腾讯社交广告
Tencent Social Ads

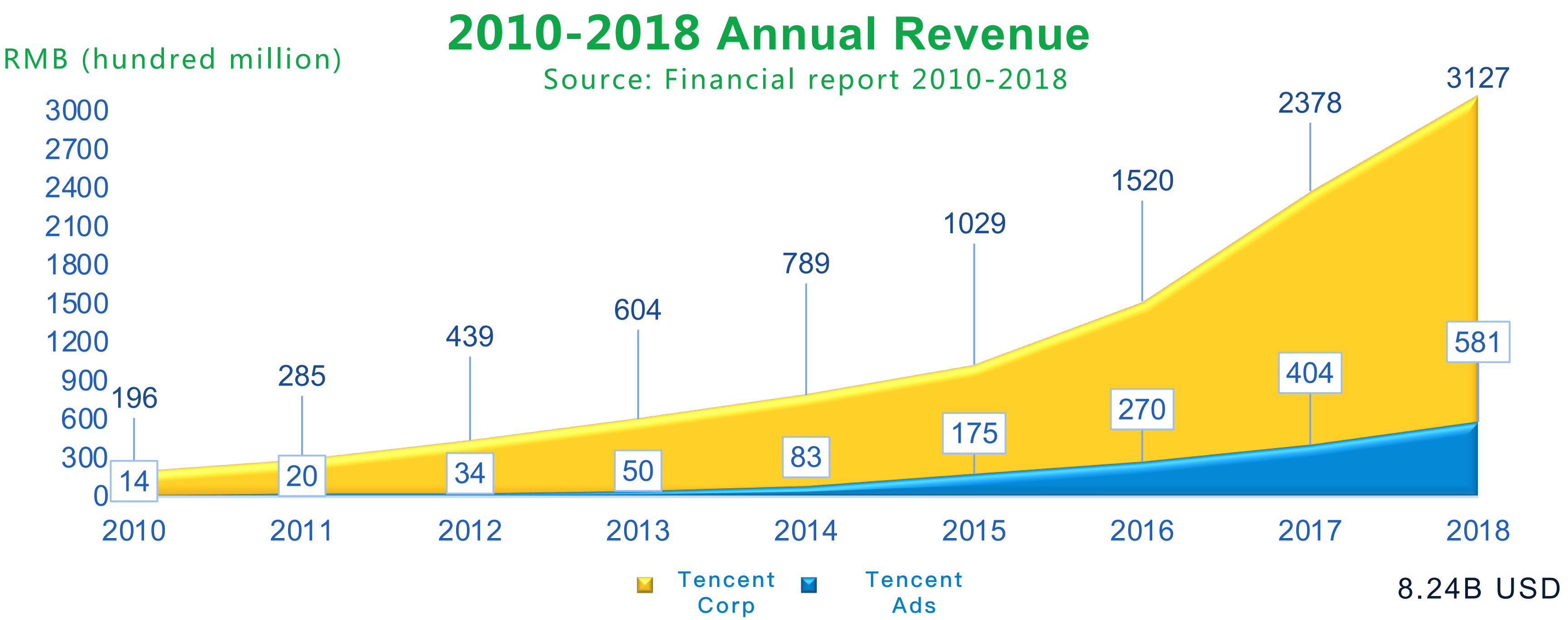
The Power to
Connect Businesses and People
赋能商业 | 始终于人

Tencent Ads: Interesting Problems and Unique Challenges

Haishan Liu

Tencent Marketing Solution – An Overview

As the leading marketing platform in China, Tencent Ads has experienced a rapid growth over the recent years

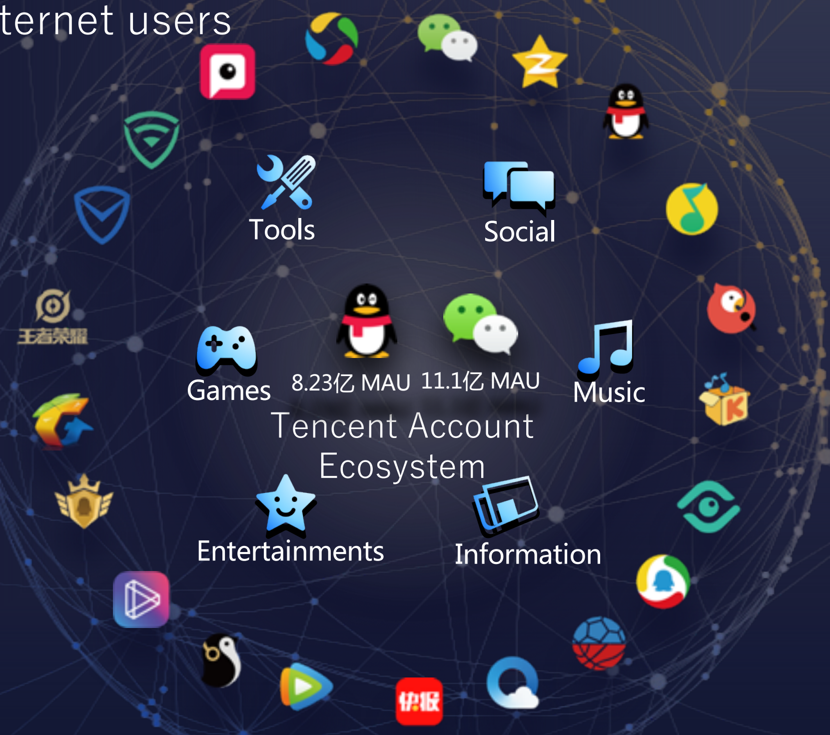


Tencent Marketing Solution Core Capability – Panorama Connection

Top Internet Traffic Sites in China

Tencent Marketing Solution Core Capability – Digital Intelligence

Covers 1B+ Chinese internet users



source ：Penguin 2018 Data
Tencent 2019年Q1 Financial Report

Tencent Marketing Solution Core Capability – Diverse & Engaging Formats

Xiaomi x Moment

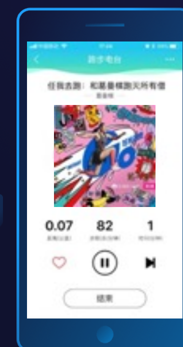
Celebrities in Moment

Lancôme x QQ

AI Cosmetic Try-up

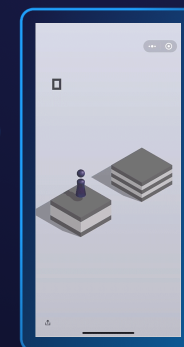
Cadillac x News Feed

Commercial Content



Nike x QQ Music

Runner’ s Radio



McDonald x WeChat Game

Gamelet

腾讯广告

Tencent Marketing Solution Core Capability – Diverse & Engaging Formats

Enhanced Branding via Social Endorsement

WeChat Moment @Friend

QQ Red Packet

Tencent Ad Ecosystem Helps Businesses Monetize Site Traffic

优量汇

Serving ads on Tencent Audience Network with the help of Tencent Ad Platform technologies

No. of Quality APPs	Hero APP %
100k+	75%
Mobile MAU 500M	100M+ Impression / day

腾讯优量汇合作伙伴

微信广告

Certain official account and applet owners can monetize their content by utilizing the built-in ad slots

3B+	600+	40%	40%
First 3Q of 2018 Shared Revenue	No. of site owners with 100k revenue/month	No. of individual site owners	Revenue for individual site owners

Monetization for WeChat Official Accounts

Article Bottom | Article Middle | Reciprocal | Cash-back products

Monetization for WeChat Applets

Banner | Incentive video | Splash screen | Reciprocal

Tencent Ad: Technology for Good

Charity Advertising Competition

我是创益人
CREATE FOR GOOD

2017-2018 Charity Advertising Facts:

160M+	1700+	62	100M+
In funding	Participating teams	Award winning pieces	People reached for engagement

《不朽的丰碑》
以本地推广广告唤起共鸣
致敬抗战老兵

《一起捐脸，勇敢代言》
联动AI、天天P图
让每个人为唇腭裂发声代言

《一个人的球队》
直击器官捐献话题
小屏感动大屏
由社交平台发酵的社会话题

《灯山行动》
用微信小游戏体验上学路
化氪金为公益

Machine Learning Applications

Smart Targeting

Smart Creative

Smart Bidding

Example Analysis of a Free-Text Chinese Query

喜剧脱口秀节目《吐槽大会》为什么能火？

- **Segmentation:** 喜剧/ 脱口秀/ 节目/ 《/ 吐槽/ 大会/ 》/ 为什么/ 能/ 火/ ?/
- **PoS tagging:** 喜剧/n 脱口秀/n 节目/n 《/w 吐槽/v 大会/n 》/w 为什么/r 能/v 火/n ?/w
- **NER:** 喜剧/n 脱口秀/n 节目/n《/w **吐槽大会**/NS 》/w 为什么/r 能/v 火/n ?/w
- **BoW:** 喜剧:0.45, 脱口秀:0.52, 节目:0.36, 吐槽:0.39, 大会:0.38, 火:0.31
- **Keywords extraction:** 吐槽大会:0.92, 喜剧:0.78, 脱口秀:0.88
- **Keywords expansion:** 情景喜剧:0.84, 喜剧电影:0.80, 今晚80后:0.84, 相声:0.74
- **LDA:** 8867:0.15,明星(0.09) 中国(0.033) 韩国(0.03) 歌手(0.03) 香港(0.02) 演员(0.02) 8191:0.11,赵本山(0.05) 小品(0.03) 春晚(0.02) 小沈阳(0.01) 喜剧(0.01) 台词(0.01)
- **Embedding:** 1:0.32,2:0.56,3:0.01,4:0.66,5:-0.12,6:0.89,7:-0.45,8:0.23,9:0.14,10:0.54
- **Classification:** 06:娱乐休闲—综艺—喜剧:0.95, 05:娱乐休闲—综艺—脱口秀:0.93

