

Dynamic Graph Convolutional Recurrent Network for Traffic Prediction: Benchmark and Solution

作者: Fuxian Li, Jie Feng, Huan Yan, Yong Li 等

机构: 清华大学, 国防科技大学

发表期刊: IEEE Transactions on Knowledge and Data Engineering (TKDE)

Preliminary

本文旨在完成多变量时序预测中的交通预测任务。通常来说基于路网的交通预测将路网中记录交通数据（交通速度、交通流量等）的传感器建模为一个有向带权图 $g = (v, \epsilon, A)$. 其中 v 表示含有 N 个节点的集合 (本文中一个传感器代表一个节点), ϵ 是路网中节点连接关系的边集合, 并使用邻接矩阵 $A \in R^{N \times N}$ 来表示边集 ϵ .

因为路网中通常有多个传感器, 所以交通预测属于多变量时序预测。多变量时序预测与单变量时序预测的区别在于需要预测所有变量未来一段时间的状态。假设路网中某个节点的时序数据为 $X_i \in R^{T \times D}$. 其中 T 表示历史时间步, D 为特征的维度。那么路网中所有传感器记录的数据可以表示为 $X \in R^{T \times N \times D}$. 交通预测旨在通过已知路网和其观测到的 P 个历史时间步数据, 使用映射函数 f 来预测未来的 Q 个时间步的交通数据。如下所示:

$$[X_{(t-P):t}, g] \xrightarrow{f} \hat{X}_{t:(t+Q)} \tag{1}$$

Challenges

首先review了交通预测已有的大多数工作, 并做如下总结, 将已有的时空预测模型进行分类, 依据 Spatial Topology Construction, Spatial dependency, Temporal dependency, External features 四个维度将现有模型进行分类, 如图1所示:

TABLE 1: Model classification.

Model	Spatial topology construction	Spatial dependency	Temporal dependency	External features
DCRNN [5]	Distance-based graph	GCN	GRU	Time
AGC-Seq2Seq [20]	Binary graph	GCN	GRU + Attention	None
TGC-LSTM [21]	Binary graph + Others	GCN	LSTM	None
ST-MGCN [22]	Binary graph + Others	GCN	RNN	POI + Road network structure
ST-MetaNet [7]	Distance-based graph	Meta-GAT	Meta-GRU	Time + Road network structure
MRA-BGCN [8]	Distance-based graph + Others	GCN + Attention	GRU	Time
DGCN [23]	Binary graph	GCN + Attention	LSTM + Attention	Time
AGCRN [15]	Adaptive graph	GCN	GRU	Time
STGNN [13]	Adaptive graph + Distance-based graph	GCN	GRU + Transformer	Time
GMAN [9]	Distance-based graph	Graph embedding + Attention	Embedding + Attention	Time
ST-GRAT [10]	Distance-based graph	Graph embedding + Attention	Embedding + Attention	Time
STTNs [11]	Distance-based graph	Transformer	Transformer	Time
STGCN [6]	Distance-based graph	GCN	CNN	Time
ASTGCN [17]	Binary graph	GCN + Attention	CNN + Attention	Time
STSGCN [18]	Binary graph	GCN	GCN	Time
LSGCN [19]	Binary graph	GCN + Attention	CNN	Time
Graph WaveNet [12]	Adaptive graph + Distance-based graph	GCN	CNN	Time
SLCNN [16]	Adaptive graph	GCN	CNN	Time
MTGNN [14]	Adaptive graph	GCN	CNN	Time
ST-ResNet [1]	Grid-based graph	CNN	CNN	Time + Weather
DeepSTN+ [2]	Grid-based graph	CNN	CNN	Time + POI
DMVST-Net [3]	Grid-based graph + Others	CNN + Graph embedding	RNN	Time + Weather
STDN [4]	Grid-based graph	CNN	RNN + Attention	Time
DARNN [24]	None	None	RNN + Attention	None
GeoMAN [25]	None	Attention	RNN + Attention	Time + Weather + POI

可以发现已有的工作中大多使用GCN、CNN、self-attention建模复杂的空间路网依赖，使用RNN、CNN、self-attention建模复杂的非线性的时间依赖。

