One Person, One Model, One World: Learning Continual User Representation without Forgetting

SIGIR2021

Data&Code: https://github.com/fajieyuan/SIGIR2021_Conure

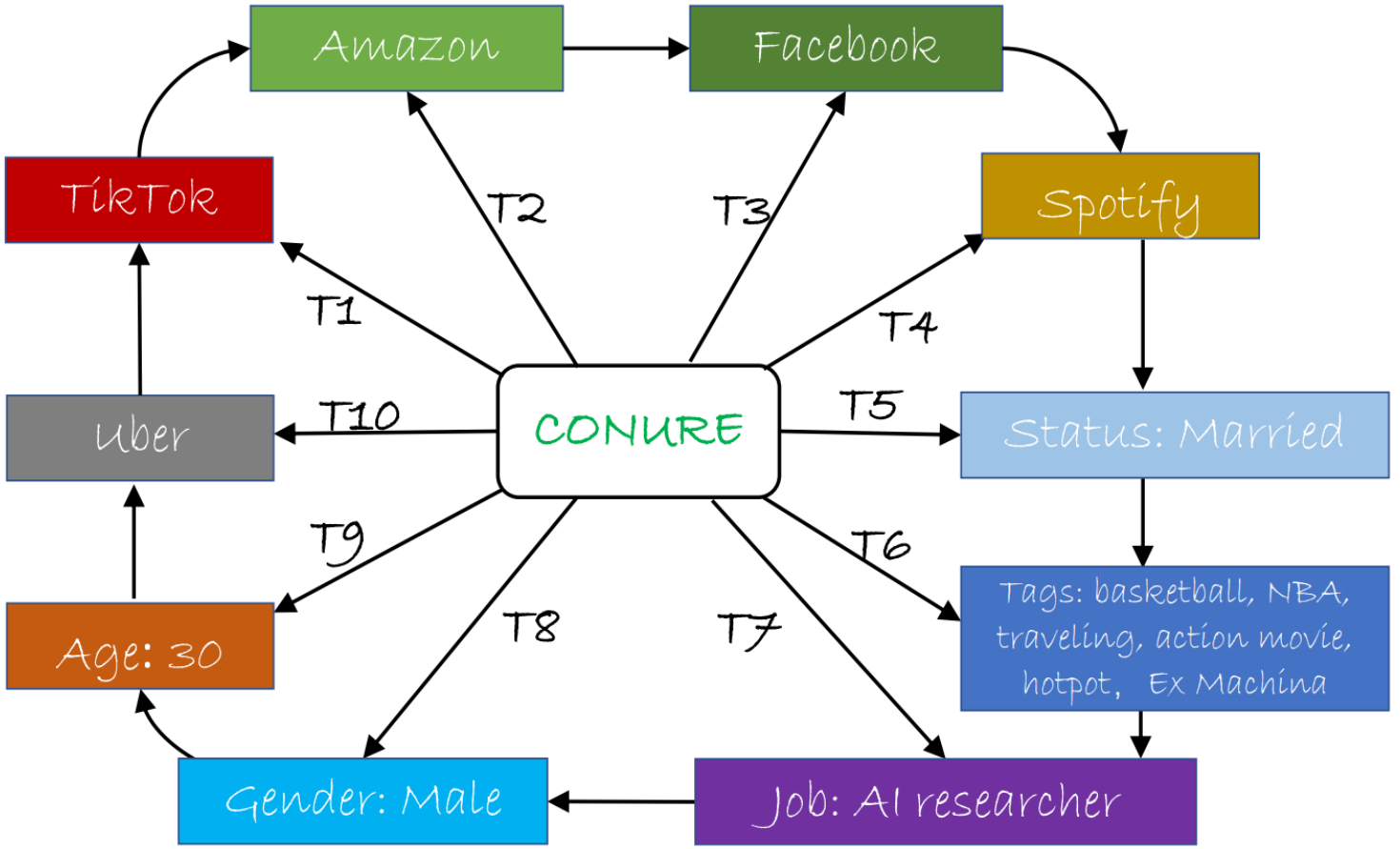
Fajie Yuan（Westlake University, Tencent）, Guoxiao Zhang（Tencent ）,
Alexandros Karatzoglou (Google Research), Joemon Jose (University of Glasgow),
Beibei Kong (Tencent), Yudong Li(Tencent)

Outline

- Motivation
- Related Work
- Conure
- Experiments

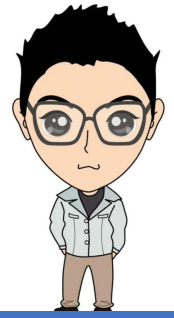
Our Motivation

A person has different roles to play in life!
But all these roles may have some commonalities, such as personalization, habits, preference.

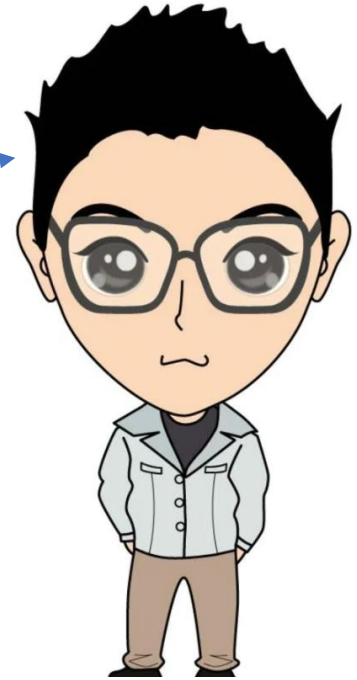


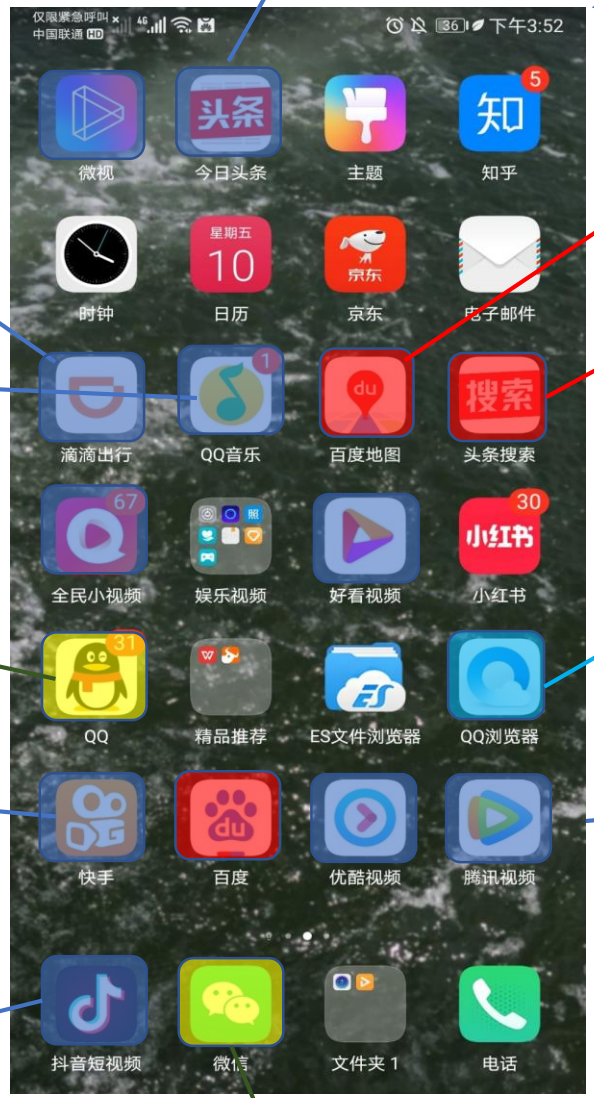
Our Focus:
Whether we can build a user representation model that could keep learning throughout all sequential tasks without forgetting

One Person, One Model, One World



A person has different roles to play in life! But all these roles may have some commonalities, such as personalization, habits, preference.





