

Social-xLSTM: Integrating Social Pooling with Extended Long Short-Term Memory for Traffic Flow Prediction

Social-xLSTM Research Team

July 13, 2025

Abstract

本研究提出了 Social-xLSTM，一個結合 Social Pooling 機制與 Extended Long Short-Term Memory (xLSTM) 架構的創新交通流量預測模型。通過深入分析 Social LSTM 原理和 xLSTM 技術細節，我們設計了一個無需預定義拓撲結構的時空建模框架。該模型為每個車輛檢測器 (VD) 維持獨立的 xLSTM，並通過座標驅動的 Social Pooling 實現空間感知的隱藏狀態交互。實驗設計預期能夠改善傳統 LSTM 的過擬合問題，同時提供更好的長期時間依賴建模能力。

關鍵詞: 時空預測, Social LSTM, xLSTM, 交通流量預測, 深度學習

1 Introduction

交通流量預測作為智慧交通系統的核心技術，一直面臨著時間依賴性和空間相關性建模的雙重挑戰。傳統的 Long Short-Term Memory (LSTM) 網路在時間序列建模方面表現優異，但在記憶容量和長期依賴處理方面存在局限性。同時，多個檢測點之間的空間交互關係往往被忽略或建模不充分。

近年來，兩個重要的技術發展為解決這些問題提供了新的可能性：

1. **Social LSTM** [?]: 提出了通過 Social Pooling 機制實現多個序列間空間交互的方法，在人體軌跡預測中取得了顯著成效。
2. **xLSTM** [?]: 擴展了傳統 LSTM 架構，引入指數門控機制和矩陣記憶體，大幅提升了記憶容量和長期依賴建模能力。

本研究的核心貢獻在於：首次將 Social Pooling 機制與 xLSTM 架構相結合，設計了 Social-xLSTM 模型，為交通流量預測提供了一個創新且理論基礎紮實的解決方案。

2 Background and Related Work

2.1 Social LSTM: Architecture and Mathematical Formulation

Social LSTM [?] 的核心創新在於解決多個相關序列之間的依賴關係問題。傳統 LSTM 無法捕捉多個相關序列間的依賴關係，因此 Social LSTM 提出了 Social Pooling 機制來連接空間上鄰近的 LSTM。

2.1.1 核心架構原則

Social LSTM 的正確理解基於以下關鍵點：

1. 個別 **LSTM** 設計: 每個個體（行人或 VD）擁有自己的 LSTM 網路
2. 權重共享: 所有 LSTM 使用相同的架構和參數

3. 隱藏狀態交互: Social Pooling 作用於隱藏狀態, 不是原始特徵
4. 分散式預測: 每個個體為自己生成預測

2.1.2 Social Pooling 數學公式

對於第 i 個個體在時間 t , Social 隱藏狀態張量的構建公式為:

$$\mathbf{H}_t^i(m, n, :) = \sum_{j \in \mathcal{N}_i} \mathbf{1}_{mn}[x_j^t - x_i^t, y_j^t - y_i^t] \mathbf{h}_j^{t-1} \quad (1)$$

其中:

$$\mathbf{H}_t^i: \text{第 } i \text{ 個個體在時間 } t \text{ 的 Social 隱藏狀態張量} \quad (2)$$

$$\mathbf{h}_j^{t-1}: \text{鄰居 } j \text{ 在時間 } t-1 \text{ 的隱藏狀態} \quad (3)$$

$$\mathbf{1}_{mn}[x, y]: \text{指示函數, 檢查位置 } (x, y) \text{ 是否在網格 } (m, n) \text{ 中} \quad (4)$$

$$\mathcal{N}_i: \text{第 } i \text{ 個個體的鄰居集合} \quad (5)$$

完整的前向傳播流程如下:

Algorithm 1 Social LSTM Forward Pass

```

for 每個個體  $i$  do
   $\mathbf{r}_t^i = \phi(\mathbf{x}_t^i; \mathbf{W}_r)$  {位置嵌入}
end for
for 每個個體  $i$  do
  構建 Social 隱藏狀態張量  $\mathbf{H}_t^i$ 
   $\mathbf{e}_t^i = \phi(\mathbf{H}_t^i; \mathbf{W}_e)$  {嵌入 Social 張量}
   $\mathbf{h}_t^i = \text{LSTM}(\mathbf{h}_{t-1}^i, [\mathbf{r}_t^i, \mathbf{e}_t^i]; \mathbf{W}_l)$ 
end for
for 每個個體  $i$  do
   $\text{prediction}_i = \text{predict}(\mathbf{h}_t^i)$  {個別預測}
end for