作者总结了现有工作的两个缺点：

1.对于空间依赖，CNN无法对非欧图数据进行建模，所以在基于路网的交通预测任务中大多不适用。GNN在基于路网的交通预测任务被广泛使用，但目前不管是基于路网还是语义的静态邻接矩阵都无法很好的表达动态变化的路网结构。self-attention虽然能动态建模每个时间的邻接关系，但其完全基于语义忽略了图的拓扑关系。

2.对于时间依赖，CNN可以用来有效地捕获时间相关性，但由于其隐式时间建模使得时间步长固定而牺牲了灵活性。self-attention虽然可以捕捉全局时间依赖关系，但短期预测效果不佳。RNN相比于self-attention有更好的短期预测表现但相对来说时间消耗更多。

本文Contributions

- 本文提出一种基于GCN和RNN的模型结构，首先利用传感器记录的实时速度数据 V_t 、时间戳信息 T_t 和历史隐藏状态 H_{t-1} ，生成动态的邻接矩阵，接着将生成的动态矩阵与原始的路网邻接矩阵合并。
- 设计了一种Dynamic Graph Convolutional Recurrent Module(DGCRM)，并以此为单元模块构造了一个RNN网络作为整体模型。
- 在三个交通数据集上进行实验(包括两个公开数据集和一个私有数据集)，将DGCRM与现有的方法进行比较。DGCRM获得了SOTA表现和较快的训练速度，体现了本文方法的优越性。

The Proposed Method

如图2所示，DGCRN基于seq2seq的架构并且将传统的GRU中的线性单元更换为动态图卷积命名为DGCRM以便捕获时空依赖性。同时可以注意到在RNN的每一步模型都会使用'Graph Generator'模块动态生成语义相关的邻接矩阵捕捉数据的空间变化性。

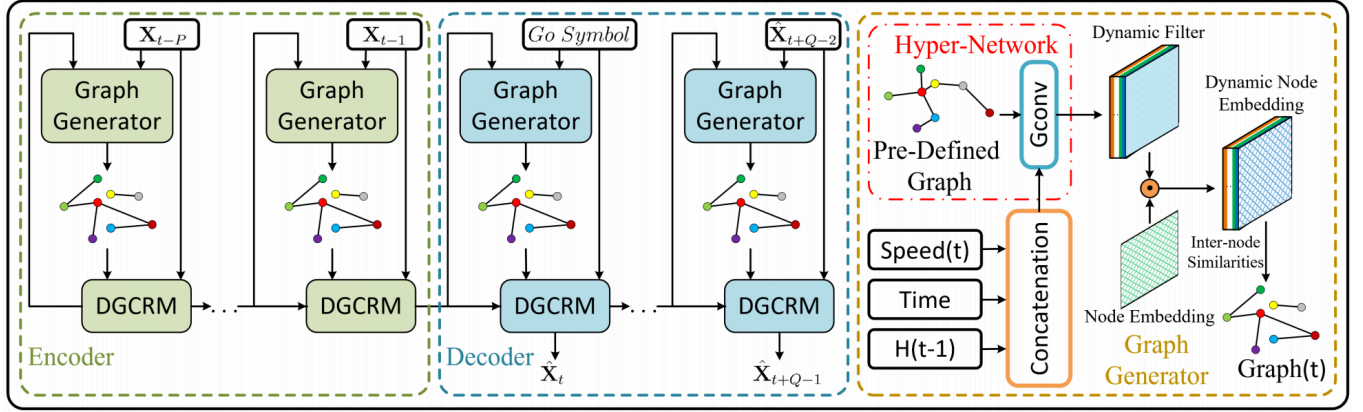


Fig. 1: The architecture of DGCRN.

Graph Generator

图生成模型的核心在于对每一个时间步 t ，使用现在以及历史信息生成一个动态的邻接矩阵 DA^t 。此文首先将当前 t 时刻速度 V_t 与时间戳信息 T_t 和历史隐藏状态 H_{t-1} 拼接到一起：

$$I_t = V_t || T_t || H_{t-1} \quad (2)$$

接着利用一个图卷积层处理拼接得到的张量 I_t ：

$$DF^t = \Theta_{*G}(I_t) \quad (3)$$

然后此文将富含空间特征的 DF^t 与可学习的参数 $E_1 \in R^{N \times D}$, $E_2 \in R^{N \times D}$ 做handmard-product以此表达源节点的嵌入向量以及目标节点的嵌入向量，如下所示：

$$\begin{aligned} DE_1^t &= \tanh(\alpha(DF_1^t \odot E_1^t)) \\ DE_2^t &= \tanh(\alpha(DF_2^t \odot E_2^t)) \end{aligned} \quad (4)$$

