

Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering

“ChebNet” NIPS’16 cited 2000+ by today

Overview

- 研究目标:

泛化/推广规则格网上的卷积神经网络(CNN)⇒非欧式域上的图卷积神经网络(graph convolution network)

涉及到的一对核心运算 - 卷积 (convolution): 特征提取; 池化 (pooling): 下采样 (进一步卷积发掘高阶特征)

- CNN性质: 提取局部静态特征; 平移不变性 (shift-/translation-invariant filter)
- 挑战: 如何推广到图卷积? 局部的图卷积核如何定义? 如何提升参数计算效率?
- 贡献: 以图谱理论为基础, 定义了图上严格局部化且低运算复杂度的卷积运算 (被称为“ChebNet”/被视为GCN研究的一座里程碑), 并提出了一种图上的池化策略。
- 代码:

Tensorflow (original): https://github.com/mdeff/cnn_graph

PyTorch: https://github.com/xbresson/spectral_graph_convnets

Research Background

图谱理论

- 图的定义:

给定一无向图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ ，其中 \mathcal{V} 是节点组成的有限集，基数 $|\mathcal{V}| = n$ ； \mathcal{E} 是边的集合； $W \in \mathbb{R}^{n \times n}$ 加权邻接矩阵，表示任意两节点之间边的权重；节点上的图信号 $x : \mathcal{V} \rightarrow \mathbb{R}$ 因而可以被视作一个向量 $x \in \mathbb{R}^n$ 。

- 在图谱理论中，图的拉普拉斯矩阵 L 被当做一重要的算子。其物理意义可被理解为图信号在图上的二阶导数(即梯度的散度)， $\Delta x = \nabla^2 x = \text{div}(\text{grad } x) = L \cdot x$ 。本文中用到的 L 的定义有两种, combinatorial Laplacian: $L = D - W \in \mathbb{R}^{n \times n}$ 和 Symmetric normalized Laplacian: $L = I_n - D^{-1/2} W D^{-1/2}$ ，其中 $D \in \mathbb{R}^{n \times n}$ 是 W 的对角化度矩阵 $D_{ii} = \sum_j W_{ij}$ ， I_n 是单位矩阵。
- 由于 L 是半正定对称矩阵，因而一定可进行特征分解 (\Leftrightarrow 存在 n 个线性无关的特征向量) $L = U \Lambda U^T$ ； $U = [u_0, \dots, u_{n-1}] \in \mathbb{R}^{n \times n}$ 是 L 的 n 个两两正交的特征向量组成的矩阵 (被称为基 Fourier basis/mode)，而 $\Lambda = \text{diag}([\lambda_0, \dots, \lambda_{n-1}]) \in \mathbb{R}^{n \times n}$ 是 L 的 n 个非负特征值构成的对角矩阵 (升序排列)，被称为该图的频率。

Research Background

从图上的离散傅立叶变换到GCN

- 在图的信号处理研究领域, D. Shuman et al. (The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains) 将传统傅立叶变换推广到离散图上。离散傅立叶变换因为可记为, 正变换: $\hat{x} = U^T x \in \mathbb{R}^n$; 逆变换: $x = U \hat{x} \in \mathbb{R}^n$ 。

- 由卷积定理, 两个图信号 x, y 在时/空域的卷积等于两者在傅立叶频域内的点积, 故常写做 $x *_G y = U((U^T x) \odot (U^T y))$ 其中 $*_G$ 表示在图上的卷积运算, \odot 表示矩阵的哈达马积。

- 若令上式中的 y 为卷积核 (spectral filter/kernel), 将其傅立叶变换写做对角矩阵形式, 则有

$$y = g_\theta(L)x = g_\theta(U \Lambda U^T)x = U g_\theta(\Lambda) U^T x$$

- J. Bruna et al. (Spectral Networks and Deep Locally Connected Networks on Graphs) 直接令

$$g_\theta(\Lambda) = \text{diag}(\theta) = \begin{pmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_n \end{pmatrix}$$