News Rec

Video Rec

Car Rec

Music Rec

Social APP

Video Rec

Video Rec

Social APP

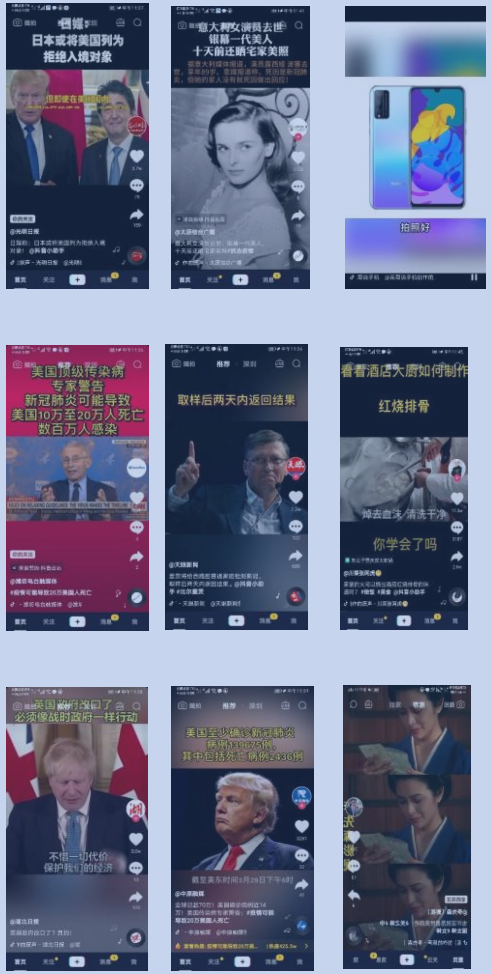
Map APP

Search Engine

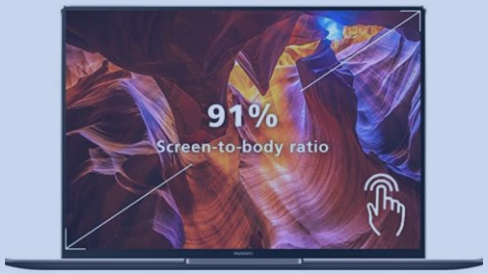
Browser

Video Rec

Clicking logs



TikTok -- warm user



13.9" LTPS DISPLAY | 4.4mm ULTRA-THIN BEZEL | 91% SCREEN-TO-BODY RATIO | 178° VIEWING ANGL

Amazon —cold users

No interaction

Ads --- new users

Using Lifelong learning techniques to solve recommendation tasks

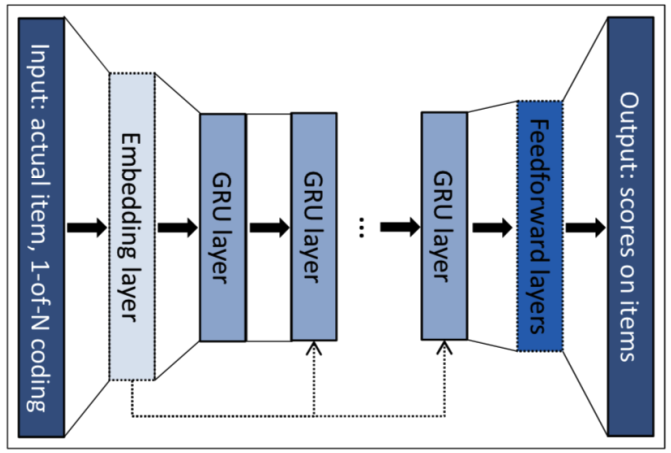
Keypoints

- Necessity and possibility why lifelong learning for UR learning?
- Lifelong learning paradigm throughout all tasks.
- Performance gain for tasks have certain correlations.

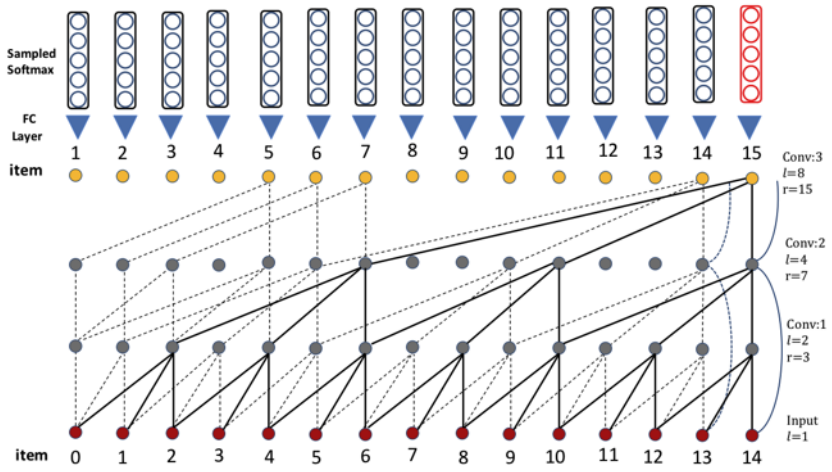
Outline

- Motivation
- Related Work
- Conure
- Experiments

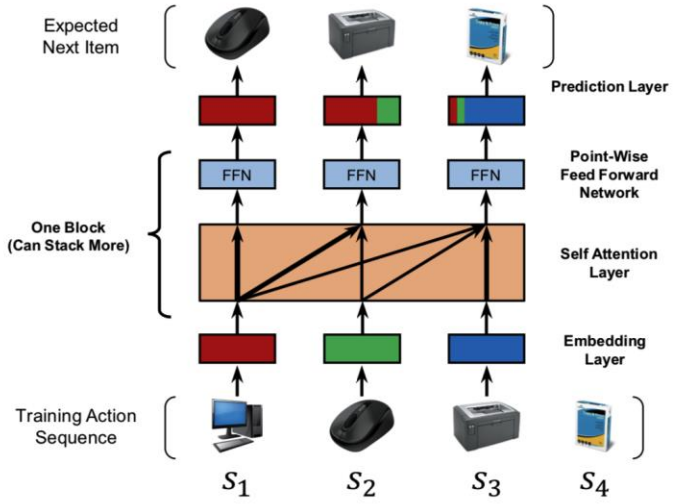
- **Classical UR models** (works well but is specific to only one task)



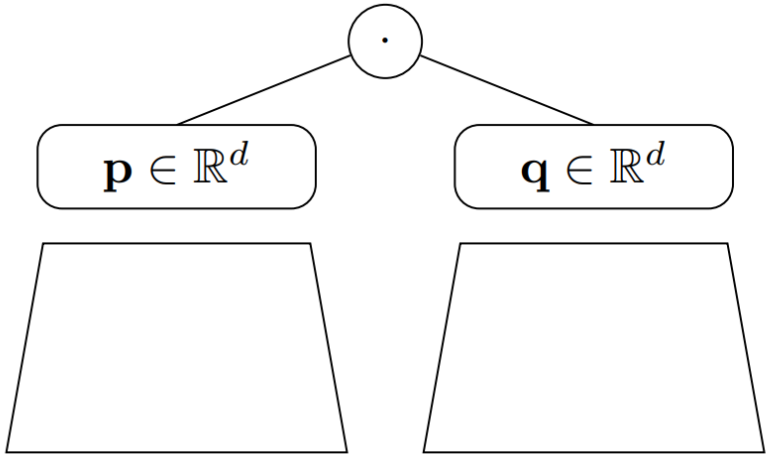
GRU4Rec (Hidasi et al ICLR2016)