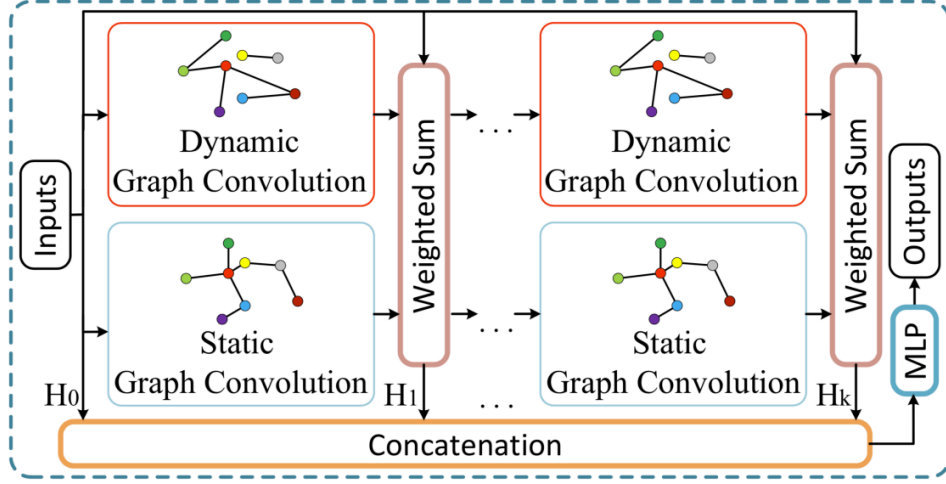
最后此文将源节点与目标节点的嵌入向量做点积得到最后的语义相似度矩阵也即语义的邻接矩阵：

$$DA^t = \text{ReLU}(\tanh(\alpha(DE_1^t DE_2^{tT} - DE_2^t DE_1^{tT}))) \quad (5)$$

其中 α 是为了控制边的数量的超参数， $DA^t \in R^{B \times N \times N}$ 是在 t 时刻基于语义相似度的动态邻接矩阵。

Dynamic Graph Convolutional Recurrent Module (DGCRM)

基于路网距离的静态邻接矩阵和基于节点语义相似度的动态邻接矩阵从不同方面反映了节点间的关系。此文使用 weighted average 来合并两个矩阵所进行图卷积后得到的张量从多方面捕捉空间依赖性。图卷积流程如图3所示：