Syntactic

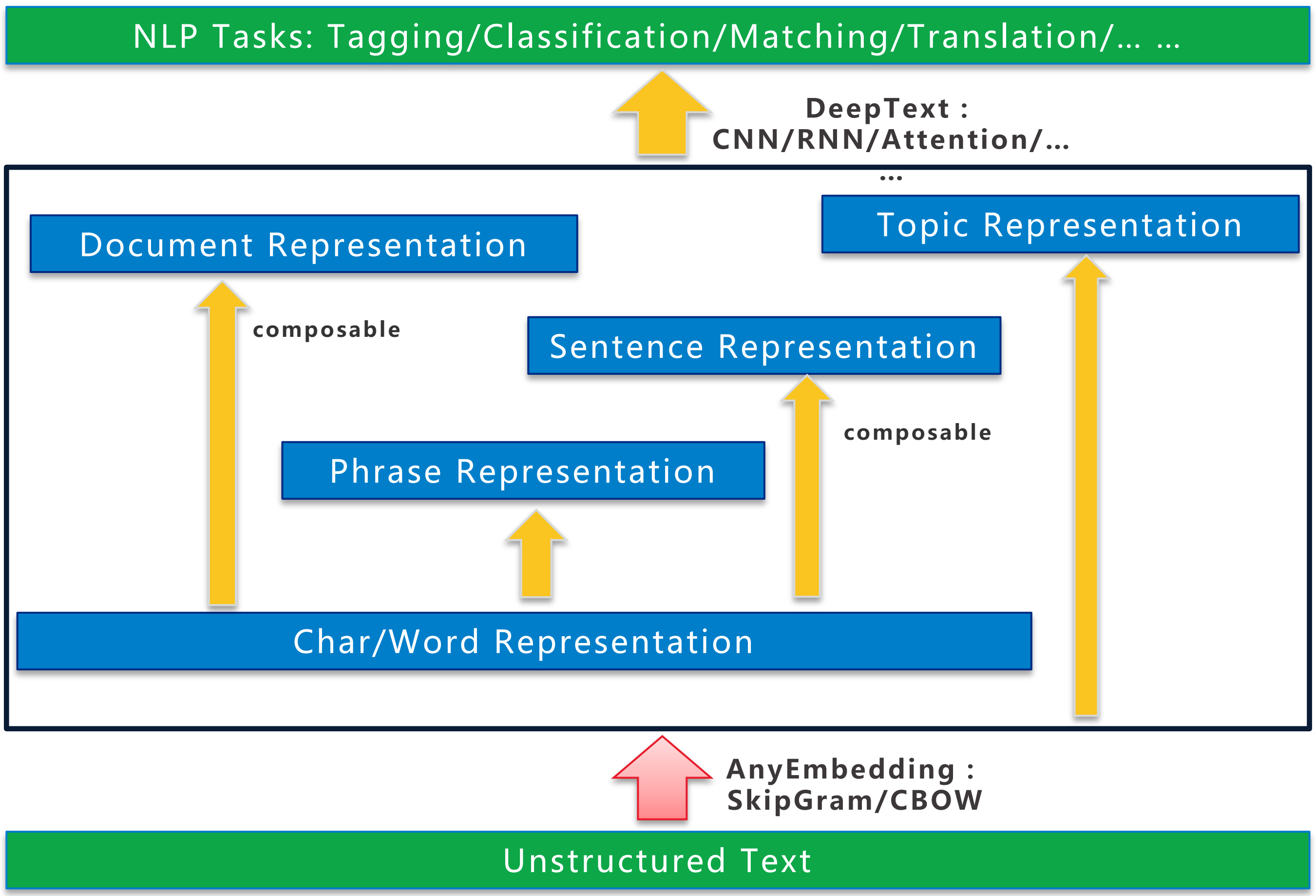
?

Semantic

NLP at Tencent Ads – an Overview

Service	TextMiner RPC Service (C++/Python/Java/Scala)		TextMiner Lib (C++)	
Module	Text Classification			
	BoW	Keyword extraction	Keyword expansion	LDA
	Chinese Segmentation	PoS Tagging	NER	Embedding
Tool	TensorFlow	PyTorch	FastText	FAISS
	Crawler (TEG)	Data Kitchen (MIG)	XGBoost/ LightGBM	MaxEnt
	Labeling: LabelMe	Distributed LDA: Peacock/LightLDA	Distributed Embedding: AnyEmbedding	DNN + NLP: DeepText
Data	Multi-context behavioral data	Dictionary	Industry Knowledge Repo	Word Bank

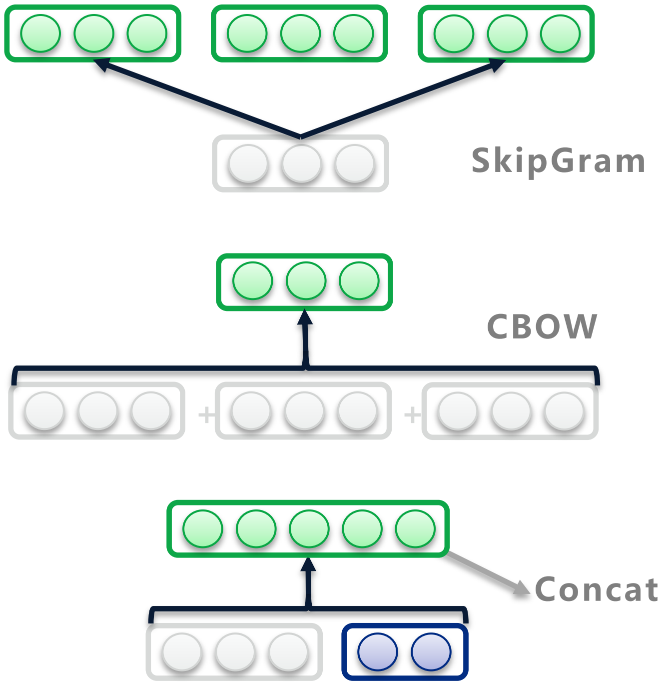
AnyEmbedding – A Unified Framework for NLP



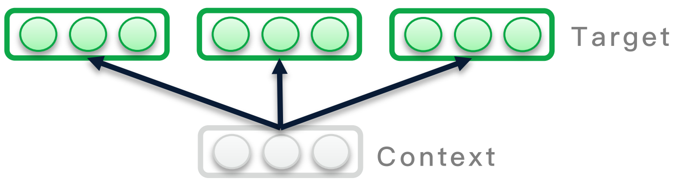
- Generate embeddings for any entities of interest
- A unified framework to tackle various NLP problems

AnyEmbedding – A Distributed Embedding System

- Goal: Design arbitrary models and express arbitrary entities via “connections”
  - word2vec, sent2vec, doc2vec, FastText, Hierarchical Document Vector model
  - TransE, TransR
  - node2vec
- Flexible mode of connections
  - 1-to-1：special case of 1-to-many
  - 1-to-many：SkipGram
  - Many-to-1：CBOW、 Concat
  - Many-to-many：extended many-to-1
- Supports both single-node and distributed computation
  - Scales up to 100k words & 1b users
  - Async distribution
- Optimization method
  - supports Negative Sampling（NCE loss）



SkipGram Negative Sampling (SGNS)



SkipGram

V : Vocab Dict
B_h : Set of vocab in a batch
σ : sigmoid function
$w_{i,jb}$: words in B_h
$b_{i,j}$: sliding window size
$w_{i,k}$: words in the context of $w_{i,j}$
$\hat{w}$: negative sampling for $w_{i,j}$
$N_{i,j,k}$: set of all negative samples
u : input embedding
v : output embedding

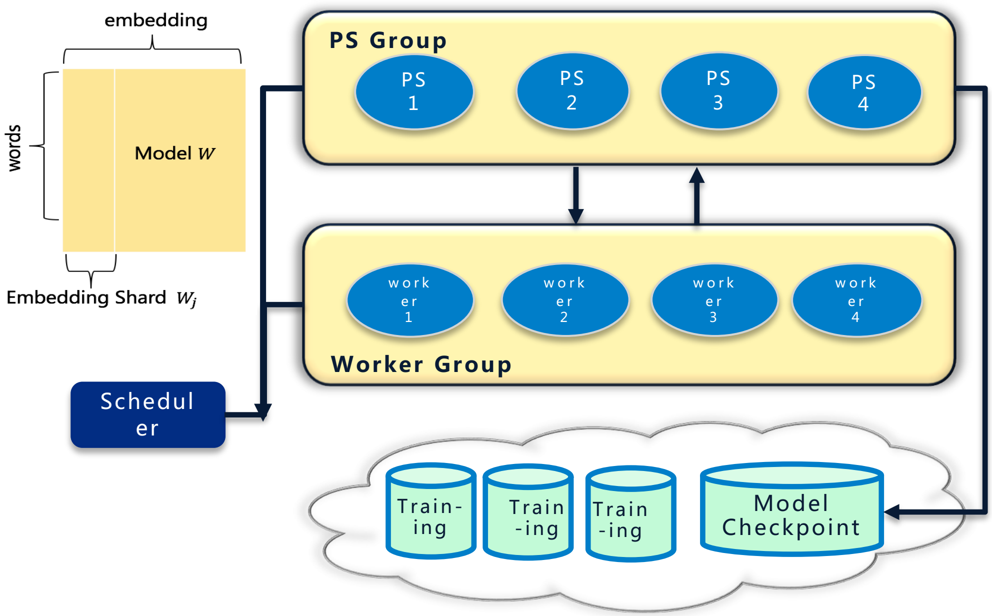
- Objective Function

$$L = \sum_{i,j\in B_h} \sum_{\substack{k\neq j:|k-j|\le b_{i,j} \\ w_{i,k}\in V}} [log \ \underbrace{\sigma(u(w_{i,j})v^T(w_{i,k}))}_{prob\ of\ predicting\ w_{i,k}\ as\ pos} + \sum_{\hat{w}\in N_{i,j,k}} log(1 - \underbrace{\sigma(u(w_{i,j})v^T(\hat{w}))}_{prob\ of\ predicting\ \hat{w}\ as\ neg})]$$

- Gradient

$$\nabla L|u(w_{io,jo}) = \sum_{\substack{(i,j)\in B_h: \\ w_{i,j}=w_{io,jo}}} \sum_{\substack{k\neq j,|k-j|\le b_{i,j}, \\ w_{i,k}\in V}} [\underbrace{(1-\sigma(\overbrace{u(w_{io,jo})v^T(w_{i,k})}^{dot\ \ product}))v(w_{i,k})}_{linear\ \ combination} - \sum_{\hat{w}\in N_{i,j,k}} \underbrace{\sigma(\overbrace{u(w_{io,jo})v^T(\hat{w})}^{dot\ \ product})v(\hat{w})}_{linear\ \ combination}]$$

Distributed Computation



- Parameter Server Group
  - Each PS stores embedding shards
  - Receives minibatches from Worker Group, calculates the dot product with the context, and sends back to Worker Group
  - Receives the result of linear combination from Worker Group and update parameters
- Worker Group
  - Loads minibatches, and sends to PS
  - Receives computed dot products from the PS Group, computes the linear combination and sends to PS
- Network transmission
  - Minibatch data, integer vectors
  - Dot product, Linear combination, real-valued scalars

Ordentlich, E.; Yang, L.; Feng, A.; Cnudde, P.; Grbovic, M.; Djuric, N.; Radosavljevic, V.; and Owens, G. Network-efficient distributed word2vec training system for large vocabularies. CIKM 2016.