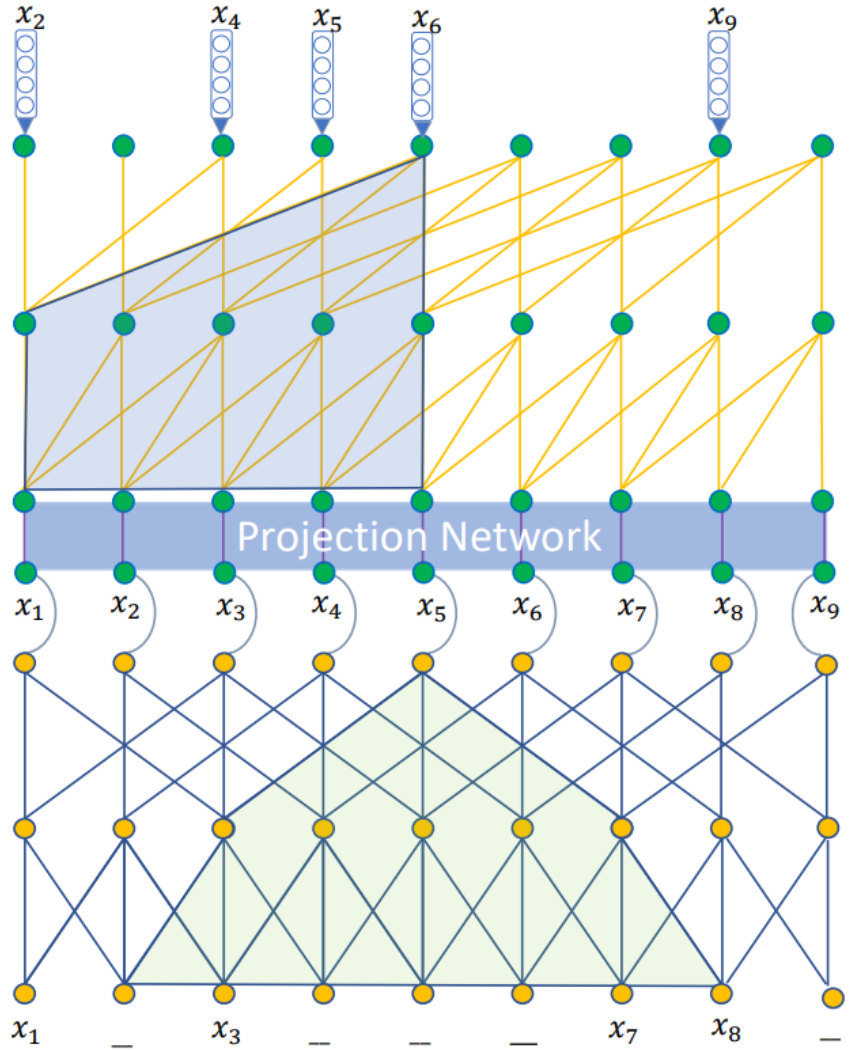
NextItNet (Yuan et al WSDM2019)



SASRec(Kang et al ICDM2018)



DSSM(Huang et al CIKM2013)

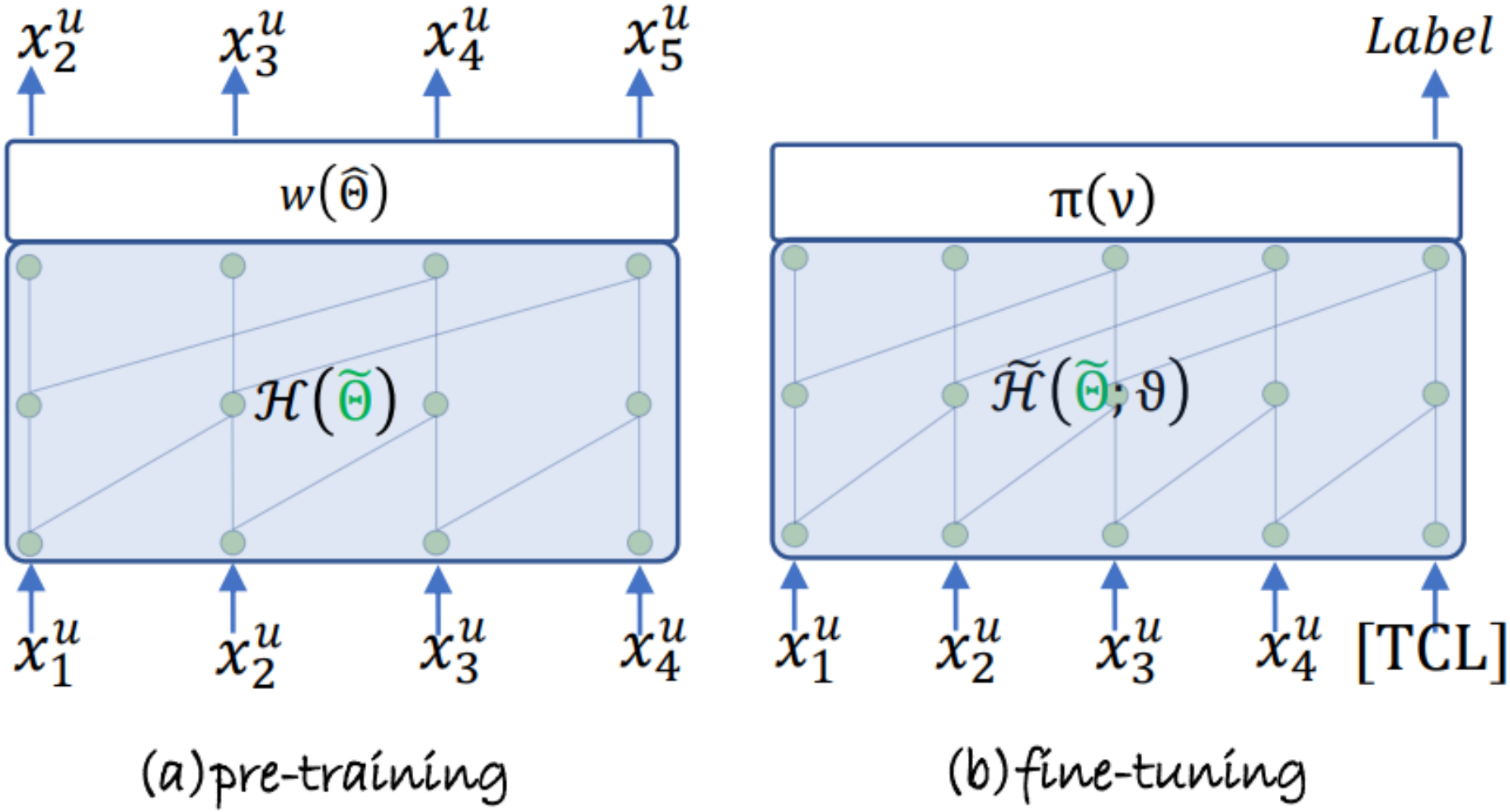


Grec (Yuan et al WWW2020)