图卷积的公式为：

$$H_{out} = \Theta_{*G}(H_{in}, DA^t, A) \quad (6)$$

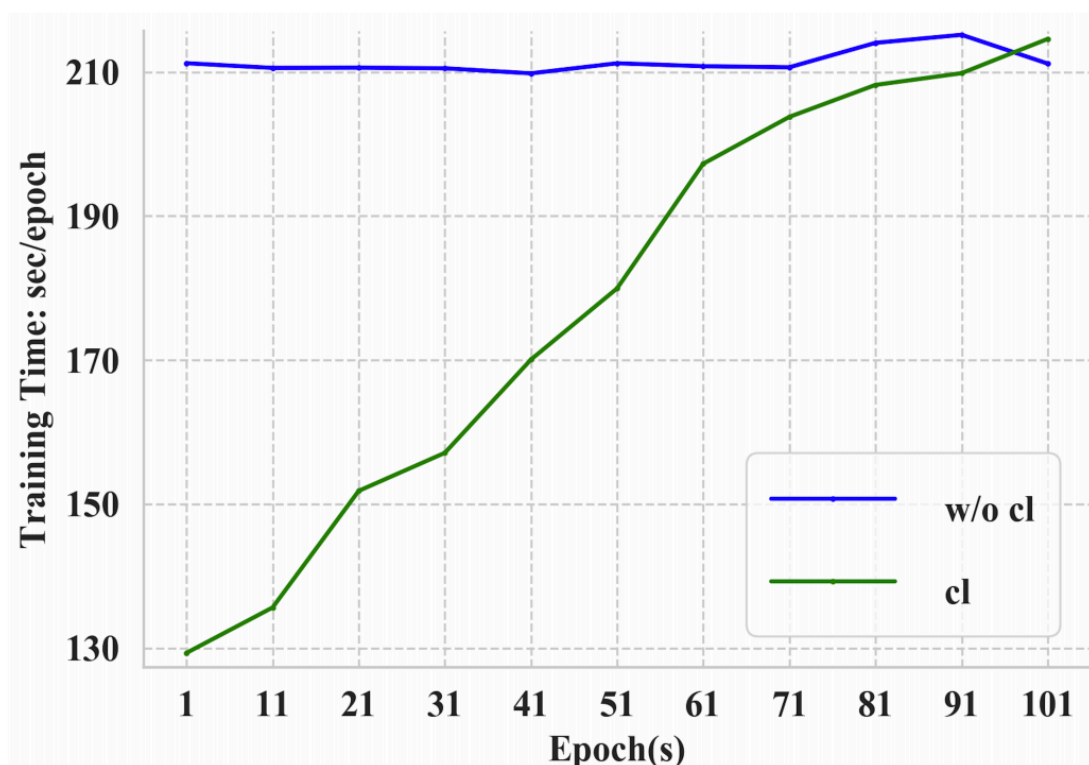
整体模型构成RNN框架

最后本文将GRU中的线性运算单元替换为以上提出的DGCRM模块，以提取复杂的时空依赖性：

$$\begin{aligned} z^{(t)} &= \sigma(\Theta_{z*G}(X_t || H_{t-1})) \\ r^{(t)} &= \sigma(\Theta_{r*G}(X_t || H_{t-1})) \\ h^{(\tilde{t})} &= \tanh(\Theta_{h*G}(X_t || (r^{(t)} \odot H_{t-1}))) \\ H_t &= z^{(t)} \odot H_{t-1} + (1 - z^{(t)}) \odot h^{(\tilde{t})} \end{aligned} \quad (7)$$

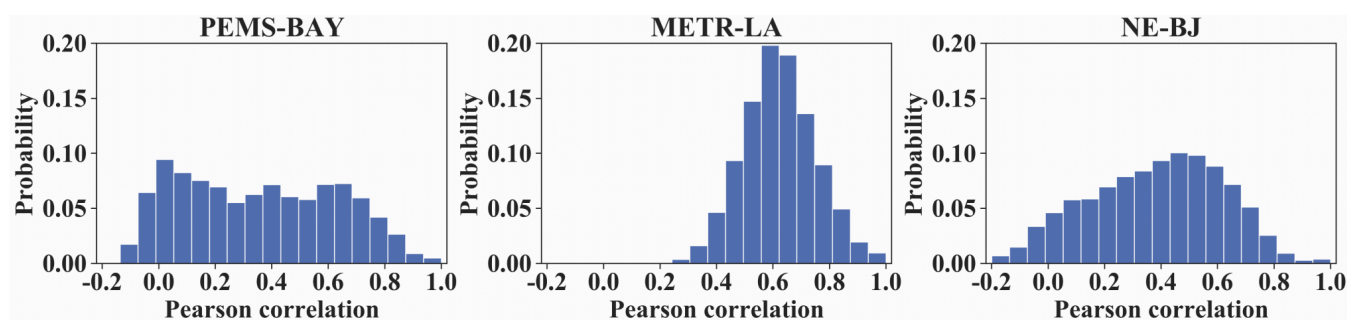
模型训练策略

本文使用了与MTGNN一致的训练策略，也即训练时从一开始并不是预测未来所有步数的信息。而是从未来的第一个时间片开始训练，训练收敛后在训练前两个，直到所有时间片都训练完成停止训练。这样的 curriculum learning 方法可以减少早期训练所需要的时间，使用了 curriculum learning 方法与不使用的时间比较如下图所示。同时本文使用了RNN中的 scheduled sampling 方法提升模型准确度。