Stergios Stergiou, Zygimantas Straznickas, Rolina Wu, and Kostas Tsioutsiouliklis. Distributed Negative Sampling for Word Embeddings. AAAI 2017.

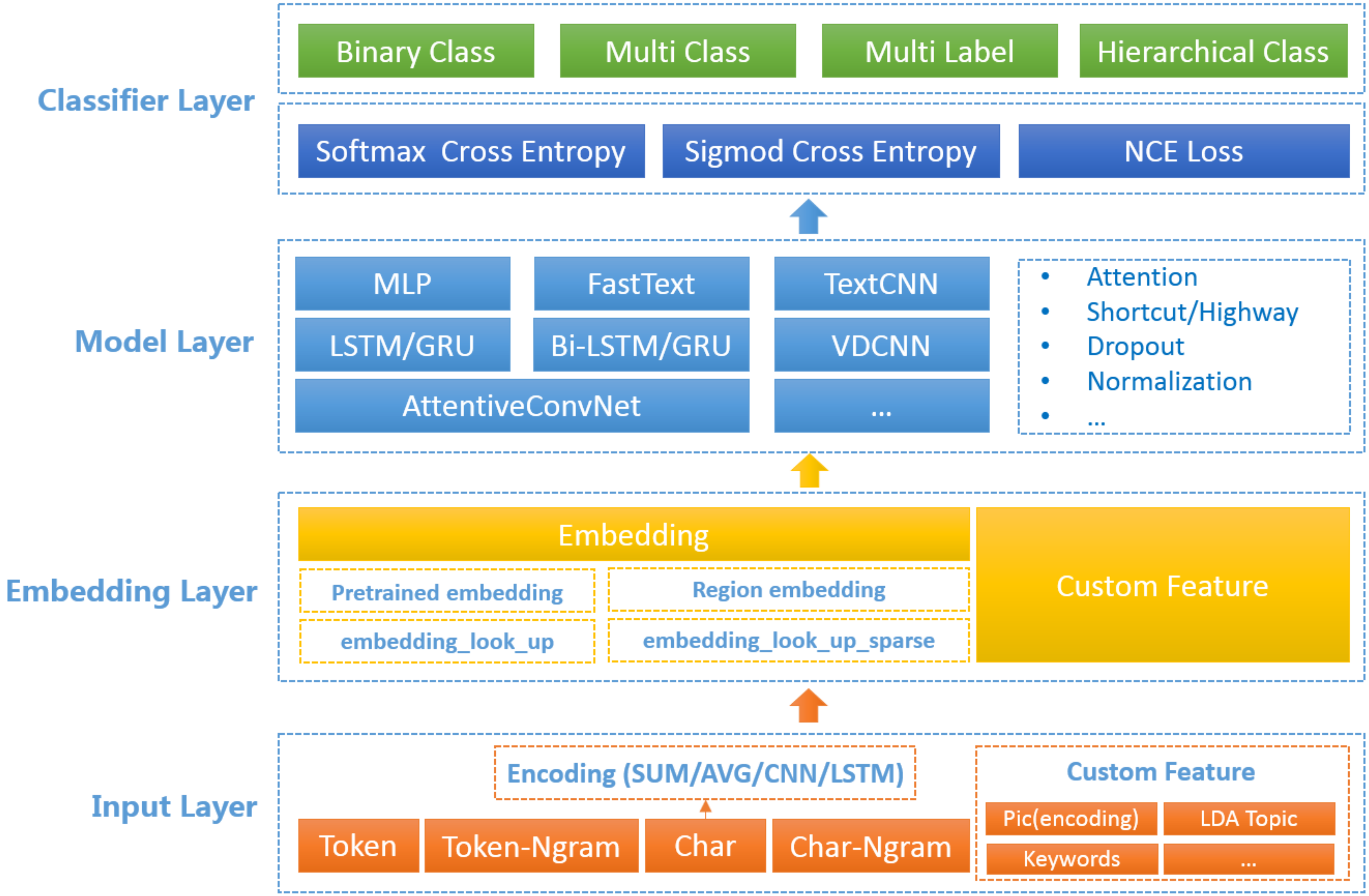
DeepText: An Open Platform for Deep NLP

Type	Description	Model	Common methods	Example tasks
Text Classification	Classify input texts to predefined categories	$s \rightarrow c$ s: string, c: label	SVM , MaxEnt , TextCNN , TextRNN	Text classification, sentiment analysis, intent understanding
Structural prediction	Derive structured information from free text	$s \rightarrow [s]$ s: string, [s]: structure	HMM , MEMM , CRF , Structure Perceptron , LSTM-CRF , CNN-LSTM-CRF , Transition-based	Chinese segmentation, PoS tagging, NER, syntax analysis
Text Matching	Match texts based on various criteria	$s, t \rightarrow R^+$ s, t: string, R^+: non − negative real values	RankNet , GBRank , LambdaMART , DSSM , CDSSM , MatchPyramid	Information retrieval, Search engine, QA bot
Text Transformation	Convert a input sequence to an output sequence	$s \rightarrow t$ s, t: string	SMT , NMT (Seq2seq+Attention)	Machine translation, QA bot, text generation

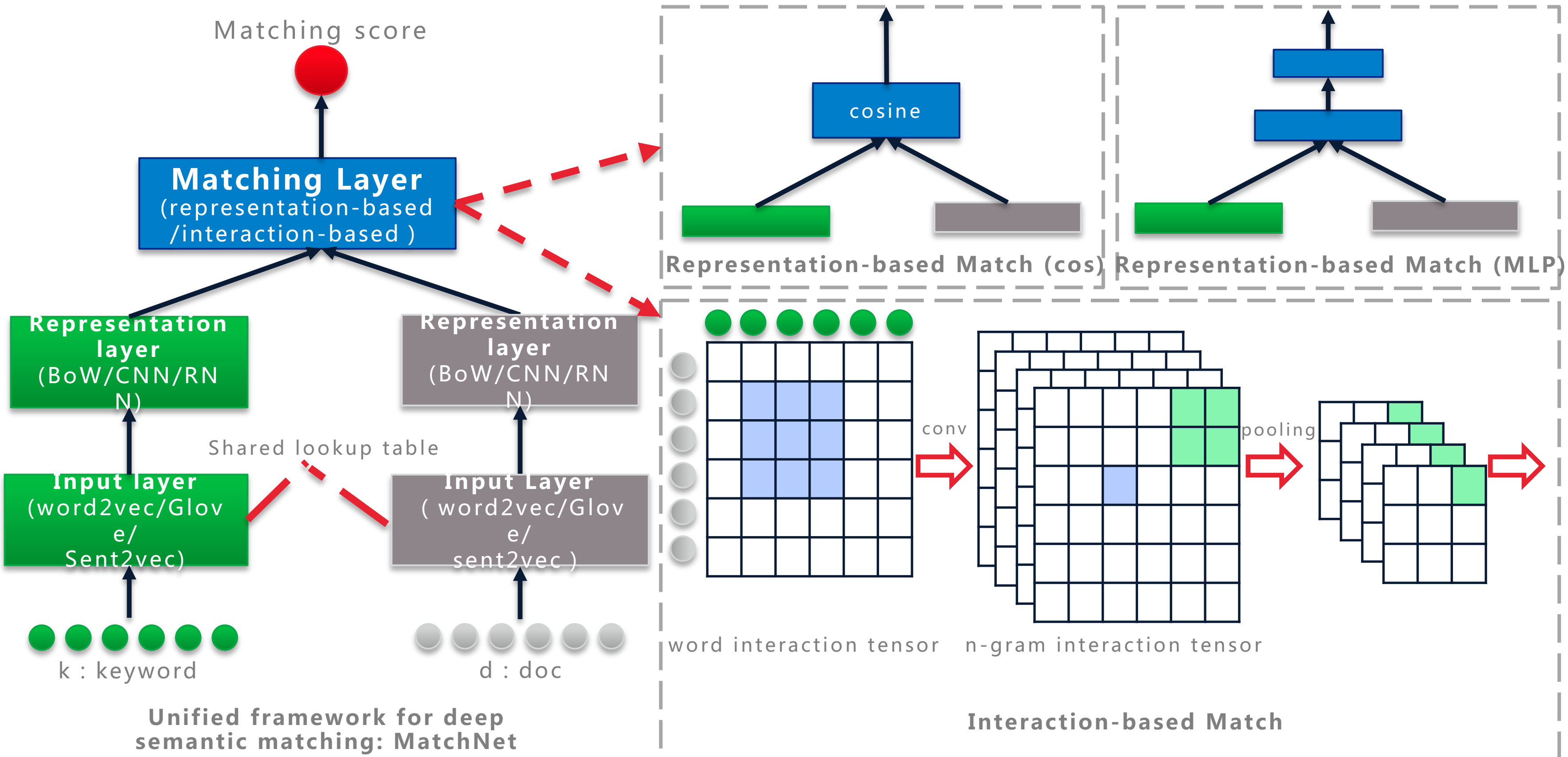
DeepText

Goal: To develop an open NLP experiment platform to tackle the problem of text classification, tagging, and matching, and support a wide range of models and optimization methods proposed in the field for fast proof of concept and experimentation