因而初代GCN可写做 $y_{\text{output}} = \sigma(U g_\theta(\Lambda) U^T x)$

Methodology

初代GCN的弊端及解决方案 1

- 卷积核不具有空域的局部连接性，且需要学习复杂度为 $\mathcal{O}(n)$ (因为需要学习的参数数量 $\theta \in \mathbb{R}^n$)
- 解决方案：使用多项式表示卷积核，即 $g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k$ ，其中 $\theta \in \mathbb{R}^K$ 为多项式系数。一方面，学习复杂度由 $\mathcal{O}(n)$ 降为 $\mathcal{O}(K)$ ；另一方面，使卷积核在空域具有局部连接性，即 $d_G(i, j) > K$ 时有 $(L^K)_{i,j} = 0$ (感受野为 K ，如下图所示)，证明由D. Hammond et al. (Wavelets on Graphs via Spectral Graph Theory)引理5.2给出。

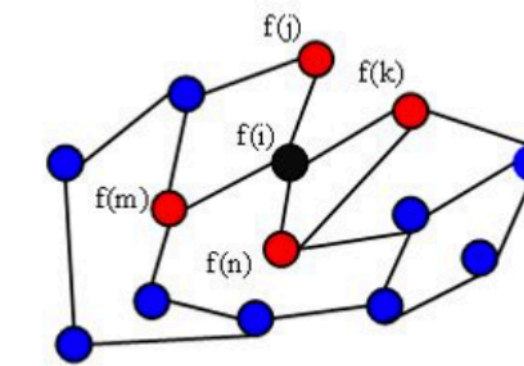
Lemma 5.2. Let G be a weighted graph, \mathcal{L} the graph Laplacian (normalized or non-normalized) and $s > 0$ an integer. For any two vertices m and n , if $d_G(m, n) > s$ then $(\mathcal{L}^s)_{m,n} = 0$.

Proof. First note that $\mathcal{L}_{i,j} = 0$ if i and j are distinct vertices that are not connected by a nonzero edge. By repeatedly expressing matrix multiplication with explicit sums, we have

$$(\mathcal{L}^s)_{m,n} = \sum \mathcal{L}_{m,k_1} \mathcal{L}_{k_1,k_2} \cdots \mathcal{L}_{k_{s-1},n} \quad (32)$$

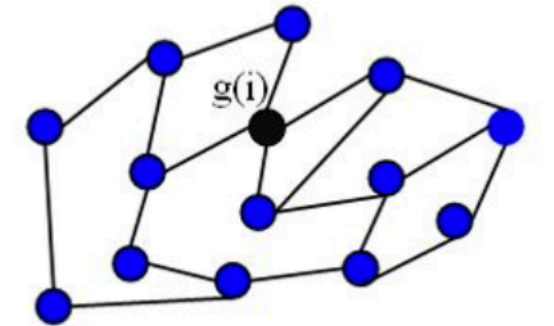
where the sum is taken over all $s-1$ length sequences k_1, k_2, \dots, k_{s-1} with $1 \leq k_r \leq N$. Assume for contradiction that $(\mathcal{L}^s)_{m,n} \neq 0$. This is only possible if at least one of the terms in the above sum is nonzero, i.e. there exist k_1, k_2, \dots, k_{s-1} such that $\mathcal{L}_{m,k_1} \neq 0, \mathcal{L}_{k_1,k_2} \neq 0, \dots, \mathcal{L}_{k_{s-1},n} \neq 0$. After removing possibly repeated values of the k_r 's, this implies the existence of a path of length less than or equal to s from m to n , so that $d(m, n) \leq s$, which contradicts the hypothesis. \square