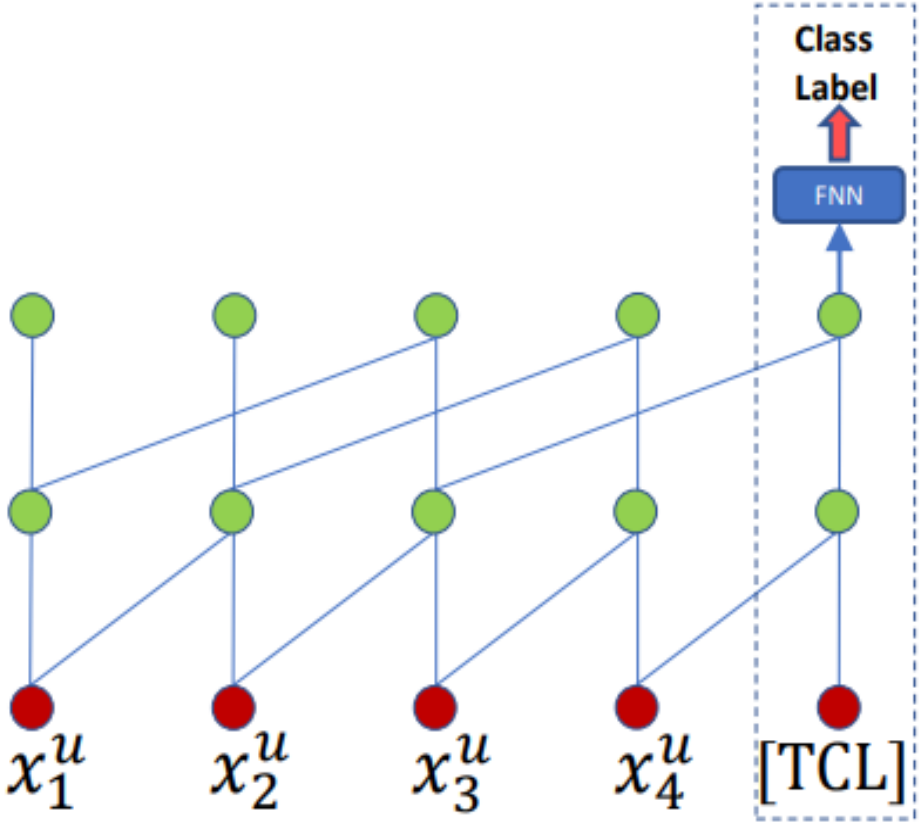
Outline

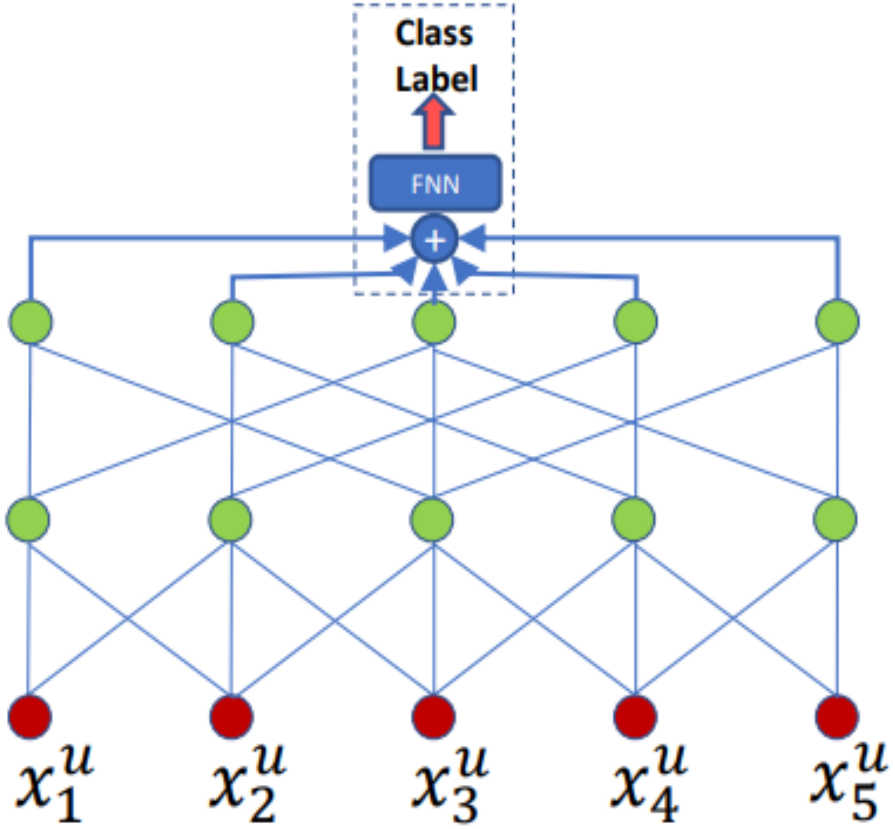
- Motivation
- Related Work
- Conure
- Experiments

- PeterRec (Two-stage Transfer Learning):



- PeterRec (Finetuning):

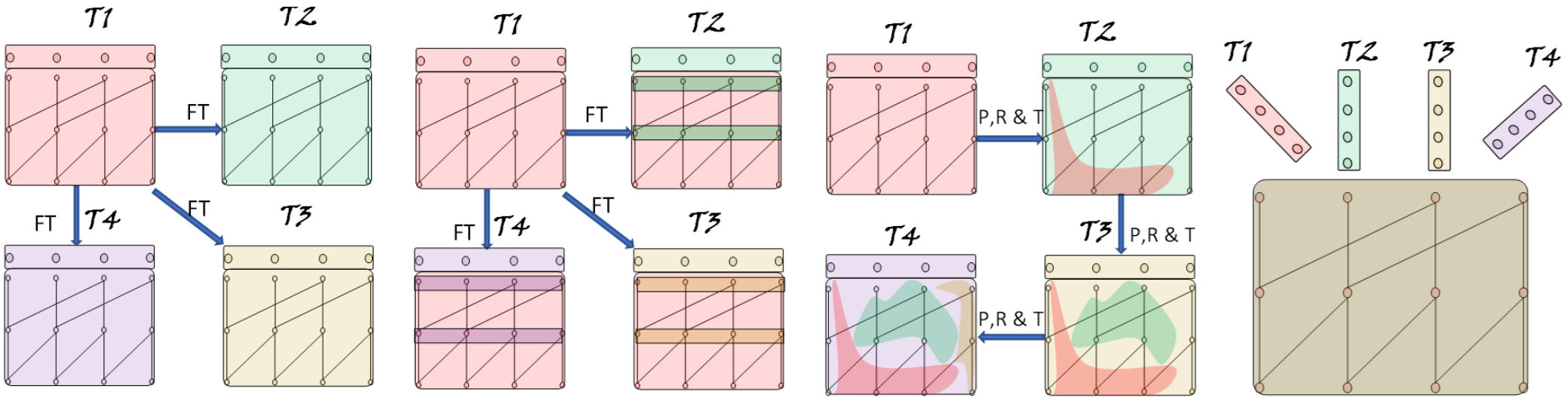




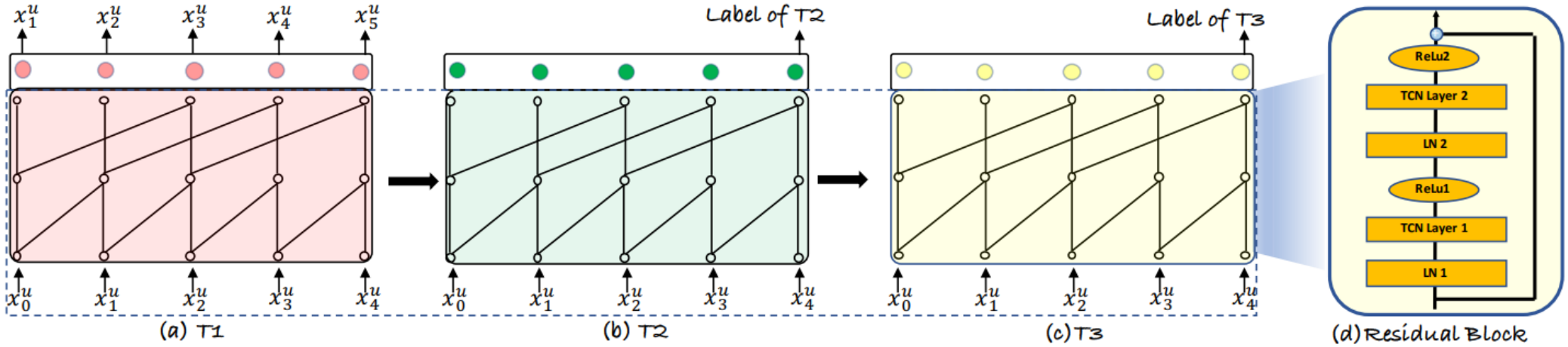
$$p(\mathbf{x}^u;\Theta) = \prod_{i=1}^{n} p(x_i^u | x_1^u, .., x_{i-1}^u;\Theta)$$

$$p(\mathbf{x}_\triangle^u;\Theta) = \prod_{i=1}^{m} p(x_{\triangle_i}^u | \tilde{\mathbf{x}}^u;\Theta)$$

- Transfer Learning Paradigm Comparisons:

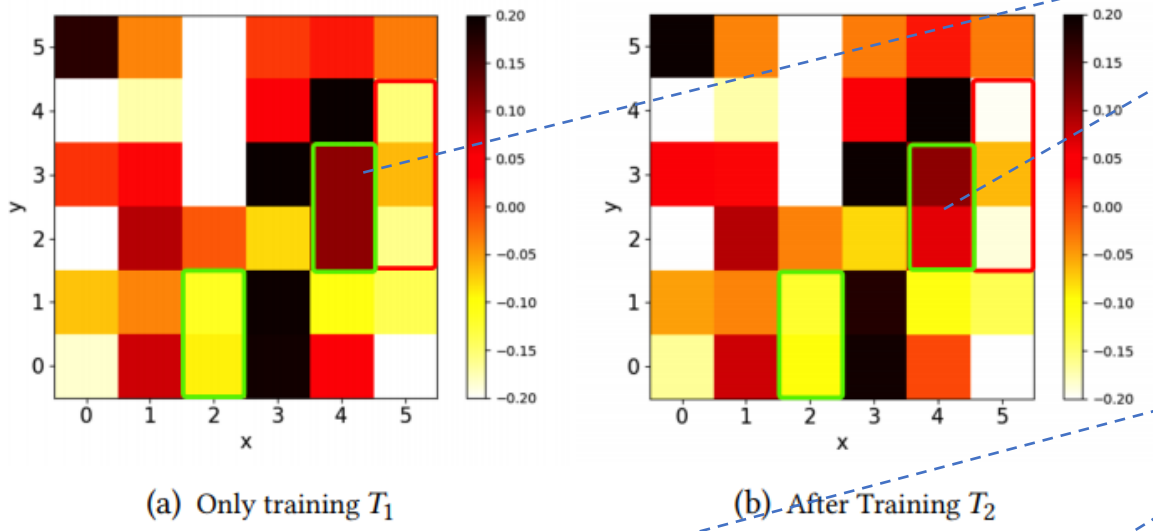


(a) Standard TF (b) PeterRec (b) Conure (b) MTL



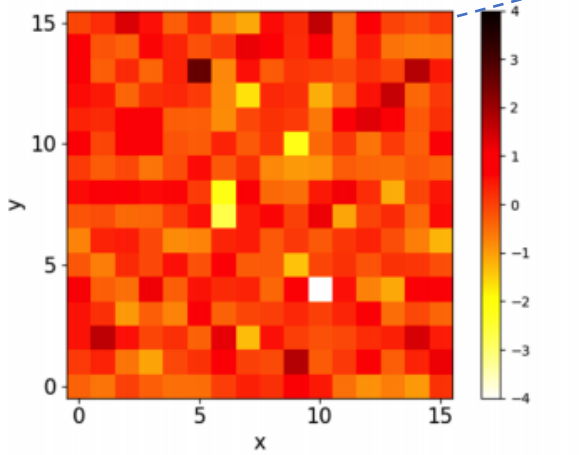
Lifelong learning without parameter preserving

- Catastrophic Forgetting :

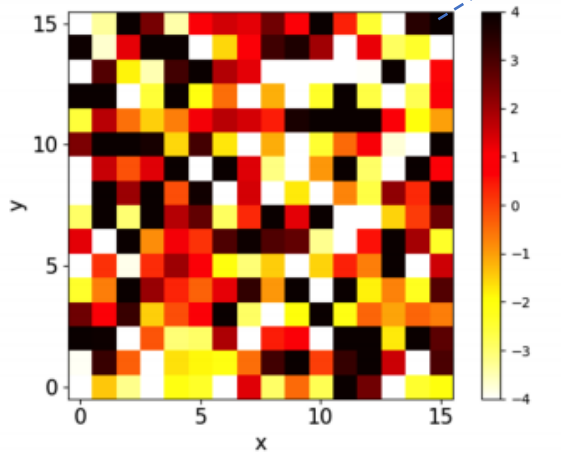


(a) Only training T_1

(b) After Training T_2



(c) Only training T_1

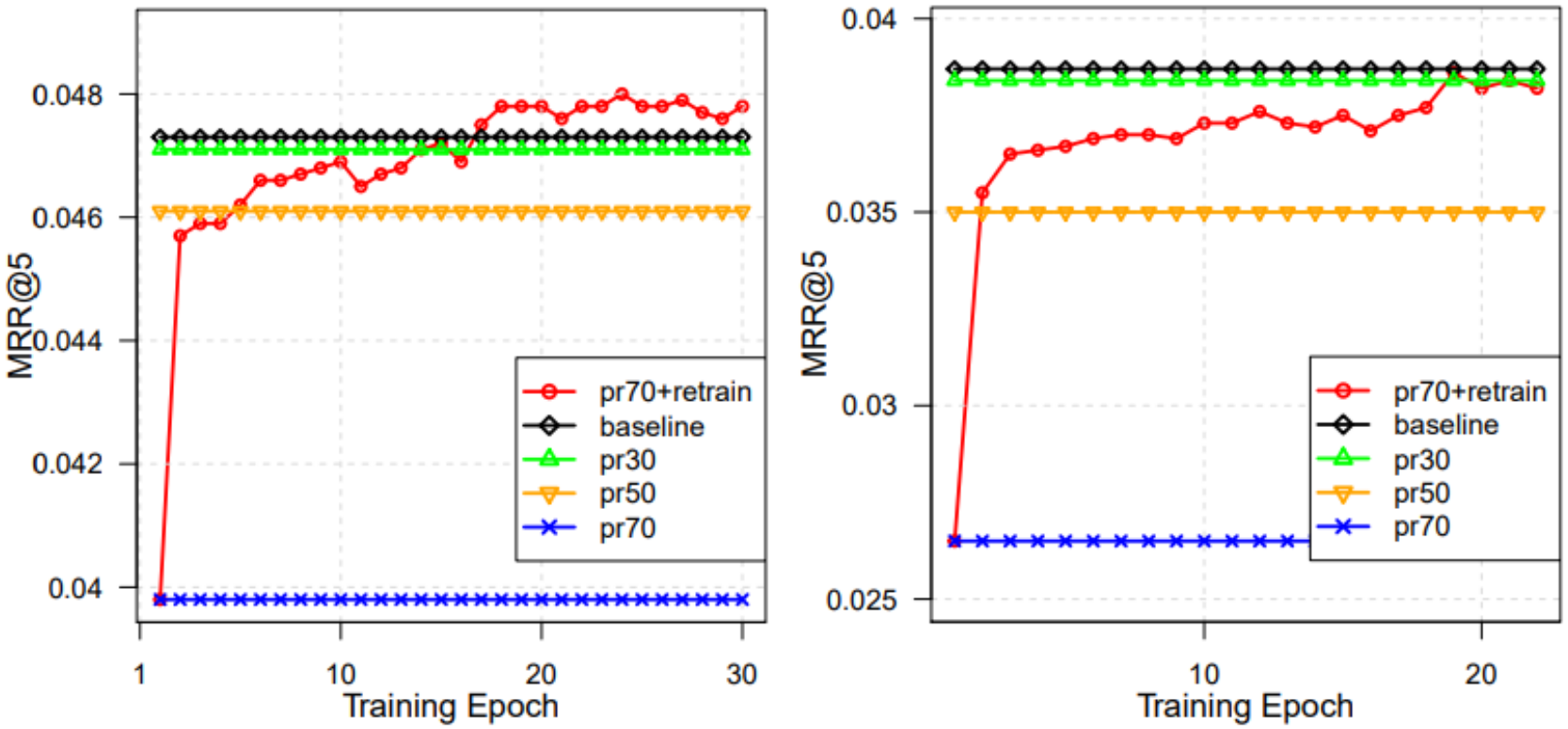


(d) After training T_2

Parameter Changes

Last hidden Vector Changes

- Over-parameterization:



(a) $f = 256$ with 16 CNN layers

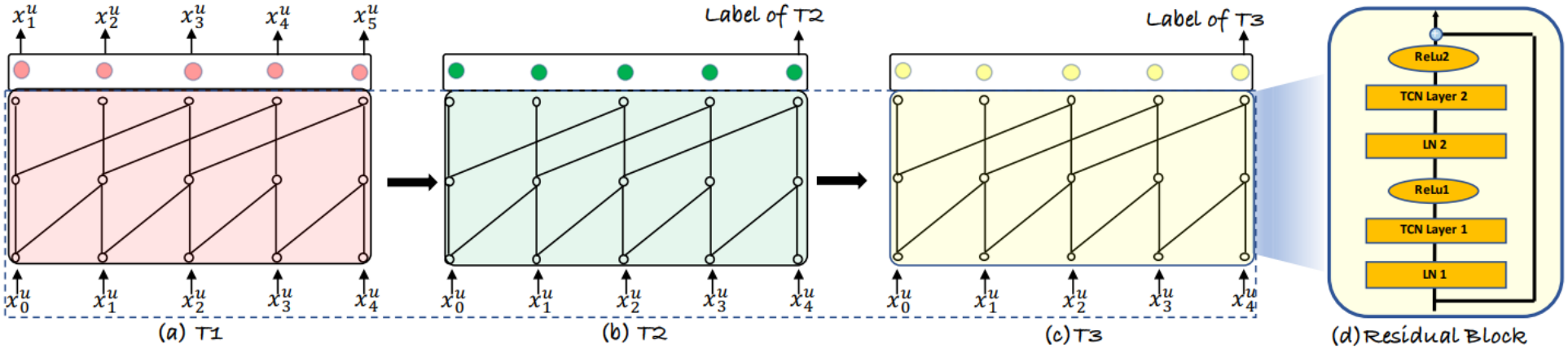
(b) $f = 64$ with 8 CNN layers

— (1) the more parameters are pruned, the worse it performs

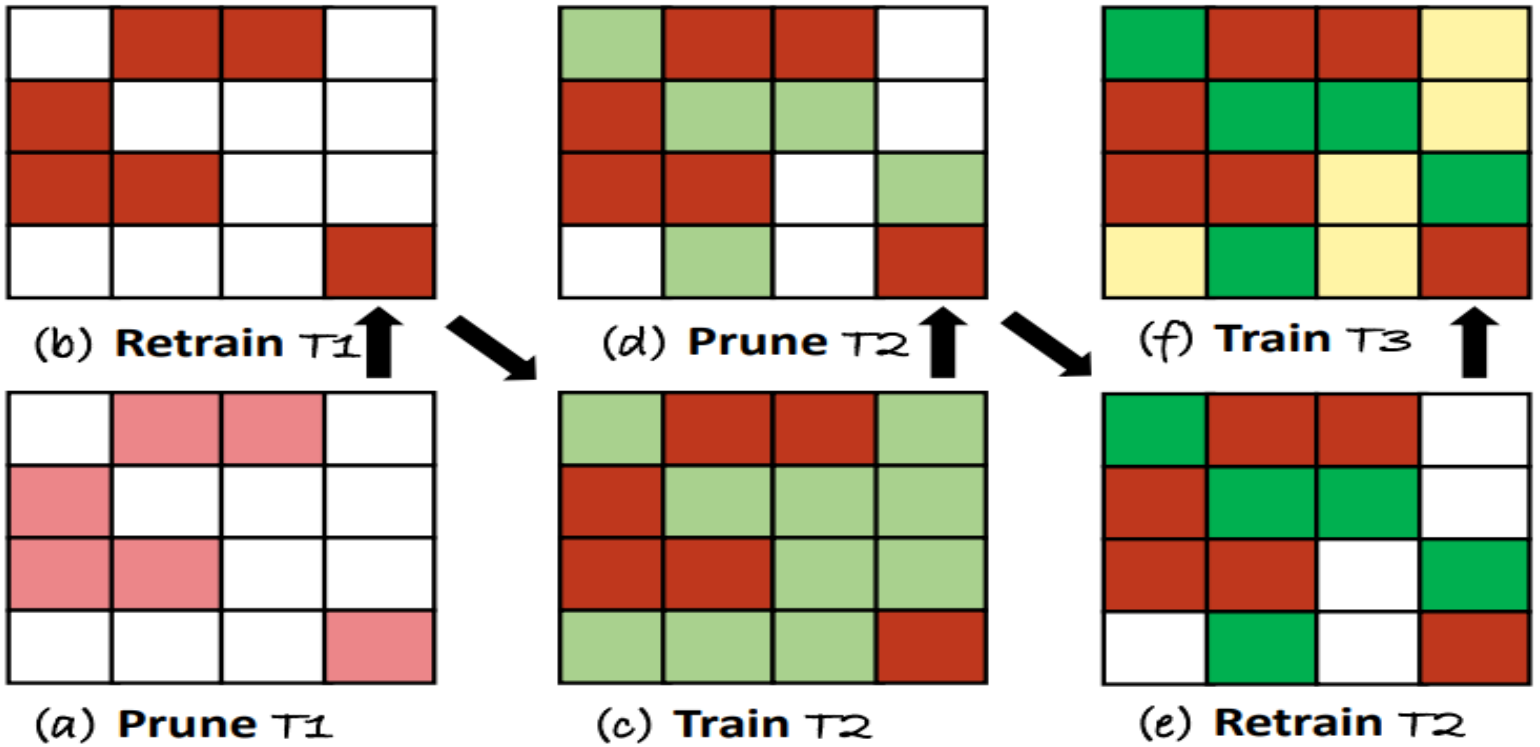
— (2) performing retraining on the pruned network (i.e., "pr70+retrain") regains its original accuracy quickly

— (3) smaller models (i.e., (b)) are also highly over-parameterized

- Conure architecture and learning process.



(d) Residual Block



Conure is conceptually very simple, easy to implement, and applicable to various sequential encoder networks.

Outline

- Motivation
- Related Work
- Conure
- Experiments

- Datasets:

TTL: https://drive.google.com/file/d/1imhHUsivh6oMEtEW-RwVc4OsDqn-xOaP/view
ML: https://drive.google.com/file/d/1-_KmnZFaOdH11keLYVcgkf-kW_BaM266/view

Table 1: Number of instances. The number of distinct items $|\mathcal{X}|$ in T_1 for TTL and ML is $646K$ and $54K$ ($K = 1000$), respectively. The number of labels $|\mathcal{Y}|$ is $18K$, $8K$, 8, 2, 6, respectively from T_2 to T_6 in TTL, and $26K$, $16K$, respectively from T_2 to T_3 in ML. $M = 1000K$.

Data	T_1	T_2	T_3	T_4	T_5	T_6
TTL	$1.47M$	$2.70M$	$0.27M$	$1.47M$	$1.47M$	$1.02M$
ML	$0.74M$	$3.06M$	$0.82M$	-	-	-

- Results:

Table 2: Accuracy comparison. #B is the number of backbone networks. The left and right of '||' represent TTL and ML, respectively. *Conure-* denotes *Conure* that has not experienced the pruning operation after training on the current task. The worse and best results are marked by '▽' and '△', respectively.

Model	T_1	T_2	T_3	T_4	T_5	T_6	#B	T_1	T_2	T_3	#B
DNN	0.0104	0.0154	0.0231	0.7131	0.8908	0.6003	**6**	0.0276	0.0175	0.0313	**3**
SinMo	0.0473	0.0144	0.0161	0.7068	0.8998	$0.5805^{\triangledown}$	**6**	0.0637	0.0160	$0.0259^{\triangledown}$	**3**
SinMoAll	$0.0009^{\triangledown}$	$0.0079^{\triangledown}$	$0.0124^{\triangledown}$	$0.5640^{\triangledown}$	$0.7314^{\triangledown}$	0.6160	**1**	$0.0038^{\triangledown}$	$0.0145^{\triangledown}$	0.0310	**1**
FineSmax	0.0473	0.0160	0.0262	0.6798	0.8997	0.6070	**1**	0.0637	0.0150	0.0262	**1**
FineAll	0.0473	0.0172	0.0271	$0.7160^{\triangle}$	0.9053	0.6132	**6**	0.0637	0.0189	0.0325	**3**
PeterRec	0.0473	0.0173	0.0275	0.7137	0.9053	0.6156	**1**	0.0637	0.0182	0.0308	**1**
MTL	-	0.0151	0.0172	0.7094	0.8979	0.6027	**1**	-	0.0167	0.0276	**1**
Conure-	0.0473	0.0174	0.0286	0.7139	0.9051	0.6180	-	0.0637	0.0183	0.0347	-
Conure	$0.0480^{\triangle}$	$0.0177^{\triangle}$	$0.0287^{\triangle}$	0.7146	$0.9068^{\triangle}$	$0.6185^{\triangle}$	**1**	$0.0656^{\triangle}$	$0.0197^{\triangle}$	$0.0353^{\triangle}$	**1**

— (1) Conure largely outperforms other models on T3 because of the positive transfer from T1 and T2
— (2) Conure, PeterRec and FineAll largely outperforms SimMo because of of the positive transfer from T1
— (3) SinMoAll performs much worse on most tasks (except the last one) because of catastrophic forgetting

- Ablation study- T2 for T3:

Table 3: Impact of T_2 on T_3. *Conure_noT_2* denotes training *Conure* on T_3 after T_1. *Conure_noT_2* and *Conure* both are the *Conure-* versions. TTL20% and ML20% denote the 20/80 train/test split.