DeepText-Classification: a Wide Range of Supported Methods



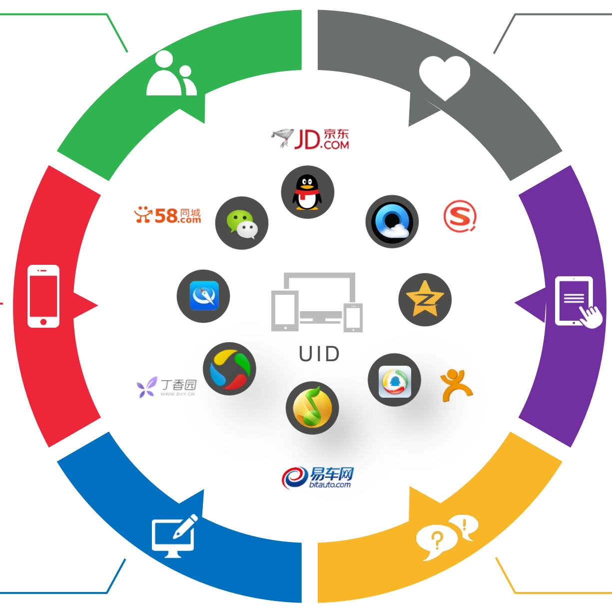
DeepText-Matching: a Variety of Flexible Matching Models



User Profile Understanding & Accurate Audience Insights

Demographics
- inherit
- location
- education
- marriage
- wealth
- employment

Interest
- purchasing intent
- hobby



Device
- carrier
- network
- brand
- type
- OS
- price

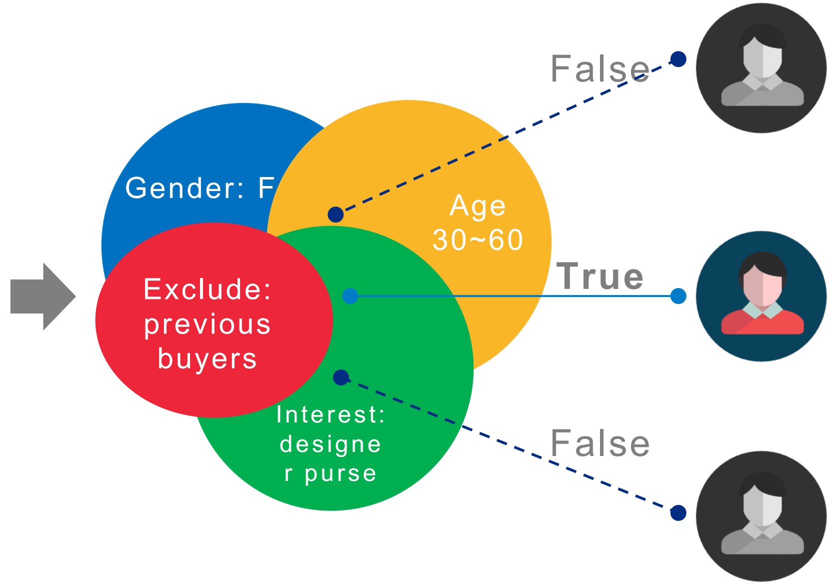
Behavior
- travel
- App
- eCommerce
- O2O
- Ad engagement
- purchase

Custom
- seed
- expansion
- 1-party labels

Vertical
- auto
- education
- eCommerce
- game
- travel
- 3C
- finance
- FMCG

Traditional Audience Definition - Manual Targeting

Gender, age, geo, education.	Interests, keywords
Marriage, employment, wealth, culture	Followers, APP installs, payments, engagements
Context, device, weather	APP, eCommerce, Search, and Social behaviors



Disadvantages:

- Domain knowledge
- Information loss
- Low granularity
- Manual optimization

Large number of targetable attributes

Advertisers manually specify the targeting criteria

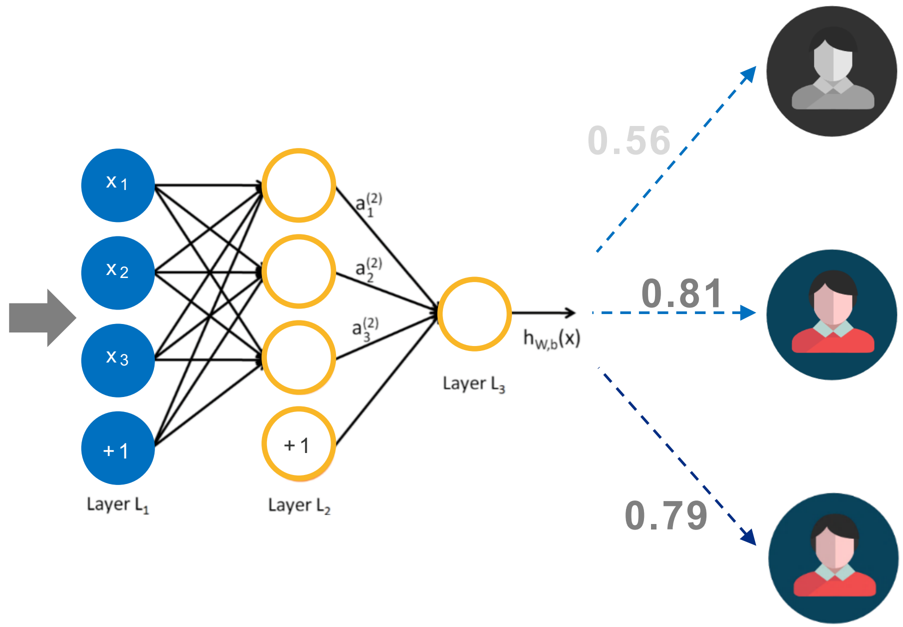
Boolean match during ads retrieval

Auto Targeting: Automatic Audience Identification by AI

existing leads from 1st-party data

Previously Converted users

User/Ad features



A change of mindset

- Targeting by user-ad relevance

Multiple ways to define the seed audience

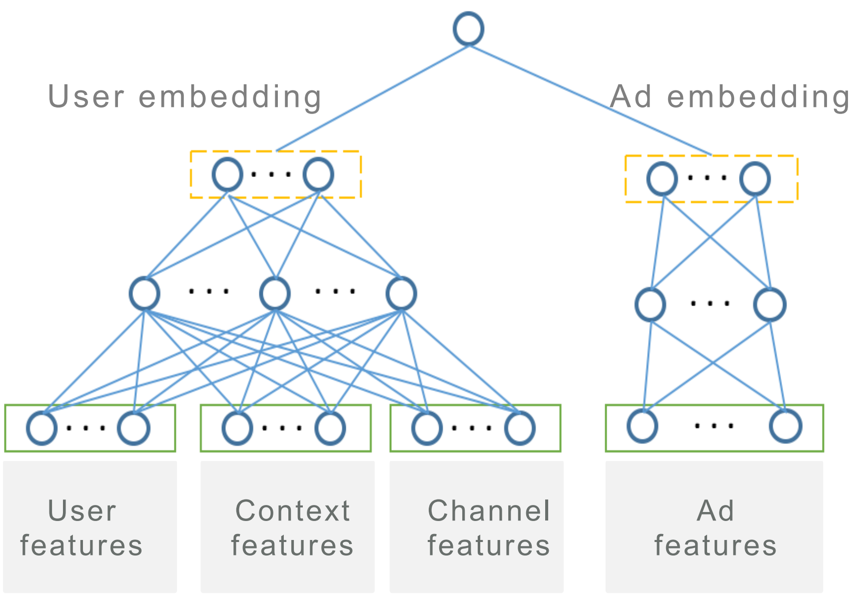
Targeting by machine learning models

Matching based on relevance

Auto-Targeting：Representation Learning & ANN

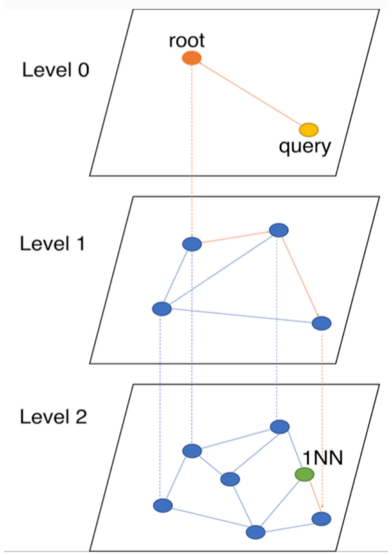
Representation learning

Learns embeddings for users and ads;
can be used during ad retrieval



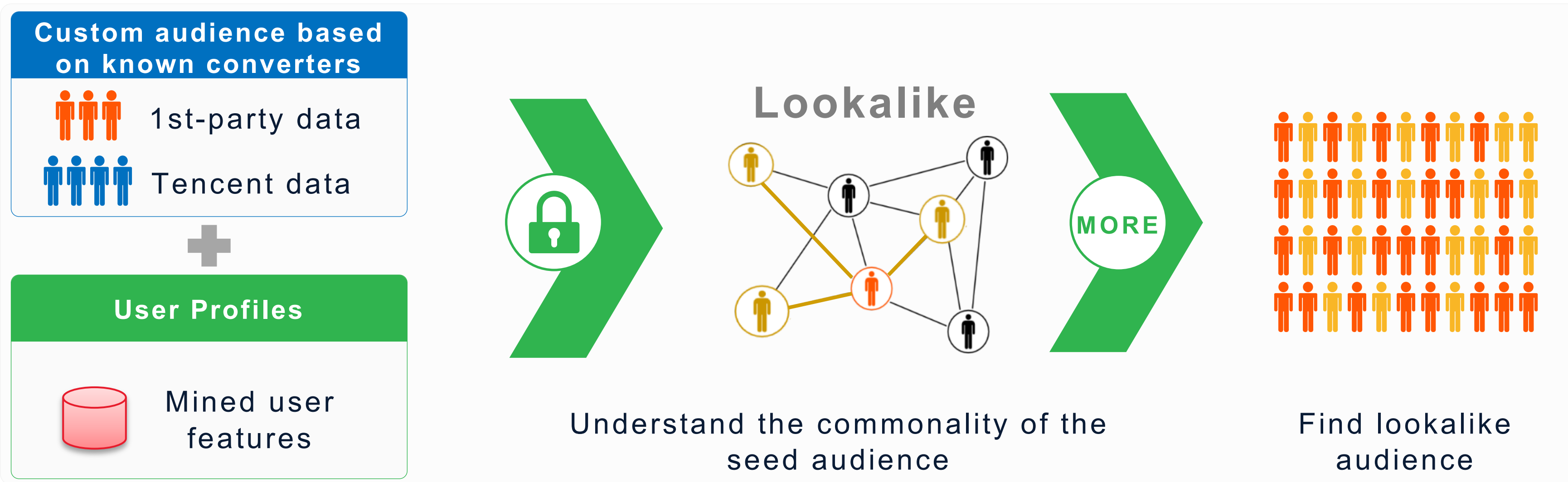
Approximate Nearest Neighbor Search

Efficient ad retrieval based on user/ad
embeddings under scalability constraints



"Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs", 2016

Lookalike Expansion for Targeting: Identify Potential New Customers



2013

Peacock
- Large scale distributed LDA

2014

Logistic Regression
- One model per seed audience, expansion done offline

2015

DNN
- Deep embedding for each seed audience, expansion done online

AI in Tencent Social Ads

Smart Targeting

Smart Creative

Smart Bidding

Manual Creative Review: Laborious, Inefficient, and Error-Prone

1. **Text** suggestive, vulgar, over the top, low effort
2. **Image** blurry、deformed、IP violation、% of text
3. **Landing page** relevance to the creative content
4. **Official Account** title, description
5. **Published content** history, compliance

The manual review procedure consists of
Hundreds of individual rules

More than 4x increase of the review request in 1 year

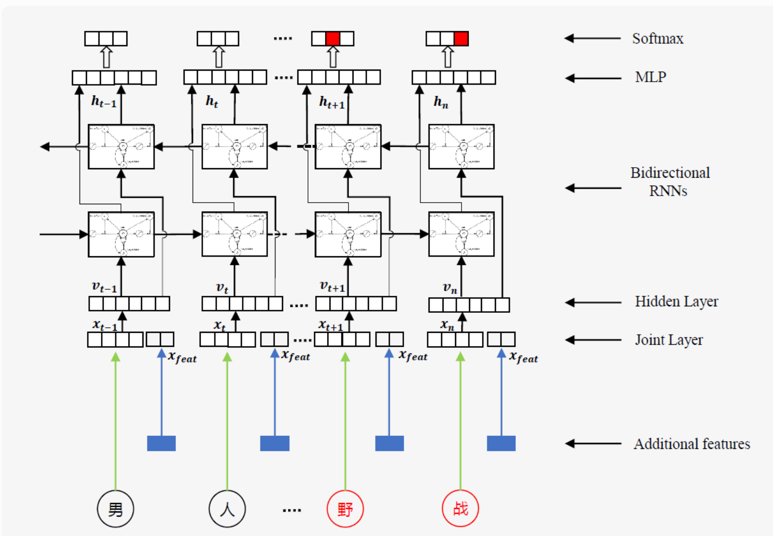
1. **hard to scale with human labor**
   Extremely high volume, Repeated ad submission
2. **Steep learning curve, prone to inconsistency**
   Complicated set of rules, high training cost,
   implicit knowledge that is hard to be passed down
3. **Some cases are hard for human**
   Celebrity/IP/Plagiarism detection

Low-Quality Content Detection based on Texts or Images Individually

① Text-based detection

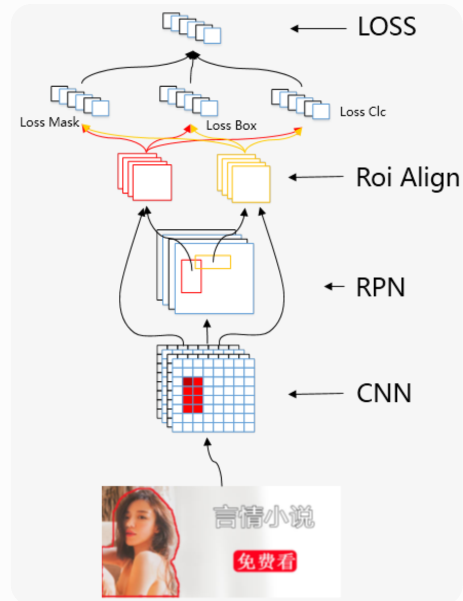
"是男人就来**野战**一场，一小时的**持久**战斗！"

Be a man and do a field battle, an hour-long game of stamina



Review result: Reject
Reason: sexual innuendo

② Image-based detection



Goal is not to reject

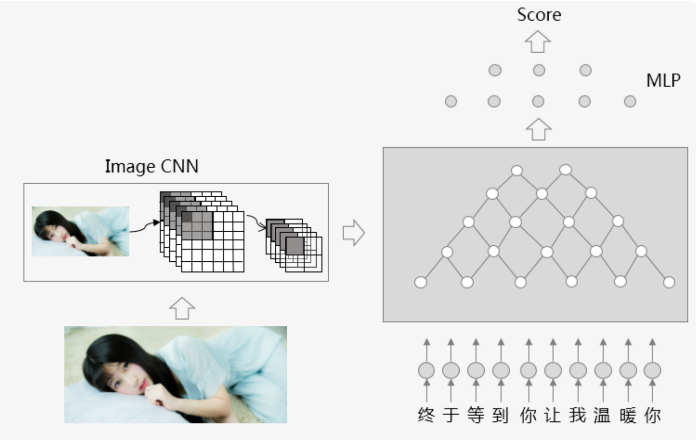
...but to help advertisers improve

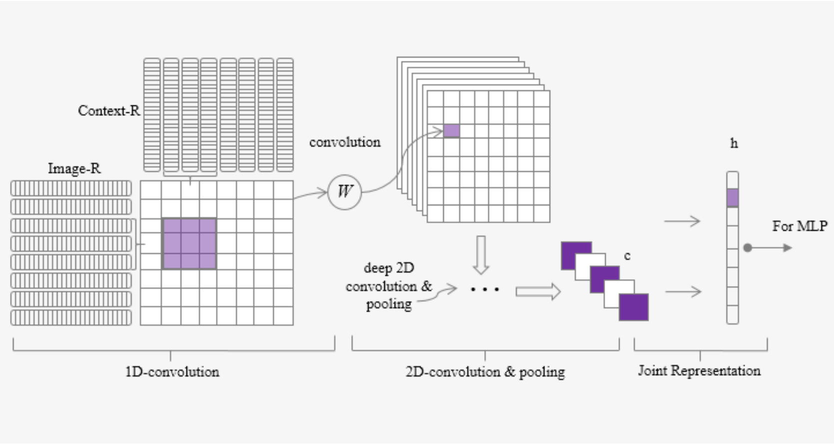
Review result: Reject
Reason: explicit/suggestive content

Low-Quality Content Detection based on Joint Image/Text model

3 Joint image/text models

"Finally you are here.
Let me give you some warmth."





10+ similar models

Expert System for Rejection

① Expert system to reason about the cause of rejection

LCS = 0.85

Expert System

Rule engine → Template engine

Rule DB | Cause DB

Rule Configuration System

LCS <= 0.7

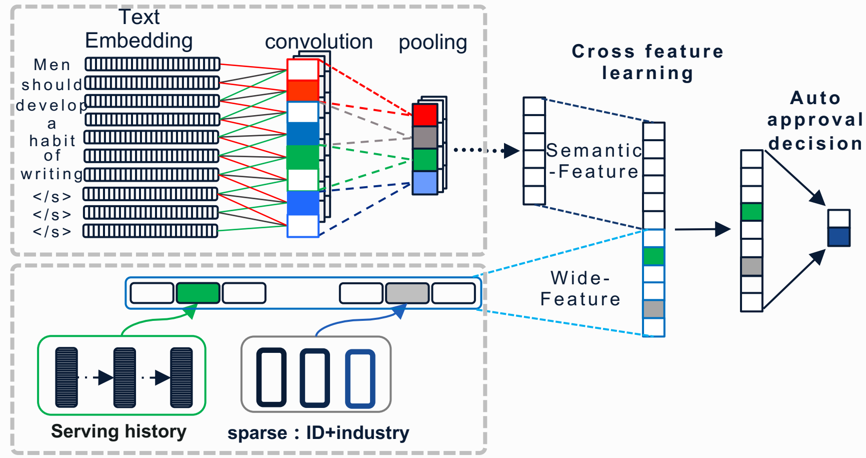
Review result: Rejection

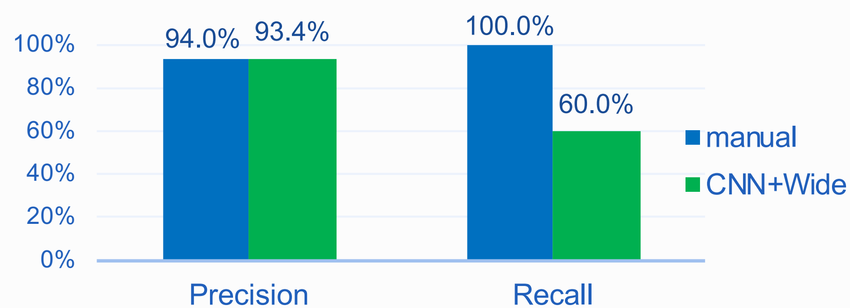
Cause of rejection:

