

RICONOSCIMENTO ACUSTICO DEL FONDO STRADALE ATTRAVERSO ALGORITMI DI DEEP LEARNING

S. Squartini, L. Gabrielli, L. Ambrosini, F. Vesperini, D. Ferretti, D. Droghini

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche
Via Brecce Bianche 12, 60131, Ancona, Italia, e-mail s.squartini@univpm.it

Le emissioni sonore di un veicolo dipendono da molteplici fattori, tra i quali il rumore del motore e delle componenti meccaniche, il rumore aerodinamico e il rumore dovuto al rotolamento degli pneumatici sulla superficie stradale [1]. Il bilanciamento tra le componenti e il loro spettro sono dipendenti dalla velocità. In particolar modo, oltre i 30 km/h il rumore da rotolamento è quello prevalente rispetto agli altri [2]. I meccanismi di generazione di tale componente sono legati al contatto tra pneumatico e superficie stradale, che crea energia acustica caratterizzata in frequenza dalle geometrie tra pneumatico e asfalto e dalla composizione fine dell'asfalto [3].

Lo scopo di questo lavoro è sviluppare un sistema automatico per la classificazione della rugosità stradale a partire dal segnale audio prelevato da microfoni in prossimità dello pneumatico. La ruvidità della superficie stradale è caratterizzata dalla dimensione media della base di ghiaia e la presenza del materiale inerte di riempimento (cemento asfaltato o calcestruzzo). Questi elementi hanno un grande impatto sul livello e sullo spettro delle emissioni acustiche. Un lavoro recente [4], riporta l'utilizzo di Support Vector Machine con lo scopo di classificare diversi tipi di strada (asfalto, ghiaia, neve, strada sassosa) utilizzando feature acustiche. SVM sono impiegate anche in [5], dove l'obiettivo principale è la distinzione tra le condizioni di asfalto asciutto e asfalto bagnato. In [6] sono state utilizzate reti neurali ricorrenti LSTM abbinate alle Auditory Spectral Features (ASF) per lo stesso scopo. Questi lavori motivano ad approfondire ulteriori approcci deep learning per il problema del rilevamento della superficie stradale.

Uno degli algoritmi proposti di recente dagli autori consiste nell'utilizzo di reti neurali artificiali di tipo convoluzionale (CNN), che facciano uso di features ricavate dal segnale audio ricavato dai microfoni. Le features utilizzate in questa fase sono le Auditory Spectral Features (ASF) [6], ricavate dalla giustapposizione dei coefficienti Log-Mel, e dall'energia di una breve finestra dell'audio. Successivamente, i vettori delle features relativi ai frame adiacenti vengono uniti in un blocco con lo scopo di sfruttare la correlazione temporale dei dati. Quindi, la rete neurale viene addestrata su dati etichettati per stimare la classe a cui appartiene il frammento, nelle due classi liscio e ruvido.

Un passo necessario per questo lavoro è l'acquisizione di un dataset ampio e completo al fine di progettare e validare questo ed ulteriori algoritmi. L'auto scelta per la campagna di acquisizione dati è quella utilizzata in uno dei paper di riferimento [6]. L'auto è stata equipaggiata con 3 microfoni posti in posizioni fisse in prossimità degli pneumatici, 1 in prossimità del motore (per utilizzi futuri) e 2 microfoni posti all'interno dell'abitacolo. La raccolta dei dati è stata effettuata nelle aree urbane ed extra-urbane di Ancona includendo diverse velocità, condizioni del traffico e rugosità della pavimentazione. Le operazioni sono state svolte in giorni differenti. In primo luogo l'acquisizione è stata eseguita con il manto stradale completamente asciutto. Il dataset consiste in un totale di 52 minuti di registrazione, di cui rispettivamente 30 minuti acquisiti su manto stradale liscio e 22 minuti su asfalto ruvido. Si misura, perciò, uno sbilanciamento delle classi. Questo dataset viene perciò più avanti definito *unbalanced*.