$K = 1$

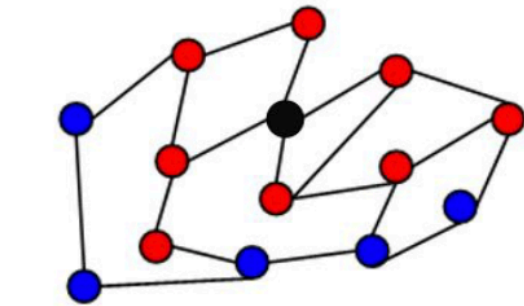


$$g(i) = \sum_{j \in N(1)} a_j f(j)$$

Graph Convolution

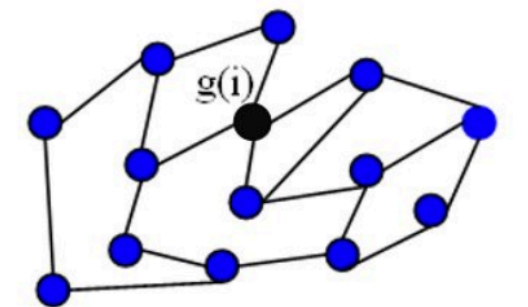


$K = 2$



$$g(i) = \sum_{j \in N(2)} a_j f(j)$$

Graph Convolution



Methodology

初代GCN的弊端及解决方案 2

- 前向传播计算矩阵乘积 $Ug_{\theta}(\Lambda)U^T$ 的时间复杂度为 $\mathcal{O}(n^2)$ (对于大规模的图而言代价过高)
- 解决方案：利用递归形式多项式表示卷积核以降低计算复杂度。在图信号处理领域GSP, D. Hammond et al. (Wavelets on Graphs via Spectral Graph Theory) 提出用第一类切比雪夫多项式近似图上小波变换中的kernel。受到启发，这里将卷积核进而写为 $g_{\theta}(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\Lambda})$ ，式中 k 阶切比雪夫多项式可由递归得到， $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ 且 $T_0 = 1, T_1 = x$ 。另外，由于卷积核由特征值对角矩阵 Λ 得出，进而 $U \sum_{j=0}^{K-1} \theta_j \Lambda^j U^T = \sum_{j=0}^{K-1} \theta_j U \Lambda^j U^T = \sum_{j=0}^{K-1} \theta_j L^j$ ，因此拉普拉斯矩阵不需要再做特征分解，可直接作为卷积核的输入，从而降低了运算复杂度 $\mathcal{O}(K|\mathcal{E}|) \ll \mathcal{O}(n^2)$ (对于稀疏拉普拉斯矩阵 L)。
- 注意第一类切比雪夫多项式本身的形式为 $T_k(x) = \cos(k \cdot \arccos(x))$ ，而反余弦函数的定义域是 $[-1, 1]$ 。所以作为输入，拉普拉斯矩阵 L 需先投影到 $[-1, 1]$ ，记作 $\tilde{L} = 2L/\lambda_{max} - I_n$ ，式中 λ_{max} 是 L 最大的特征值。
- 综上，诞生了新一代GCN - “ChebNet”，卷积层可以写做

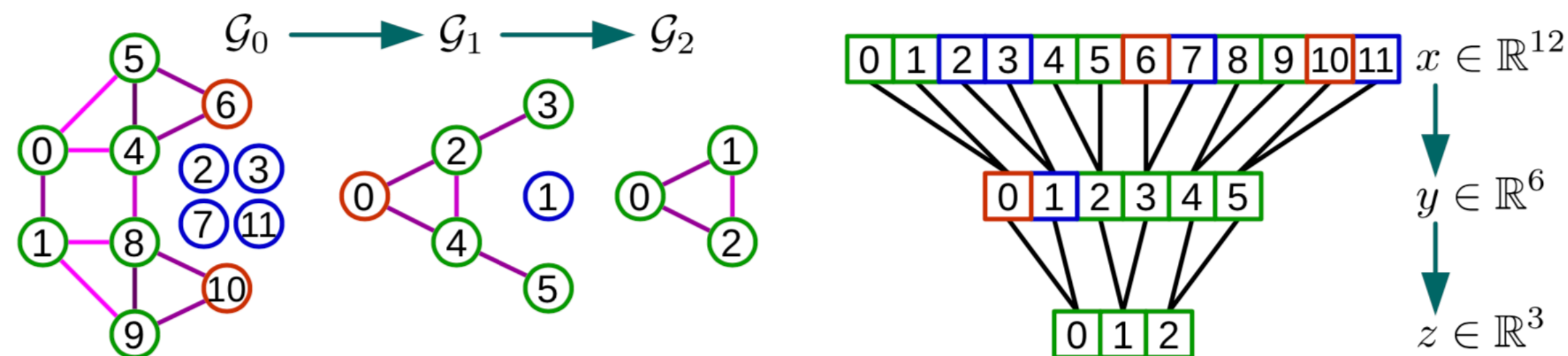
$$y = g_{\theta}(L)x = \sigma \left(\sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x \right)$$