suggestive content in context

If no applicable rule

② Auto approval

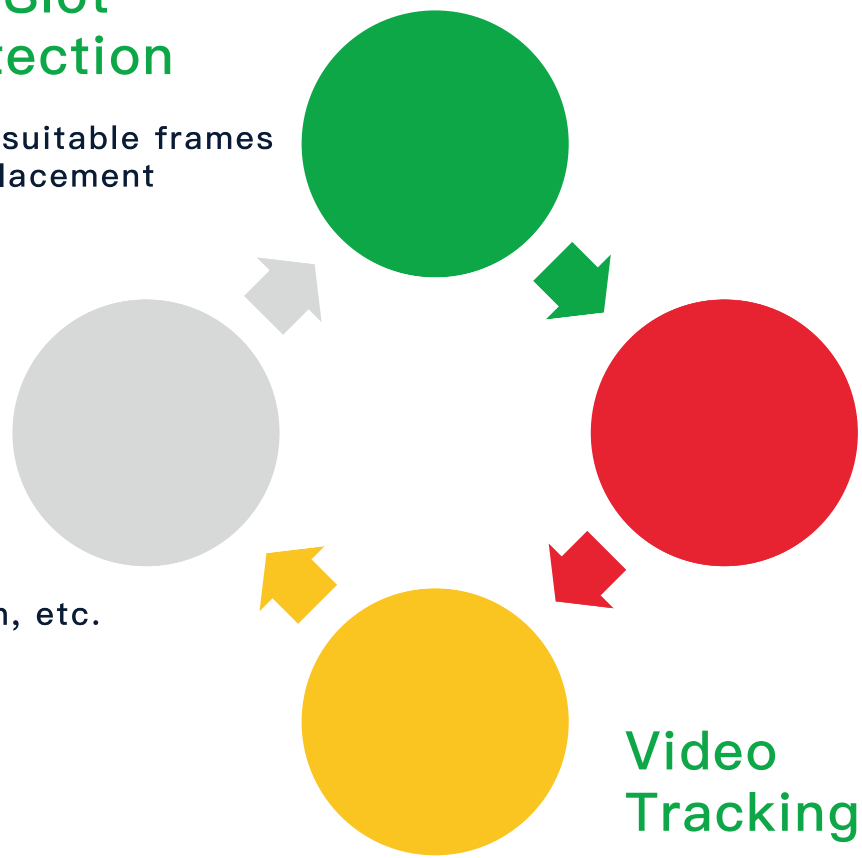




VideoIn Ads: Smart Product Placement

Ad Slot Detection

Find suitable frames for placement



Object & Asset Orientation

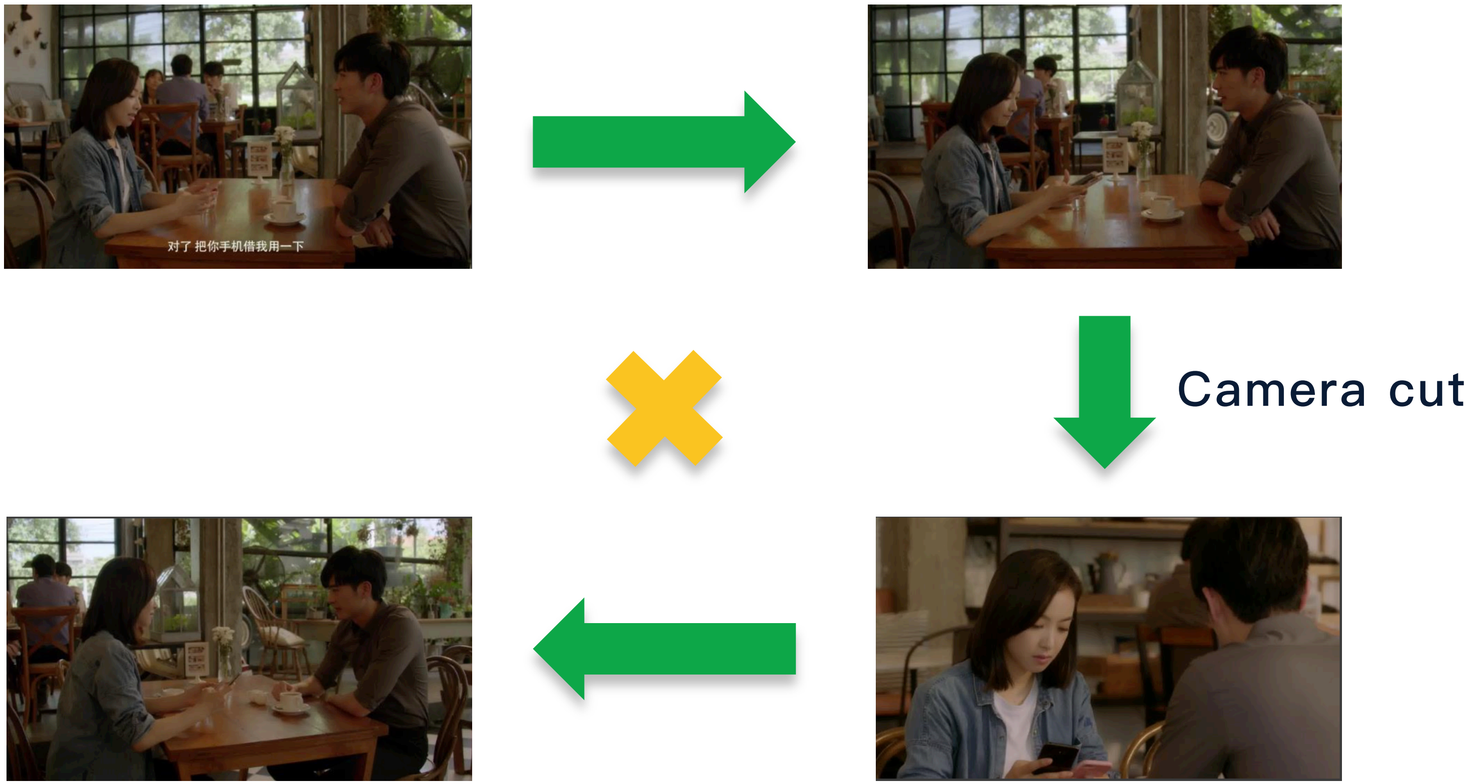
Adjust the perspective of 3D models and assets

Video Tracking

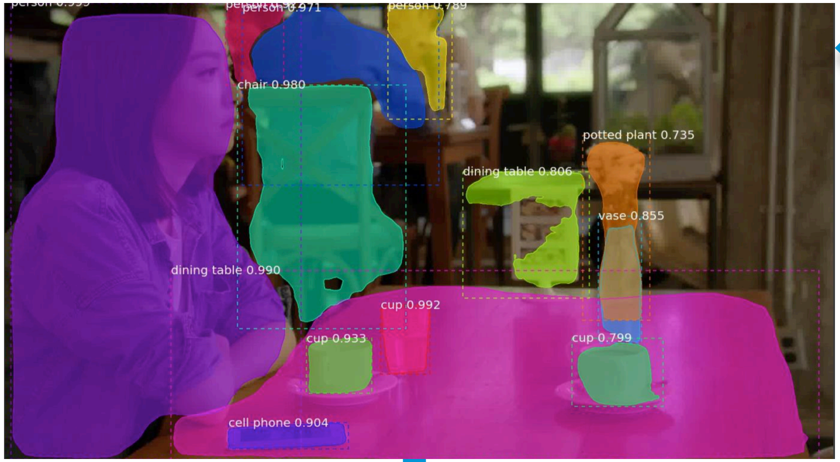
Position the rendered creative in the frames

Ad Creative Rendering

Render the asset with the ambient brightness, saturation, etc.



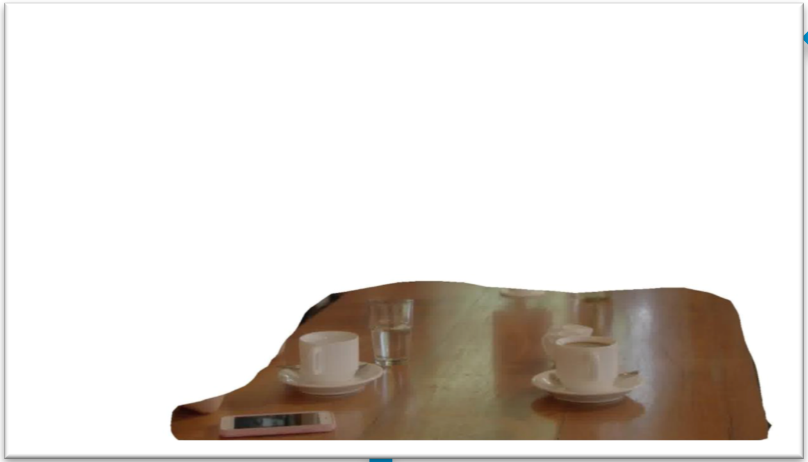
Exclude unsuitable sections by detecting camera cuts from a grouping of key frames



Detection of the first frame of a potential ad slot

Non-central object masking algorithm based on object segmentation with mask R-CNN