```

2.2 xLSTM: Extended Long Short-Term Memory

xLSTM [?] 通過兩個主要創新擴展了傳統 LSTM: 指數門控機制和修改的記憶體結構。xLSTM 包含兩種變體: sLSTM (scalar LSTM) 和 mLSTM (matrix LSTM)。

2.2.1 sLSTM: Scalar LSTM with Exponential Gating

sLSTM 引入了指數門控和適當的正規化技術:

門控計算:

$$\tilde{i}_t = \mathbf{W}_i \mathbf{e}_t + \mathbf{R}_i \mathbf{h}_{t-1} + \mathbf{b}_i \quad (6)$$

$$\tilde{f}_t = \mathbf{W}_f \mathbf{e}_t + \mathbf{R}_f \mathbf{h}_{t-1} + \mathbf{b}_f \quad (7)$$

$$\tilde{z}_t = \mathbf{W}_z \mathbf{e}_t + \mathbf{R}_z \mathbf{h}_{t-1} + \mathbf{b}_z \quad (8)$$

$$\tilde{o}_t = \mathbf{W}_o \mathbf{e}_t + \mathbf{R}_o \mathbf{h}_{t-1} + \mathbf{b}_o \quad (9)$$

指數門控:

$$i_t = \exp(\tilde{i}_t) \quad (10)$$

$$f_t = \sigma(\tilde{f}_t) \quad (11)$$

$$z_t = \tanh(\tilde{z}_t) \quad (12)$$

$$o_t = \sigma(\tilde{o}_t) \quad (13)$$

細胞狀態和隱藏狀態:

$$c_t = f_t \odot c_{t-1} + i_t \odot z_t \quad (14)$$

$$n_t = f_t \odot n_{t-1} + i_t \quad (15)$$

$$\tilde{h}_t = \frac{c_t}{n_t} \quad (16)$$

$$h_t = o_t \odot \tilde{h}_t \quad (17)$$

其中 n_t 是正規化狀態以防止溢出。

2.2.2 mLSTM: Matrix LSTM with Covariance Update

mLSTM 使用矩陣值記憶體以提高存儲容量:

查詢、鍵、值計算:

$$\mathbf{q}_t = \mathbf{W}_q \mathbf{e}_t \quad (18)$$

$$\mathbf{k}_t = \mathbf{W}_k \mathbf{e}_t \quad (19)$$

$$\mathbf{v}_t = \mathbf{W}_v \mathbf{e}_t \quad (20)$$

矩陣記憶體更新:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot (\mathbf{v}_t \mathbf{k}_t^T) \quad (21)$$

$$\mathbf{n}_t = \mathbf{f}_t \odot \mathbf{n}_{t-1} + \mathbf{i}_t \odot \mathbf{k}_t \quad (22)$$

其中 $\mathbf{C}_t \in \mathbb{R}^{d \times d}$ 是矩陣記憶體, $\mathbf{n}_t \in \mathbb{R}^d$ 是正規化向量。
記憶體檢索:

$$\mathbf{h}_t = \mathbf{o}_t \odot \frac{\mathbf{C}_t \mathbf{q}_t}{\max(\mathbf{n}_t^T \mathbf{q}_t, 1)} \quad (23)$$

2.2.3 混合 xLSTM 區塊堆疊

一個 Social-xLSTM 區塊結合了 sLSTM 和 mLSTM 組件:

$$\text{Block}(\mathbf{x}) = \text{LayerNorm}(\mathbf{x} + \text{mLSTM}(\text{sLSTM}(\mathbf{x}))) \quad (24)$$