Methodology

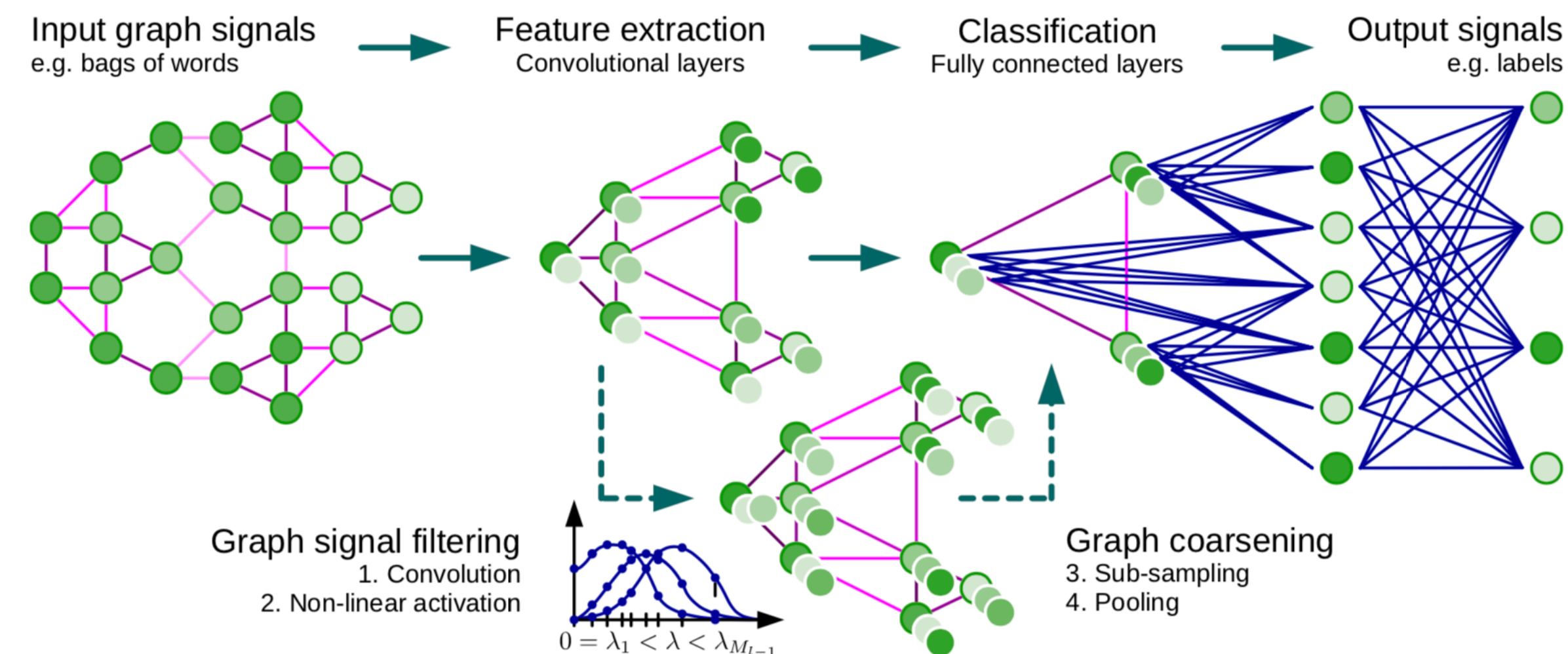
池化层 Pooling Layer

- 池化运算：等价于↔进行图的聚类 (graph clustering) 的同时保持图的局部几何结构
- 这里利用I. Dhillon et al. (Weighted Graph Cuts Without Eigenvectors: A Multilevel Approach) 提出的图的粗化算法 (Graclus multilevel clustering algorithm), 基于Metis。由于图的聚类本身属于NP-hard问题, 故此处用一贪心算法求解局部最优解: 通过连续多级粗化, 每一级粗化最大化任意两节点间的normalized cut $W_{ij}(1/d_i + 1/d_j)$ 至每对节点被遍历。
- Graclus算法视频讲解链接: <https://www.youtube.com/watch?v=kksLJ2D5wdU>
- 图信号的快速池化:

转化为1D Pooling: (1) 构建平衡二叉树 (Balanced Binary Tree); (2) 节点重排列



Experiments



vs. CNN on MNIST (图像分类)

Accuracy

Model	Architecture	Accuracy
Classical CNN	C32-P4-C64-P4-FC512	99.33
Proposed graph CNN	GC32-P4-GC64-P4-FC512	99.14

Speed

Model	Architecture	Time (ms)		
		CPU	GPU	Speedup
Classical CNN	C32-P4-C64-P4-FC512	210	31	6.77x
Proposed graph CNN	GC32-P4-GC64-P4-FC512	1600	200	8.00x

Filter
Variants

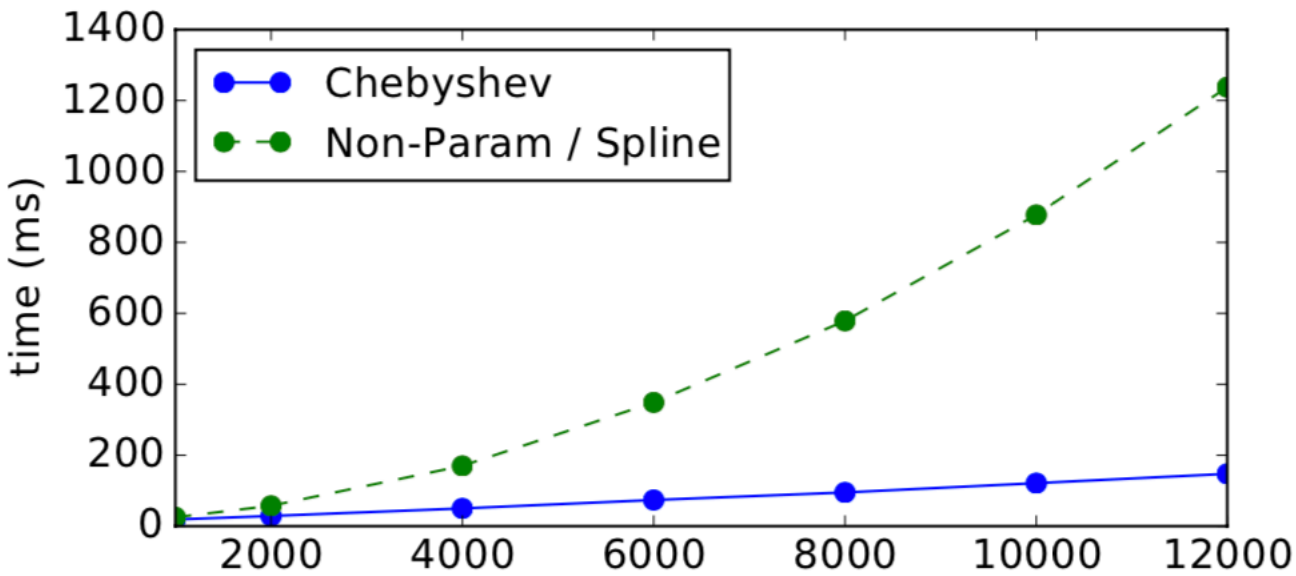
Dataset	Architecture	Accuracy		
		Non-Param (2)	Spline (7) [4]	Chebyshev (4)
MNIST	GC10	95.75	97.26	97.48
MNIST	GC32-P4-GC64-P4-FC512	96.28	97.15	99.14

vs. Baselines on 20NEWS (文本分类)

Model	Accuracy
Linear SVM	65.90
Multinomial Naive Bayes	68.51
Softmax	66.28
FC2500	64.64
FC2500-FC500	65.76
GC32	68.26

Accuracy

Efficiency



Thanks for your attention!

Research discussion now 