Potential Ad Area Location



Detection of the first frame
of a potential ad slot

Color clustering and bounding box
search to detect empty table surface

Potential Ad Area Location

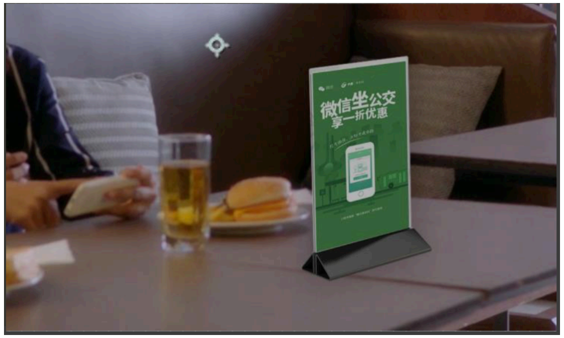
Picture frame detection and tracking

Object recognition and segmentation
Based on DaSiamRPN

In-house labeling of 17
classes of video contents

Typical Ad Format

User supplied pictures

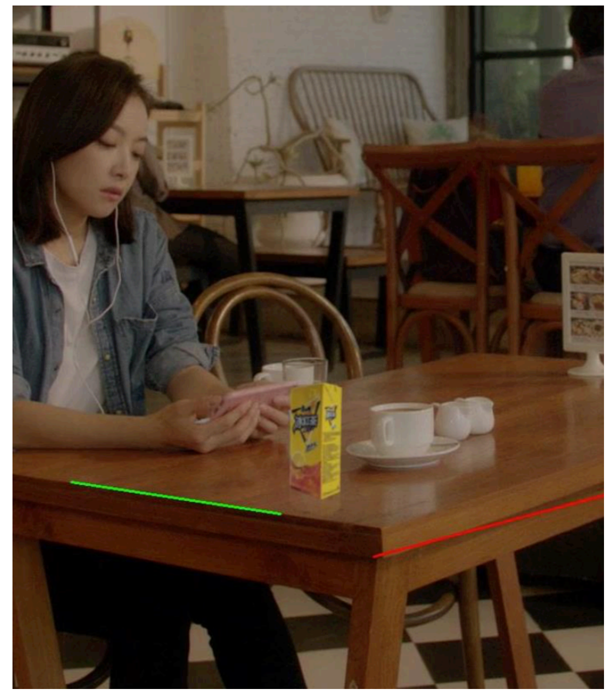


User supplied pictures
&
platform supplied 3D assets

User supplied 360-degree pictures

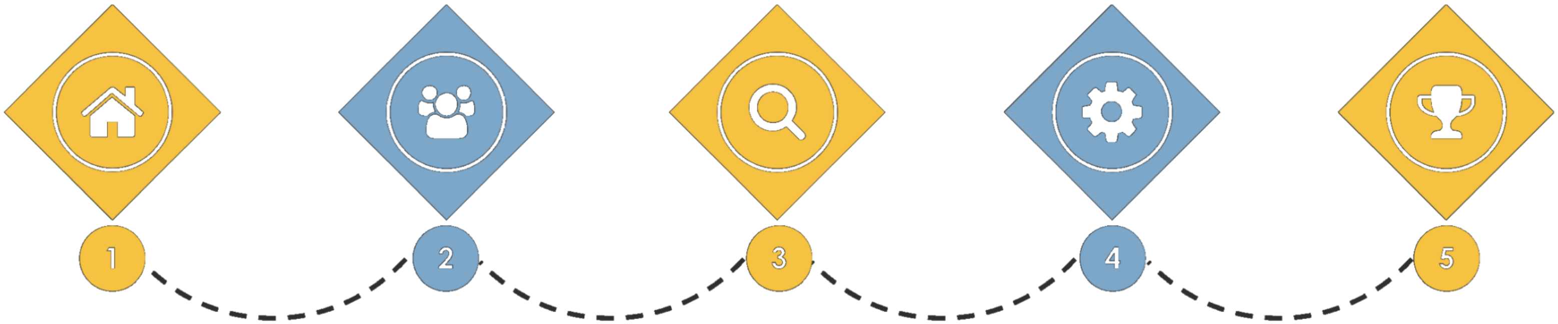
Detecting the bottom edge of the table corner

Matching the bottom edge of the object



In: properly oriented stock picture

Out: adjusted picture with background harmonization



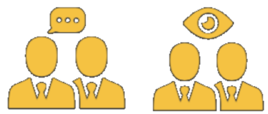
Original

Brightness

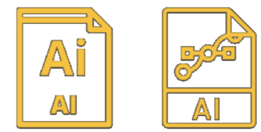
Saturation

Blurriness

Shadow



Pre–adjustment



Post–adjustment


青檀姐


纯牛奶


腾讯视频 独播


腾讯视频 独播
Cadillac
DARE GREATLY


腾讯视频
对


腾讯视频
HUMAN MADE
对

Static Creative Is Hard To Optimize For Advertisers with A Large Inventory

User

Impression

Serving

Creation

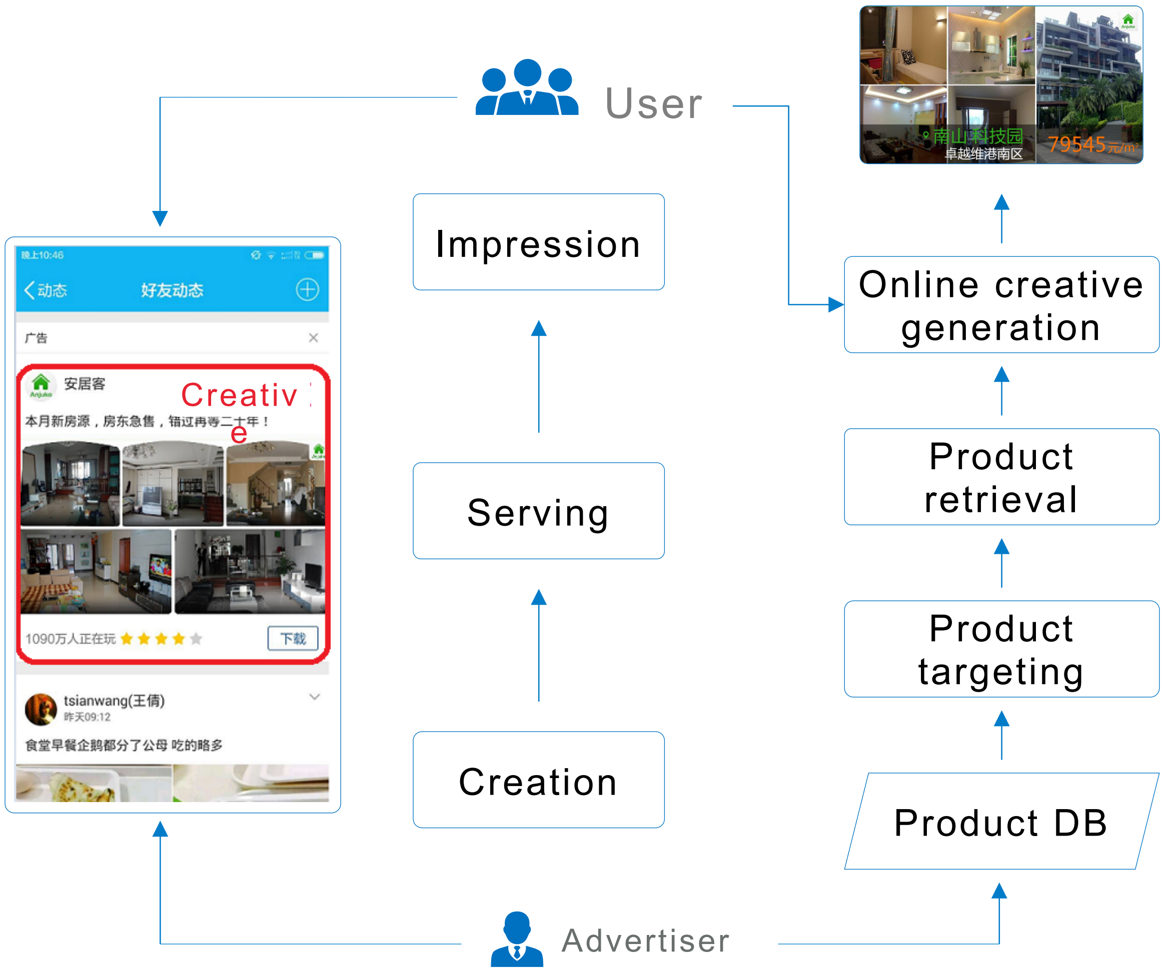
Advertiser

Static Creative :

Challenges

- Hit-or-miss
- Coarse-grained targeting
- Static content with no personalization

DPA: Automatic creation, retrieval, and rendering of personalized Ad Creatives



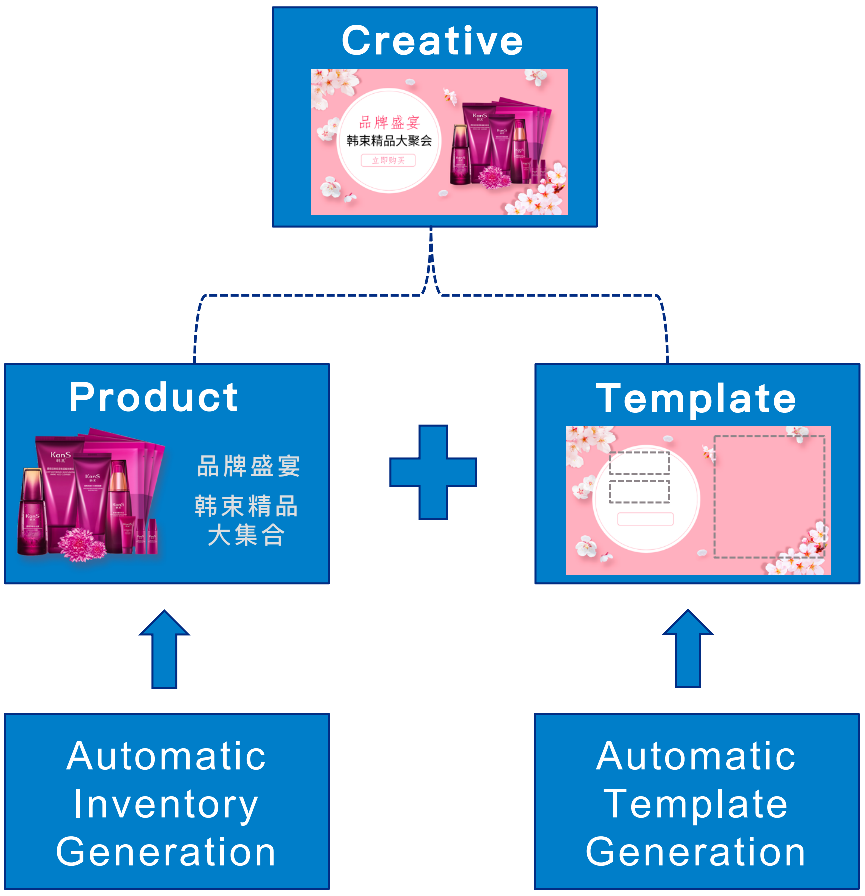
Advantages

- **Automatic creative generation**
  - No need for trial-and-error optimization from advertisers
- **Fine-grained targeting**
  - 100m targetable products
  - More detailed and accurate than ad targeting
- **Truly personalized**
  - Different experience even for the same ad

Anatomy of a Creative

- **Efficiency**： Automatic creative generation based on product + template anatomy
- Product： Programmatic interface to sync from advertisers’ inventory
- Template： Coverage over more than 3000 ad slots, automatically generated by AI
- **Performance**： optimize