完整模型包含 L 個堆疊區塊:

$$\mathbf{h}^{(0)} = \text{InputEmbedding}(\mathbf{e}_t) \quad (25)$$

$$\mathbf{h}^{(\ell)} = \text{Block}_\ell(\mathbf{h}^{(\ell-1)}), \quad \ell = 1, \dots, L \quad (26)$$

$$\mathbf{y}_t = \text{OutputLayer}(\mathbf{h}^{(L)}) \quad (27)$$

3 Social-xLSTM: Methodology

3.1 Problem Formulation

考慮一個包含 N 個車輛檢測器 (VDs) 的交通網路，每個 VD i 位於固定座標 (x_i, y_i) 。在時間步 t ，每個 VD 的輸入特徵向量定義為：

$$\mathbf{x}_i^t = [q_i^t, v_i^t, o_i^t]^T \quad (28)$$

其中：

$$q_i^t = \text{VD } i \text{ 在時間 } t \text{ 的流量} \quad (29)$$

$$v_i^t = \text{VD } i \text{ 在時間 } t \text{ 的速度} \quad (30)$$

$$o_i^t = \text{VD } i \text{ 在時間 } t \text{ 的佔有率} \quad (31)$$

我們的目標是預測每個 VD 在下一個時間步 $t + 1$ 的交通狀態：

$$\hat{\mathbf{y}}_i^{t+1} = [\hat{q}_i^{t+1}, \hat{v}_i^{t+1}, \hat{o}_i^{t+1}]^T \quad (32)$$

3.2 Social-xLSTM Architecture

Social-xLSTM 架構包含四個主要組件：

1. **Individual xLSTM Processor:** 每個 VD 的獨立 xLSTM 處理
2. **Social Pooling Layer:** 座標驅動的空間特徵聚合
3. **State Fusion Layer:** 個體與社交特徵融合
4. **Output Layer:** 每個 VD 的個別預測

3.2.1 Individual xLSTM Processing

每個 VD i 使用共享的 xLSTM 架構進行獨立處理：

$$\mathbf{h}_i^{(\text{ind})} = \text{xLSTM}_{\text{shared}}(\mathbf{x}_i^{1:t}) \quad (33)$$

其中 $\text{xLSTM}_{\text{shared}}$ 是所有 VD 共享的 xLSTM 模型，確保參數效率和知識共享。

3.2.2 Coordinate-Driven Social Pooling

鄰居發現：對於目標 VD i ，我們基於歐幾里得距離找到其空間鄰居：

$$\mathcal{N}_i = \{j : d(\mathbf{c}_i, \mathbf{c}_j) \leq R, j \neq i\} \quad (34)$$

其中 $\mathbf{c}_i = (x_i, y_i)$ 是 VD i 的座標， R 是空間半徑參數。

空間網格構建：我們構建一個 $M \times N$ 的空間網格，中心位於目標 VD：

$$\mathbf{S}_i^t(m, n, :) = \sum_{j \in \mathcal{N}_i} w_{ij} \cdot \mathbf{1}_{mn}[\Delta x_{ij}, \Delta y_{ij}] \cdot \mathbf{h}_j^{(\text{ind})} \quad (35)$$

其中：

$$\Delta x_{ij} = x_j - x_i, \quad \Delta y_{ij} = y_j - y_i \quad (36)$$

$$w_{ij} = \exp\left(-\frac{d_{ij}}{R/3}\right) \quad (\text{距離權重}) \quad (37)$$

$$\mathbf{1}_{mn}[\cdot] : \text{網格指示函數} \quad (38)$$

網格坐標計算為:

$$m = \left\lfloor \frac{\Delta x_{ij} + R_x}{2R_x/M} \right\rfloor \quad (39)$$

$$n = \left\lfloor \frac{\Delta y_{ij} + R_y}{2R_y/N} \right\rfloor \quad (40)$$

空間特徵投影:

$$\mathbf{s}_i^t = \text{Proj}(\text{flatten}(\mathbf{S}_i^t)) \quad (41)$$

其中投影函數定義為:

$$\text{Proj}(\mathbf{x}) = \text{ReLU}(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) \quad (42)$$

3.2.3 State Fusion

我們使用多層融合策略結合個體和社交特徵:

$$\mathbf{h}_i^{(\text{fused})} = \text{LayerNorm}(\mathbf{W}_f[\mathbf{h}_i^{(\text{ind})}; \mathbf{s}_i^t] + \mathbf{b}_f) \quad (43)$$

