Bento: A Toolkit for Subcellular Analysis of Spatial Transcriptomics Data

Enhancing Spatial Transcriptomics Analysis Using BERT Transformer Models

Anjali Jangir

Introduction to Improvements Using BERT Transformers

Spatial transcriptomics provides spatial context to gene expression within individual cells, but traditional methods often struggle to capture the complex spatial and molecular relationships in this data. To address this, we propose using a BERT-based transformer model, adapted from natural language processing, to enhance RNA localization prediction, colocalization analysis, and subcellular domain segmentation on the MERFISH dataset.

Transformers like BERT are well-suited for this analysis due to their attention mechanisms, which model dependencies across large input spaces. By treating each RNA molecule as a token and incorporating spatial coordinates and gene expression levels, BERT can dynamically learn which RNA molecules are contextually related based on spatial and molecular features. This approach yields several improvements over traditional models:

- **Enhanced RNA Localization**: Prediction: with context-aware attention, enabling more accurate subcellular classification[1].

- **Improved Colocalization Analysis**: through attention-based metrics that consider both spatial and molecular proximity[2].

- **Refined Subcellular Domain Segmentation**: via unsupervised learning of spatial features, revealing transcriptionally distinct regions.

These improvements lead to better interpretability, generalization across spatial transcriptomic datasets, and deeper insights into cellular organization and function.

Mathematical Derivations for BERT-Based Model in Spatial Transcriptomics

1. Self-Attention Mechanism

In the BERT model, each input X representing an RNA molecule's spatial and gene expression information is transformed into three vectors: the **query** Q, **key** K, and **value** V. These are computed as follows:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

where W_Q, W_K, and W_V are learned weight matrices that allow the model to project the input X into different subspaces for calculating attention.

2. Scaled Dot-Product Attention

The attention score between each pair of RNA molecules i and j is calculated using scaled dot-product attention, which adjusts for the dimensionality of the vectors:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where d_k is the dimensionality of the key vectors, used to stabilize the gradient. The attention weights α_{ij} reflect the relevance of RNA molecule j to molecule i.

3. Spatially-Aware Attention Scores

To incorporate spatial information, we adjust the attention scores based on the spatial distance d_{ij} between molecules i and j. This results in spatially-aware attention weights α_{ij}:

$$\alpha_{ij} = \text{softmax}\left(\frac{(Q_i K_j^T)}{\sqrt{d_k}} - \lambda d_{ij}\right)$$

where λ is a hyperparameter that controls the influence of spatial distance in the attention calculation, balancing between spatial proximity and gene expression similarity [**?**].

4. Multi-Head Attention

To capture multiple aspects of spatial and molecular relationships, BERT uses multi-head attention. Each head h has separate learned parameters, and the final attention output is a concatenation of all heads:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O$$

where each $\text{head}_i = \text{Attention}(QW_Q^{(i)}, KW_K^{(i)}, VW_V^{(i)})$, and W_O is a learned projection matrix.

5. Positional Encoding for Spatial Data

To provide spatial context to each RNA molecule, we add positional encoding based on the 2D or 3D coordinates:

$$PE(x, y) = \left[\sin\left(\frac{x}{10000^{2i/d}}\right), \cos\left(\frac{x}{10000^{2i/d}}\right), \sin\left(\frac{y}{10000^{2i/d}}\right), \cos\left(\frac{y}{10000^{2i/d}}\right)\right]$$

These encodings are added to the input embeddings, enabling BERT to understand the spatial layout of the RNA molecules within the cell.

6. Loss Function for Localization Prediction

To train the model for RNA localization prediction, we use a cross-entropy loss $\mathcal{L}$ for supervised classification:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability for each localization category [**?**].

Expected Improvements and Results

By using a BERT-based model for spatial transcriptomics analysis, we expect the following improvements over traditional methods:

- **Improved RNA Localization Prediction**: Context-aware predictions from BERT's attention mechanism enable accurate classification of subcellular localizations, distinguishing subtle patterns such as nuclear vs. cytoplasmic localization.
- **Enhanced Colocalization Analysis**: BERT's attention weights can represent colocalization patterns by considering both spatial distance and gene expression similarity, resulting in more biologically relevant gene associations.
- **Refined Subcellular Domain Segmentation**: Using BERT's attention to capture spatial context allows for more accurate segmentation of subcellular domains, revealing distinct transcriptional regions within cells.
- **Interpretability and Insight**: Attention weights provide interpretability, allowing researchers to trace spatial dependencies and infer functional relationships between RNA molecules.
- **Cross-Platform Generalization**: BERT's flexible architecture enables the model to generalize across various spatial transcriptomic platforms, making it suitable for cross-comparative analyses.

In summary, the BERT-based approach enhances spatial transcriptomics by providing contextually informed, spatially aware predictions that improve both the accuracy and interpretability of RNA localization, colocalization, and subcellular segmentation tasks. This approach is expected to yield deeper insights into cellular organization and molecular interactions within cells.

References

[1]
[2]
[3]