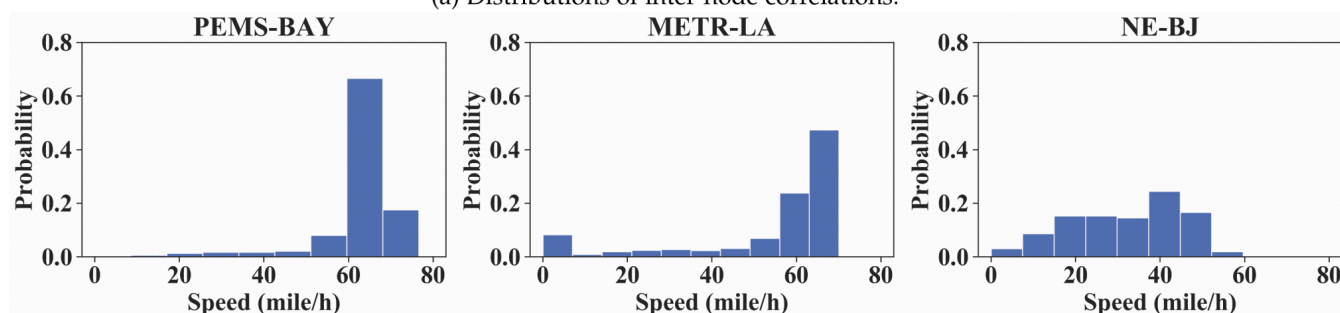


Experiments

作者首先对实验使用的三个交通数据集进行了统计学分析，如下图所示METR-LA和NE-BJ数据集中点对间的 Pearson correlation 比PEMS-BAY数据集更大，也即存在更多的空间依赖关系，其次PEMS-BAY存在更多的缺失数据。



(a) Distributions of inter-node correlations.



(b) Distributions of speed.

下图与baselines的比较结果显示DGCRN在三个数据集的大多数任务上均取得了最好的结果。

	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA									
HA	4.16	7.80	13.00%	4.16	7.80	13.00%	4.16	7.80	13.00%
ARIMA _{kal}	3.99	8.21	9.60%	5.15	10.45	12.70%	6.90	13.23	17.40%
VAR	4.42	7.89	10.20%	5.41	9.13	12.70%	6.52	10.11	15.80%
SVR	3.99	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%
FNN	3.99	7.94	9.90%	4.23	8.17	12.90%	4.49	8.69	14.00%
FC-LSTM	3.44	6.30	9.60%	3.77	7.23	10.90%	4.37	8.69	13.20%
DCRNN	2.77	5.38	7.30%	3.15	6.45	8.80%	3.60	7.60	10.50%
STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%
Graph WaveNet	2.69	5.15	6.90%	3.07	6.22	8.37%	3.53	7.37	10.01%
ST-MetaNet	2.69	5.17	6.91%	3.10	6.28	8.57%	3.59	7.52	10.63%
ASTGCN	4.86	9.27	9.21%	5.43	10.61	10.13%	6.51	12.52	11.64%
STSGCN	3.31	7.62	8.06%	4.13	9.77	10.29%	5.06	11.66	12.91%
AGCRN	2.87	5.58	7.70%	3.23	6.58	9.00%	3.62	7.51	10.38%
GMAN	2.80	5.55	7.41%	3.12	6.49	8.73%	3.44	7.35	10.07%
MTGNN	2.69	5.18	6.86%	3.05	6.17	8.19%	3.49	7.23	9.87%
DGCRN	2.62	5.01	6.63%	2.99	6.05	8.02%	3.44	7.19	9.73%
PEMS-BAY									
HA	2.88	5.59	6.80%	2.88	5.59	6.80%	2.88	5.59	6.80%
ARIMA _{kal}	1.62	3.30	3.50%	2.33	4.76	5.40%	3.38	6.50	8.30%
VAR	1.74	3.16	3.60%	2.32	4.25	5.00%	2.93	5.44	6.50%
SVR	1.85	3.59	3.80%	2.48	5.18	5.50%	3.28	7.08	8.00%
FNN	2.20	4.42	5.19%	2.30	4.63	5.43%	2.46	4.98	5.89%
FC-LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%
DCRNN	1.38	2.95	2.90%	1.74	3.97	3.90%	2.07	4.74	4.90%
STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%
Graph WaveNet	1.30	2.74	2.73%	1.63	3.70	3.67%	1.95	4.52	4.63%
ST-MetaNet	1.36	2.90	2.82%	1.76	4.02	4.00%	2.20	5.06	5.45%
ASTGCN	1.52	3.13	3.22%	2.01	4.27	4.48%	2.61	5.42	6.00%
STSGCN	1.44	3.01	3.04%	1.83	4.18	4.17%	2.26	5.21	5.40%
AGCRN	1.37	2.87	2.94%	1.69	3.85	3.87%	1.96	4.54	4.64%
GMAN	1.34	2.91	2.86%	1.63	3.76	3.68%	1.86	4.32	4.37%
MTGNN	1.32	2.79	2.77%	1.65	3.74	3.69%	1.94	4.49	4.53%
DGCRN	1.28	2.69	2.66%	1.59	3.63	3.55%	1.89	4.42	4.43%
NE-BJ									
HA	6.00	10.95	26.40%	6.00	10.95	26.40%	6.00	10.95	26.40%
VAR	5.42	8.16	19.28%	5.76	9.07	21.53%	6.14	9.65	23.33%
FNN	4.08	7.22	13.31%	5.14	9.27	18.24%	6.47	11.35	25.57%
FC-LSTM	3.97	7.05	13.05%	4.93	9.04	17.74%	6.06	10.88	23.52%
DCRNN	3.84	6.84	12.82%	4.51	8.49	15.84%	5.15	9.77	19.08%
STGCN	5.02	8.34	19.31%	5.10	8.55	19.82%	5.39	9.09	22.14%
Graph WaveNet	3.74	6.54	12.49%	4.41	8.08	15.79%	4.99	9.20	19.45%
ST-MetaNet	3.82	6.69	13.05%	4.50	8.46	16.93%	5.05	9.74	20.00%
ASTGCN	4.43	7.34	14.65%	5.31	8.86	18.24%	6.29	10.31	22.70%
AGCRN	3.84	6.75	13.80%	4.48	8.41	16.70%	4.99	9.44	19.94%
GMAN	4.08	7.63	14.94%	4.42	8.45	16.51%	4.80	9.18	18.36%
MTGNN	3.75	6.71	12.91%	4.39	8.33	16.07%	4.90	9.38	19.79%
DGCRN	3.56	6.27	12.01%	4.23	7.96	15.10%	4.79	9.23	17.98%