3.2.4 Individual Prediction

每個 VD 使用其獨立的輸出層生成預測:

$$\hat{\mathbf{y}}_i^{t+1} = \mathbf{W}_{\text{out}}^{(i)} \mathbf{h}_i^{(\text{fused})} + \mathbf{b}_{\text{out}}^{(i)} \quad (44)$$

3.3 Complete Algorithm

Algorithm 2 Social-xLSTM Forward Pass

Input: VD features $\{\mathbf{x}_i^{1:t}\}_{i=1}^N$, coordinates $\{(x_i, y_i)\}_{i=1}^N$

Output: Predictions $\{\hat{\mathbf{y}}_i^{t+1}\}_{i=1}^N$

for each VD $i = 1, \dots, N$ **do**

 // Individual xLSTM Processing

$\mathbf{h}_i^{(\text{ind})} = \text{xLSTM}_{\text{shared}}(\mathbf{x}_i^{1:t})$

end for

for each VD $i = 1, \dots, N$ **do**

 // Social Pooling

 Find spatial neighbors $\mathcal{N}_i = \{j : d(\mathbf{c}_i, \mathbf{c}_j) \leq R\}$

 Construct spatial grid \mathbf{S}_i^t of size $M \times N$

for each neighbor $j \in \mathcal{N}_i$ **do**

 Compute relative position $(\Delta x_{ij}, \Delta y_{ij})$

 Assign $w_{ij} \cdot \mathbf{h}_j^{(\text{ind})}$ to grid cell (m, n)

end for

$\mathbf{s}_i^t = \text{Proj}(\text{flatten}(\mathbf{S}_i^t))$

end for

for each VD $i = 1, \dots, N$ **do**

 // State Fusion and Prediction

$\mathbf{h}_i^{(\text{fused})} = \text{LayerNorm}(\mathbf{W}_f[\mathbf{h}_i^{(\text{ind})}; \mathbf{s}_i^t] + \mathbf{b}_f)$

$\hat{\mathbf{y}}_i^{t+1} = \mathbf{W}_{\text{out}}^{(i)} \mathbf{h}_i^{(\text{fused})} + \mathbf{b}_{\text{out}}^{(i)}$

end for

4 Loss Function and Training

4.1 Multi-Objective Loss Function

我們使用多目標損失函數結合不同的預測誤差度量：

$$\mathcal{L} = \alpha \mathcal{L}_{\text{MAE}} + \beta \mathcal{L}_{\text{MSE}} + \gamma \mathcal{L}_{\text{MAPE}} \quad (45)$$

其中：

$$\mathcal{L}_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |\hat{\mathbf{y}}_i - \mathbf{y}_i|_1 \quad (46)$$

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2^2 \quad (47)$$

$$\mathcal{L}_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{\mathbf{y}}_i - \mathbf{y}_i}{\mathbf{y}_i} \right|_1 \quad (48)$$

推薦的權重設置為 $(\alpha, \beta, \gamma) = (0.4, 0.4, 0.2)$ 。

4.2 Training Strategy

訓練過程採用端到端的方式，所有 VD 的損失同時優化：

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^N \mathcal{L}_i \quad (49)$$

其中 \mathcal{L}_i 是 VD i 的個別損失。

5 Computational Complexity

5.1 Social Pooling Complexity

對於每個 VD i 有 $|\mathcal{N}_i|$ 個鄰居：

$$\mathcal{O}(\text{Social Pooling}) = \mathcal{O}(|\mathcal{N}_i| \cdot d_h + M \cdot N \cdot d_h) \quad (50)$$

5.2 xLSTM Complexity

對於單個 xLSTM 區塊：

$$\mathcal{O}(\text{sLSTM}) = \mathcal{O}(d_h^2) \quad (51)$$

$$\mathcal{O}(\text{mLSTM}) = \mathcal{O}(d_h^3) \quad (52)$$

5.3 Total Model Complexity

對於 N 個 VD， L 層， T 個時間步：

$$\mathcal{O}(\text{Social-xLSTM}) = \mathcal{O}(N \cdot T \cdot L \cdot (|\bar{\mathcal{N}}| \cdot d_h + d_h^3)) \quad (53)$$

