Tesina 2

Analisi Particle Swarm Optimization come Feature Selection su Dataset DARWIN

Contesto del Problema

Un team di ricerca deve analizzare l'impatto dei diversi parametri dell'algoritmo Particle Swarm Optimization (PSO) sulla loro performance e convergenza utilizzando il dataset DARWIN. Questo dataset contiene caratteristiche di scrittura a mano per lo studio dell'Alzheimer, offrendo un contesto reale e significativo per l'analisi dell'algoritmo PSO.

Specifiche del Dataset

- DARWIN Dataset:
  - Features di scrittura a mano
  - Caratteristiche cinematiche
  - Pressione della penna
  - Parametri geometrici
  - Caratteristiche temporali
- Complessità:
  - Multiple feature categories
  - Dati numerici continui
  - Correlazioni complesse

Obiettivi

1. Implementare un **Algoritmo PSO** base per feature selection
2. Studiare sistematicamente l'effetto dei parametri:
   - Convergenza dell'algoritmo
   - Stabilità delle soluzioni

- Velocità di esecuzione
- Robustezza della selezione

3. Determinare configurazioni ottimali per:
    - Diverse dimensioni del subset di feature
    - Vincoli computazionali
    - Requisiti di stabilità

Vincoli

- Utilizzo stesso seed per confronti equi
- Minimo 30 run per configurazione
- Tempo massimo di esecuzione per run
- Gestione appropriata missing values

Fasi del Progetto

Fase 1: Implementazione PSO Base

- Sviluppare algoritmo PSO con:
    - Rappresentazione binaria per selezione feature
    - Funzione fitness basata su correlation analysis
    - Aggiornamento velocità e posizione
- Implementare logging dettagliato:
    - Fitness per iterazione
    - Feature selezionate
    - Tempi di esecuzione
    - Diversità dello sciame

Fase 2: Analisi Parametrica

1. **Scenario Dimensione Sciame**
    - Test dimensioni: [20, 50, 100, 200, 500]
    - Metriche:

        - Velocità convergenza
        - Stabilità selezione feature
        - Costo computazionale
    - Altri parametri fissi:
        - Inerzia (w): 0.7
        - c1 (cognitive): 2.0
        - c2 (social): 2.0
2. **Scenario Coefficienti PSO**
    - Inerzia (w): [0.4, 0.6, 0.7, 0.9]
    - Coefficiente cognitivo (c1): [1.0, 1.5, 2.0, 2.5]
    - Coefficiente sociale (c2): [1.0, 1.5, 2.0, 2.5]
    - Dimensione sciame fissa: 100
    - Analisi di:
        - Bilanciamento esplorazione/sfruttamento
        - Velocità di convergenza
        - Qualità delle soluzioni
3. **Scenario Criteri di Stop**
    - Numero iterazioni fisse: [50,100,200]
    - Convergenza (no improvement):
        - Soglie: [10,20,30] iterazioni
        - Tolleranze: [1e-4, 1e-5, 1e-6]
    - Analisi di:
        - Trade-off qualità/tempo
        - Stabilità delle soluzioni
        - Efficienza computazionale

Output Richiesti per Ogni Scenario

- Curve di convergenza

- Box plot distribuzioni fitness
- Frequenza selezione feature
- Tempi di esecuzione
- Analisi del comportamento dello sciame:
    - Dispersione delle particelle
    - Velocità media
    - Evoluzione del gbest