	TTL	TTL20%	ML	ML20%
Conure_noT_2	0.0277	0.0245	0.0334	0.0295
Conure	0.0286	0.0261	0.0347	0.0309
Impro.	3.2%	6.5%	3.9%	4.7%

— (1) Without training T2, Conure shows worse results, e.g., -6.5% on TTL20%

- Ablation study- Task order:

Table 4: Impact of task orders. Order1 is the original order as mentioned in Section 5.1. KC, KT and Life denotes the clicking dataset, the thumbs-up dataset and the life status dataset of *Kandian*, respectively. Results on T_1 are omitted due to the same accuracy. The left and right of '||' are results of *Conure*- and *Conure*, respectively.

Orders	KC	KT	Life	KC	KT	Life
Order1	0.0174	0.0286	0.6180	0.0177	0.0287	0.6185
Order2	0.0174	0.0289	0.6154	0.0177	0.0290	0.6152
Order3	0.0174	0.0289	0.6145	0.0177	0.0287	0.6149

— (1) Conure is not sensitive to the task order.

- Ablation study:

Table 5: Pruning and retraining both the embedding & convolutional layers. The left & right of '||' are tasks on TTL & ML.

Models	T_1	T_2	T_3	T_1	T_2	T_3
Conure-	0.0473	0.0175	0.0290	0.0637	0.0191	0.0341
Conure	0.0474	0.0177	0.0295	0.0645	0.0196	0.0347

— (1) pruning also works for the embedding layer

Table 6: Results by specifying *Conure* with Transformer as the backbone network. The left and right of '||' represent tasks on TTL and ML, respectively. 'Mo' , 'FA', 'C-', 'C', denotes Models, FineAll, Conure- and Conure, respectively.

Mo	T_1	T_2	T_3	#B	T_1	T_2	T_3	#B
FA	0.0510	0.0161	0.0243	**3**	0.0654	0.0193	0.0321	**3**
C-	0.0510	0.0177	0.0288	-	0.0654	0.0198	0.0345	-
C	0.0513	0.0179	0.0289	**1**	0.0662	0.0200	0.0357	**1**

— (1) Conure is not restricted to specialized sequential encoder.

— (2) Conure with the Transformer backbone works a bit better than it with NextItNet.

Contributions:

- (1) providing the first lifelong learning paradigm for user representations.
- (2) providing insights for forgetting and redundancy issues in user representation models
- (3) designing Conure, the first lifelong learning algorithm - smple and easy to implement
- (4) instantiazing Conure with NextItNet and Transformer backbones
- (5) Extensive experiments with SOTA performance with many new discoveries and insights

Q&A

Case study:

user behavior on T1

新用户推荐效果： 左侧是A场景用户行为
右侧是B场景下预测出来的用户行为。A B为不同的推荐引擎。 该用户在B场景下均为新用户或者只有少量的点击被仅用作评测。

new user on T2, behavior predicted by Conure

宋晓峰一家五口生活照曝光，e3[illegible]
原来全家都是高颜值，5omi27v[illegible]
就属宋晓峰最丑！:e3072yxnw30，c306491vo1o
2014-04-17我为喜剧狂:5omi27[illegible]1tmwo
文松打出“一张4”，d08[illegible]
宋晓峰直接出“大王”，i[illegible]/c
文松顿时一脸懵圈:c3[illegible]36h0n0
宋晓峰：我媳妇为什么不让握手？就凭这一句话，m[illegible]18leoyh
台下观众们都被逗乐！:t0[illegible]6p9hx
宋小宝被谢楠质问：你什么玩意！吴京也忍不住笑了:d[illegible]

赵本山上台领奖太搞笑：硬把颁奖现场弄成相声大会，逗乐全场大[illegible]
峰哥就是屌，澡堂吹牛不打草稿，吓得壮汉赶紧跑：[illegible]
https://v.qq.com/x/page/6945f50a928674ah.html:69
https://v.qq.com/x/[illegible]
1997年，卓琳代小平同志参加香港回归盛典，现场响起经久不[illegible]
赵四谢广坤又同台演小品了，笑料百出，观众尖叫不停:95[illegible]
https://v.qq.com/x/page/2935f9fd88c009ah.html:29
江疏影，独一无二的嘟嘴，这是被谁气到啦？:0[illegible]
皮蓬生涯最“妖”一球，乔丹都愣了！这到底是打球还是耍猴？:92
CBA总决赛经典冲突1：广东男篮掌击门，杜锋头顶外援被打晕:1035fb
天龙八部：乔峰的演技有多好？看看这段你就明白了，多少人
王大拿说不差钱，结果刘大脑袋2亿拿下金店，王大拿当场傻[illegible]
副省长太嚣张，谁料军区处长一句话让他秒怂，真是太解气了
超级大山炮之夺宝奇兵：静静的看完这一段，不笑算我输。:0
https://v.qq.com/x/[illegible]487ah.html
https://v.qq.com/x/page/7855fb[illegible]480ah.html
赵四最硬气的两场打架
也是场面最大的两次
百看不厌:3855fb73d6d435ah:8054
https://v.qq.com/x/page/80[illegible]ah.html
郭德纲占于谦便宜被反杀，相声皇后真不简单！:5505f6f7d
颁奖礼与郭德纲同台的女星，范冰冰被套路，柳岩被逗得没法
赵本山颁奖晚会现场：硬把获奖感言说成了小品，把台下大腕
武警突击检查酒店，不料打开包厢，里面坐的竟是自己老婆！
媳妇去公司找丈夫，见丈夫有这么漂亮的秘书，瞬间不淡定了
https://v.qq.com/x/page/893[illegible]ah.html
https://v.qq.com/x/page/9865fb[illegible]ah.html
说实话，你是不是来砸场子的！:1[illegible]
https://v.qq.com/x/page/2865fbf52e4536ah.html
影视：朱德被污蔑，周总理直接请示伟大领袖危机顿时解除，
岳云鹏大型吹牛现场，那叫一个一本正经，就连郭德纲都被整
https://v.qq.com/x/page/86[illegible]ah.html
刘华强霸气上场：给我拼命，你有这个实力？太爷们了:7455
小伙要特别服务，没想到这个服务和想象中好像有点区别呀:5
美就算了，但是又很拽，就很过分啦！:85[illegible]
就知道谭维维不简单！连《刀剑如梦》也能翻唱超越，开嗓飒
https://v.qq.com/[illegible]html
https://v.qq.c[illegible]aw.html
小伙被老板看上，本想着干一番大事业，没想到却是保洁工作
普京大帝的伊尔96-300，可在空中发动核战争，接机场面太[illegible]
史上最嚣张的绝杀！大将军转身提前庆祝 阿里纳斯五大绝杀:
https://v.qq.com/x/[illegible]
元帅俘虏一屋子军官，进屋之后与其挨个握手，胸怀真是让人
宋晓峰被列为编外人员，这大红裤衩子把兄弟们逗笑了！:81
倚天屠龙记：不愧是明教总部的，张无忌打的都费劲，武功太
https://v.qq.com/x/page/4415fb39521640bk.html