其中 $|\bar{\mathcal{N}}|$ 是每個 VD 的平均鄰居數。

特性	Traditional LSTM	Social-xLSTM
記憶機制	標量記憶體	標量 + 矩陣記憶體
門控機制	Sigmoid	指數門控
空間建模	無	Social Pooling
長期依賴	有限	增強
過擬合風險	較高	預期較低
拓撲依賴	N/A	無需預定義

Table 1: Social-xLSTM 與傳統 LSTM 的比較

6 Theoretical Advantages

6.1 Compared to Traditional LSTM

6.2 Key Innovation Points

1. 首次將 **xLSTM** 應用於交通預測: 探索最新的記憶網路技術在時空預測中的潛力
2. 創新的時空混合建模: sLSTM 專門處理時間序列特徵, mLSTM 處理空間關係
3. 座標驅動的 **Social Pooling**: 完全基於地理座標的空間建模, 無需預定義拓撲
4. 個體化與社交化的平衡: 每個 VD 保持獨立性的同時實現空間感知
5. 無拓撲依賴的可擴展架構: 新增 VD 時無需重新設計網路結構

7 Expected Improvements

基於理論分析, Social-xLSTM 預期在以下方面取得改善:

7.1 Over-fitting Mitigation

傳統 LSTM 基準顯示嚴重過擬合 (驗證 $R^2 = -6$), xLSTM 的指數門控和更大記憶容量預期能:

- 改善訓練/驗證指標差距至 < 2 倍
- 實現正的驗證 R^2 值
- 更好的泛化能力

7.2 Long-term Dependencies

xLSTM 的矩陣記憶體和指數門控機制預期提供:

- 更好的長期時間序列建模
- 增強的記憶容量
- 更穩定的梯度傳播

7.3 Spatial Modeling

Social Pooling 機制預期實現:

- 自動發現空間關係
- 無需預定義拓撲的建模能力
- 可解釋的空間交互學習

8 Implementation Considerations

8.1 Hyperparameters

參數	符號	推薦值
網格大小	$M \times N$	8×8
空間半徑	R	25000m
隱藏維度	d_h	128
xLSTM 區塊數	L	6
sLSTM 位置	-	[1, 3]
學習率	α	0.001
損失權重	(α, β, γ)	(0.4, 0.4, 0.2)

Table 2: Social-xLSTM 超參數

8.2 Technical Dependencies

- **xlstm** 庫: 提供 sLSTM 和 mLSTM 實現
- **PyTorch**: 深度學習框架
- **CUDA**: GPU 加速支援 (推薦)

9 Conclusions and Future Work

9.1 Contributions

本研究提出的 Social-xLSTM 架構具有以下主要貢獻:

1. 理論創新: 首次結合 Social Pooling 與 xLSTM, 為交通預測提供新的理論框架
2. 技術進步: 設計了無需預定義拓撲的時空建模方法
3. 架構優化: 通過混合 sLSTM/mLSTM 實現時空特徵的專門化處理
4. 實用價值: 提供了可直接應用於實際交通系統的解決方案

9.2 Future Research Directions

1. 性能優化: 探索 TFLA 內核等優化技術以提升計算效率
2. 多尺度建模: 研究不同時空尺度下的模型適應性
3. 實時應用: 開發適用於實時交通預測的輕量化版本
4. 遷移學習: 探索跨區域、跨城市的模型遷移能力
5. 因果推理: 整合因果推理機制以提升模型可解釋性

9.3 Expected Impact

Social-xLSTM 預期在以下領域產生重要影響:

- 學術研究: 為時空預測領域提供新的研究範式
- 工業應用: 改善智慧交通系統的預測準確性
- 技術發展: 推動 xLSTM 在時空建模中的應用發展

Acknowledgments

感謝所有為 Social-xLSTM 研究做出貢獻的團隊成員，以及提供技術支援的相關開源社群。

References

- [1] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 961-971).
- [2] Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., ... & Hochreiter, S. (2024). xLSTM: Extended Long Short-Term Memory. In *Thirty-eighth Conference on Neural Information Processing Systems*.
- [3] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.