Ablation Study

作者在METR-LA这个数据集上进行消融实验，以验证本文提出的各个组件对模型性能表现的影响。

- (1)dg w/o time: 取消时间戳 T_t 作为DGCRN动态图的输入；
- (2)dg w/o speed:取消实时速度 V_t 作为DGCRN动态图的输入；
- (3)dg w/o h:取消历史隐藏状态 H_{t-1} 作为DGCRN动态图的输入；

- (4)dg2sg: 以静态图代替动态图；
- (5)wo c/l: 模型训练策略取消增量学习方式；
- (6)hypernet mul2matmul: 将Hadamard 乘积替换为matrix multiplication.

TABLE 4: Ablation study on METR-LA.

	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
DGCRN	2.62	5.01	6.63%	2.99	6.05	8.02%	3.44	7.19	9.73%
w/o dg	2.71	5.19	7.04 %	3.12	6.31	8.60 %	3.64	7.59	10.62 %
w/o preA	2.64	5.09	6.68 %	3.02	6.15	8.14 %	3.47	7.30	9.97 %
w/o hypernet	2.67	5.12	6.90 %	3.05	6.20	8.33 %	3.50	7.39	10.03 %
dg w/o speed	2.63	5.01	6.67 %	3.02	6.08	8.10 %	3.49	7.29	9.90 %
dg w/o time	2.62	5.03	6.65 %	3.00	6.09	8.08 %	3.46	7.29	9.83 %
dg w/o h	2.63	5.04	6.72 %	3.00	6.08	8.23 %	3.45	7.23	10.04 %
dg2sg	2.65	5.05	6.73 %	3.02	6.10	8.17 %	3.48	7.28	9.96 %
w/o cl	2.63	5.04	6.67 %	3.00	6.13	8.07 %	3.47	7.35	9.77 %
hypernet mul2matmul	2.67	5.12	6.86 %	3.05	6.19	8.30 %	3.50	7.36	10.06 %

从对比试验可以看出，DGCRN模型对动态图以及hyper-network设计最为敏感。当去掉动态图也即 w/o dg；去掉hyper-network 也即w/o hypernet；不再根据语义生成动态图而是生成静态图和MTGNN一样也即 dg2sg；将公式4中的点乘换为矩阵乘法也即 hypernet mul2matmul。这几个的实验结果最差，其它因素对模型亦有影响只是影响不大。