中美共同的敌人是谁？特朗普该醒醒了，
崔大凯直接给美指出明路:z31489g73sd,

|https://v.qq
|https://v.qq.com/x/page/4045fc5fca2792ah.html:4(
|https://v.qq.com/x/page/0635fb0a281550ah.html:0(
|https://v.qq.com/x/page/8675fbdcd2e423ah.html:8(
|https://v.qq.com/x/page/1785
|https://v.qq.com/x/page/8695fbb41a9492ah.html:8(
|中国又一好消息，南海30万吨黄土涌出海面？西方直言：太厉害了！:
|今天才知道，找胶带头这么简单，再也不必拿手抠了，早学会就好
|渔民看见鲨鱼数秒被"海怪"撕裂！靠近一看，竟然是这家伙！:6
|新型航母指日可待，116名舰载机飞行员待命，美感慨：这才是大
|美国卫星路过中国沙漠，发现了"神秘"黑点，科学家们坐不住了:
|震撼！成都飞尼泊尔，飞跃喜马拉雅看珠峰，这场面很难见到:25
|大国没料到，为何我国突然亮出"王牌"，让美不能制霸世界！:6
|一只猫在桥头伺机而动，接下来一幕，让人直呼太精彩:7735fc(
|偷学来的电线接头绝活，手法专业又牢固，又不用缠胶布，长见识
|狞猫敢单挑群狼，却唯独怕它，一旦遇到连喘气的机会都没有:8
|我国又一逆天工程，削平33座大山，世界上最大山区城市即将问t
|牧民上厕所咋解决？草原姑娘一本正经示范讲解
|所见之处皆是厕所!:565
|毛主席会见美国代表人，主席抽着香烟，神情自若的说了一段话!
|中国湖南有位"峭壁飞人"，专门帮人上龙头香，一天能赚上万元!
|普京表示，我从不这么等人而你是第一个，这个人会是谁呢:681
|全球最罕见的现象，堪称百年难遇，科学家至今都无法解释！:2
|云南惊现"悬浮山"，村民凑近一看吓得拔腿就跑，牛顿坐不住了!
|曾仕强：男人最怕数字"九"，逢"九"会有大劫难，这是什么原因
|20米坠桥事故完整视频来了，原因反转了，结果也反转了:2365f
|男子将野猪和家猪关在一起，没想到意外发生了，视频记录全过程
|我国"打穿"喜马拉雅山，尼泊尔举国欢庆，印度人却不乐意了！:
|这才是软线和硬线的正确接法，不看不知道，原来以前都接错了:
|马拉多纳最离谱一球，如果不是镜头记录下来，国足绝不信这是真
|农民工的新发明，在果园里真实用啊，快收藏起来:1165fb48fd
|凶残野狼追杀驼鹿，驼鹿机智跳入水里，下一秒太激烈了！:806
|假老虎放在猪面前，猪会是什么反应？男子亲自试验:8285fbf7
|中国第一高楼，耗资148亿却建"歪"了，当地希望50年后拆除！:
|马里亚纳海沟有多恐怖？把10斤铁球丢进去，结果会怎样呢:342

2020-07-22脱口秀大会 第3季:mzc00200qq2[redacted]
2020-09-16脱口秀大会 第3季:mzc00200rapy3ac, [redacted]3ac
nba搞笑配音：隆多跟教练闹矛盾故意失误，q32073b8ihg
麦基快攻翻车！:q32073b8ihg, [redacted]
nba搞笑解说：你没见过的邓肯，[redacted]
詹姆斯：你说的未来还来不来？[redacted]h
nba搞笑解说：那些用生命假摔的球员，k31649msclr
韦德演技是最差的！:i31646dtu[redacted]zr
nba搞笑解说：想谋权篡位的球员，y3[redacted]
詹姆斯场边执教泰伦卢一脸懵！:o3164hk063h, q3164b5wb5d
nba下赛季年薪前10名：詹姆斯第六，y3163rhumkp
一半在养伤，h3163d5axuw
谁是冤大头？:k3164[redacted]
nba巨星约架：从球场打到格斗场，k[redacted]o2
什么仇什么怨？:m3164i6byzr, z31[redacted]b
nba十大震撼时刻：詹姆斯爱上韦德霸气隔扣，[redacted]zph1
威少复仇绝杀热血沸腾！:y3164n02yta, r3163ex1rfy
nba经典复刻瞬间：詹姆斯复制艾弗森跨越泰伦卢，[redacted]
很多人没见过吧！:g[redacted]
谁都不服的罗德曼为何对乔丹言听计从？只因乔丹一句话，r31621drjho
神果然是神！:y316[redacted]
nba判无效的神仙球：浓眉360转身中场3分，x31622dylri
哈登这球差点创纪录！:h[redacted]
nba搞笑解说：奥尼尔和霍华德比拼罚球，f31627libny
这结果奥尼尔自己都不敢信！:q3163rx[redacted]

|施罗德：这一球，就当我去湖人的定金了！:0695fbf592a247bk:7
|未来怎么还是你的！詹姆斯10大不讲武德时刻，站帽库里羞辱格林总
|字母哥为何服他？回顾詹皇西决最燃一幕：35岁追着23岁小伙跑：[redacted]
|经典回顾：44分压哨绝杀！詹姆斯拿下天王山，卢指导差点晕倒:688
|超巨是怎么防守的？锡安吃遍NBA中锋，最终在詹姆斯面前吃瘪！:152
|当年的"四不控卫"哪个是你的最爱？他们的现役接班人分别是？:0215f
|为何区俊炫会沦为水货状元？看看他当初怎么试训北控的，难怪打不出来
|杜兰特格林一场比赛2次惹毛安东尼 甜瓜直接摔发带了:64[redacted]
|保罗生涯最"搞笑"的一次暴扣，库里安东尼笑翻了：你还有这一手？:9
|听说你叫马上腿？看看周鹏当年是怎么怼翻马布里的，真过瘾:2075fc3
|【徐静雨】新赛季NBA前二人是谁？詹库无疑，杜兰特老三位置难保:09
|最后的巅峰是打火箭！考辛斯勇士生涯最强一战 下半场20分暴走:7
|雷阿伦与纳什对飙绝杀 三次剧情反转 真怀念过去每队都有巨星的时
|当詹姆斯遇上安东尼，奥运会最快得分就此诞生，我就看了20遍！:364
|字母哥全力防詹姆斯会怎样？詹姆斯：好像也就这样:1065[redacted]144
|https://v.qq.com/x/page/[redacted]35fc0d
|皇帝保镖归来！杜德利生涯5大仗义时刻，怒推恩比德，一人硬刚魔术全
|CBA总决赛经典冲突1：广东男篮掌击门，杜锋头顶外援被打晕:1035fb8e
|原来施罗德来湖人早有预兆！倒地后起身拉起浓眉，这么懂事来湖人没
|职业与业余的差距有多大！曹芳自信满满防守孙铭徽，结果被扣得篮球梦碎:9
|ESPN评NBA历史十大晃倒：哈登才排第六欧文第十 第一真的太残暴
|https://v.qq.com/x/page/7075fc7d6e4972ab.html:7075fc7
|皮蓬生涯最恐怖一球，对手：真不敢追，谁追谁断腿！:9[redacted]2fa5
|科比对飚姚明！越看越激动 如果当年携手能拿几个总冠军？:597[redacted]
|一场比赛，两个排球式大火锅，詹姆斯就是这样征服施罗德吗？:1175f
|罗斯给哈登上课 没有斗志的人不会赢 碰瓷2+1也不会长久:9185
|NBA最被低估的一届选秀！他们能成为下一个黄金一代吗？:8865
|张宁一对一单挑胡明轩，本以为是神仙打架，没想到姜还是老的辣
|你高兴的太早了！NBA历史五大反绝杀时刻，库里姚明齐上榜 詹皇
|多年后的艾弗森单挑马布里，可惜再也没有观众可以欣赏这一刻！
|打球有点动作很正常？NBA10大肮脏时刻 前3皆出自一人之手！:
|经典回顾：骑士逆转73胜勇士，夺队史首冠:2[redacted]9250a
|当詹姆斯连进5个三分！湖人替补席都沸腾了！:4[redacted]
|史蒂芬森十大无厘头演技：隔扣裁判吹气詹姆斯，最骚的是这个:1
|联盟谁有资格单换詹皇？波波维奇：8年前有一个，5年后有好多个
|18年詹姆斯离开骑士，联盟其他29支球队都是什么反应？:9365f
|这球打的开挂了吧！詹姆斯与浓眉哥10大残暴时刻，死亡隔扣打服
|詹姆斯这球让字母哥目瞪口呆 你以前开推土机的吧？:4245fc3b
|进攻万花筒？枸杞哥厉害了，训练师晒维金斯单挑集锦，勇士这下
|4日《长暂停》 韦少体面挥笔分手信 底薪男扬言三分虐库里:815
|当威金斯真正觉悟展现天赋时！对手看到也只能无奈摊手！:6655
|詹姆斯生涯过人大赏：灭死神过科比，能用脑子决不靠身体！:66
|詹姆斯vs欧文！2019-20赛季湖人对阵篮网詹姆斯、欧文高光表现
|再见兄弟！比尔深情告别沃尔！回顾奇才双枪季后赛最强一战:35
|经典回顾：詹皇欧文同砍41分攻陷勇士，吹响史诗逆袭号角:4815