Le prestazioni dell’algoritmo proposto per la classificazione della rugosità della superficie stradale sono riportate in Tabella 1 in termini di Accuracy, F-Measure, Recall e Precision. L’algoritmo è stato valutato con numerose configurazioni di rete e iperparametri dell’ottimizzatore (ad es., l’algoritmo, learning rate, batch size). Le reti neurali sono state addestrate usando il mean squared error (MSE) come funzione costo e Adam e Adadelta come algoritmi per la ricerca del minimo. Gli esperimenti sono stati condotti con il dataset bilanciato, nel quale la classe più numerosa è stata ridotta per avere ugual numero di samples per entrambe le classi. Come si evince dalla Tabella 1, il dataset sbilanciato comporta risultati migliori. Infine utilizzando quest’ultimo è stata effettuata un’analisi con cross-validation, con un risultato superiore ad entrambi i casi precedenti.

| | F-measure (%) | Accuracy (%) | Recall (%) | Precision (%) |
|--|---------------|--------------|------------|---------------|
| Best CNN trained w/ balanced dataset | 76.9 | 70.1 | 99.2 | 62.7 |
| Best CNN trained w/ unbalanced dataset | 78.3 | 77.8 | 94.4 | 66.9 |
| Best CNN on cross-validation w/ unbalanced dataset | 86.0 | 87.1 | 93.1 | 79.9 |

Tabella 1. Risultati ottenuti dalle migliori configurazioni CNN sul dataset liscio/ruvido.

Questi risultati sono stati presentati ad una conferenza della Audio Engineering Society [7].

In conclusione, l’algoritmo basato sull’architettura CNN ha fornito prestazioni in termini di F-Measure pari all’86.00% nella valutazione finale ottenuta con cross-validation. Ulteriori miglioramenti potrebbero essere ottenuti andando ad approfondire la fase sperimentale utilizzando diversi tipi di features acustiche, eventualmente più adatte a gestire eventi non stazionari. Si introdurranno altre reti neurali deep che tengano traccia anche dell’evoluzione temporale del segnale per mezzo di strutture recurrent come i layer di tipo LSTM che potrebbero sostituire quelli di tipo fully connected utilizzati attualmente. Si potrà anche sfruttare la diversity dei microfoni introducendo un’architettura multicanale che permetta di estrarre ulteriori informazioni acustiche utili dagli altri microfoni utilizzati durante la fase di acquisizione del dataset.

BIBLIOGRAFIA

- [1] D. I. Hanson, R. S. James, and C. NeSmith, “Tire/pavement noise study” the Center, 2004.
- [2] Sandberg, U., “Tyre/road noise: myths and realities,” in International Congress and Exhibition on Noise Control Engineering, The Hague, Netherlands, 2001.
- [3] U. Sandberg and J. A. Ejsmont, “Tyre/road noise reference book”, 2002.
- [4] Doğan, D., “Road-types classification using audio signal processing and SVM method,” in 25th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, 2017.
- [5] Alonso, J., Lopez, J., Pavón, I., Recuero, M., Asensio, C., Arcas, G., and Bravo, A., “On-board wet road surface identification using tyre/road noise and support vector machines,” in Applied Acoustics, 76, pp. 407–415, 2014.
- [6] Abdić, I., Fridman, L., Brown, D. E., Angell, W., Reimer, B., Marchi, E., and Schuller, B., “Detecting road surface wetness from audio: A deep learning approach,” in Pattern Recognition (ICPR), 2016 23rd International Conference on, pp. 3458–3463, IEEE, 2016.
- [7] L. Ambrosini, L. Gabrielli, F. Vesperini, S. Squartini, L. Cattani, “Deep Neural Networks for Road Surface Roughness Classification from Acoustic Signals”, 144th AES Convention, Milan, 2018.