponde a trovare il parametri del modello secondo il eventi acustici etichettati utilizzati per la formazione. Soluzione: l'algoritmo Baum-Welch [Baum 69], [Welch 03].

Conclusioni Molti componenti (e anche molte variabili) compaiono nella definizione di un sistema di classificazione. Al momento non esiste una ricetta specifica per il valutazione di quali elementi sono necessari per a corretta classificazione del problema in questione e questo è un problema di ricerca aperta. - Si può definire un insieme arbitrario di descrittori audio: come possiamo valutare qual è il set migliore? –Diverse tecniche di classificazione esistono in letteratura: qual è la migliore per il problema in mano?