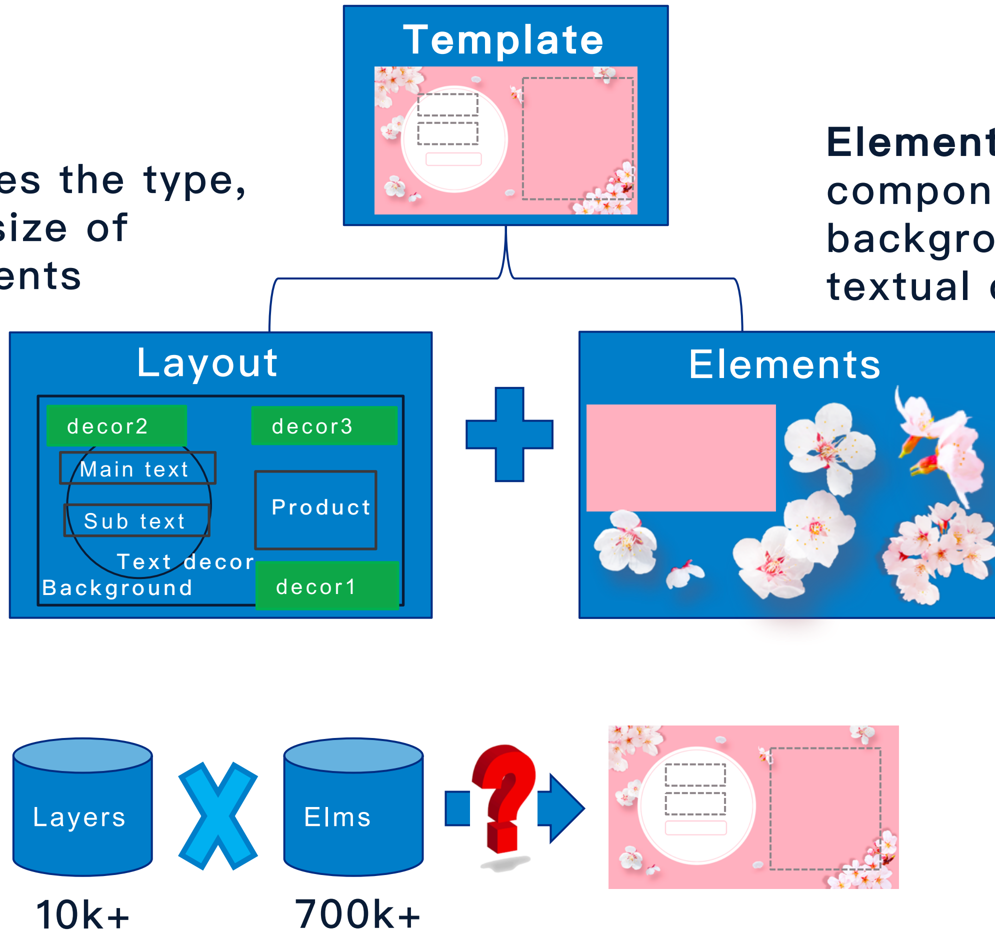
$$argmax\left\{\begin{matrix} pCTR(\text{Product}, \\ Template\ |User) \end{matrix}\right\}$$



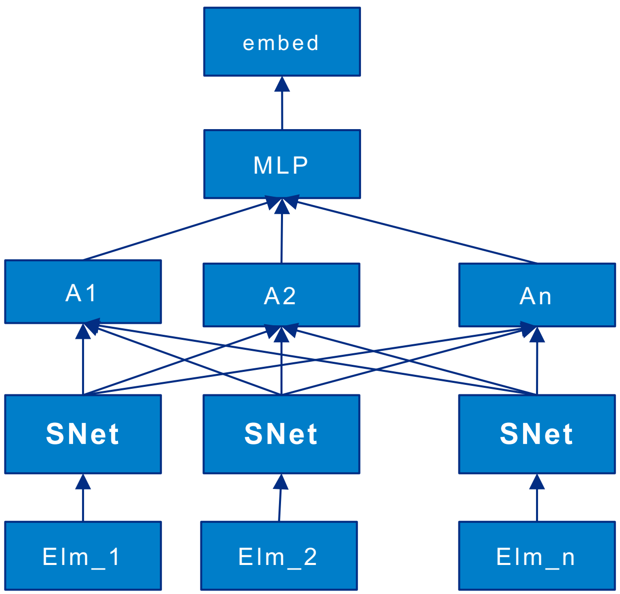
智能生成引擎之模板生成

Layout：Defines the type, position, and size of template elements

Element：Individual graphic components，e.g., background, decoration, textual decorations



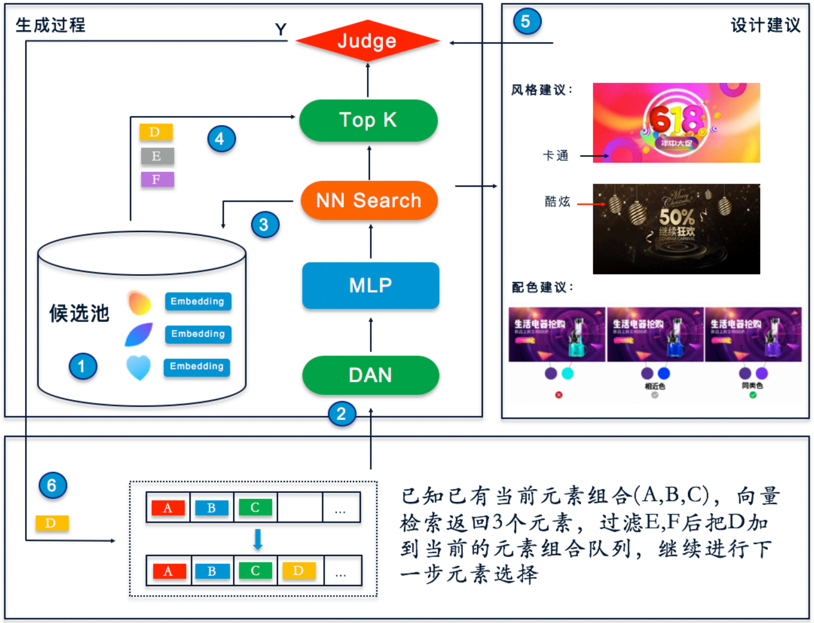
Template Generation：Embedding for Element Combination



- Treat element selection as a combinatorial optimization
- Map elements through a Deep Attention Network

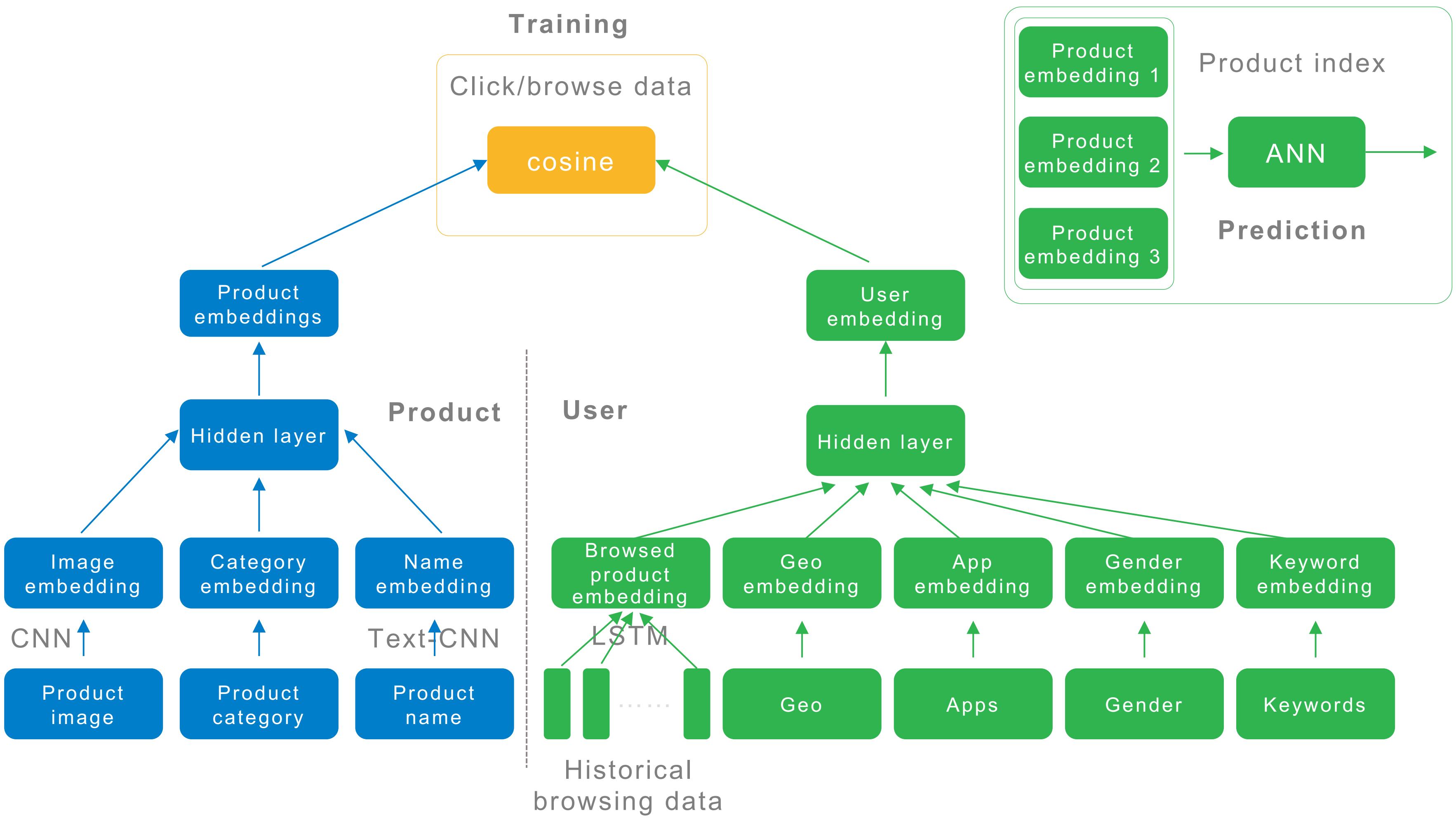
Template Generation

✓ Generation with design constraints; ML + Expert System

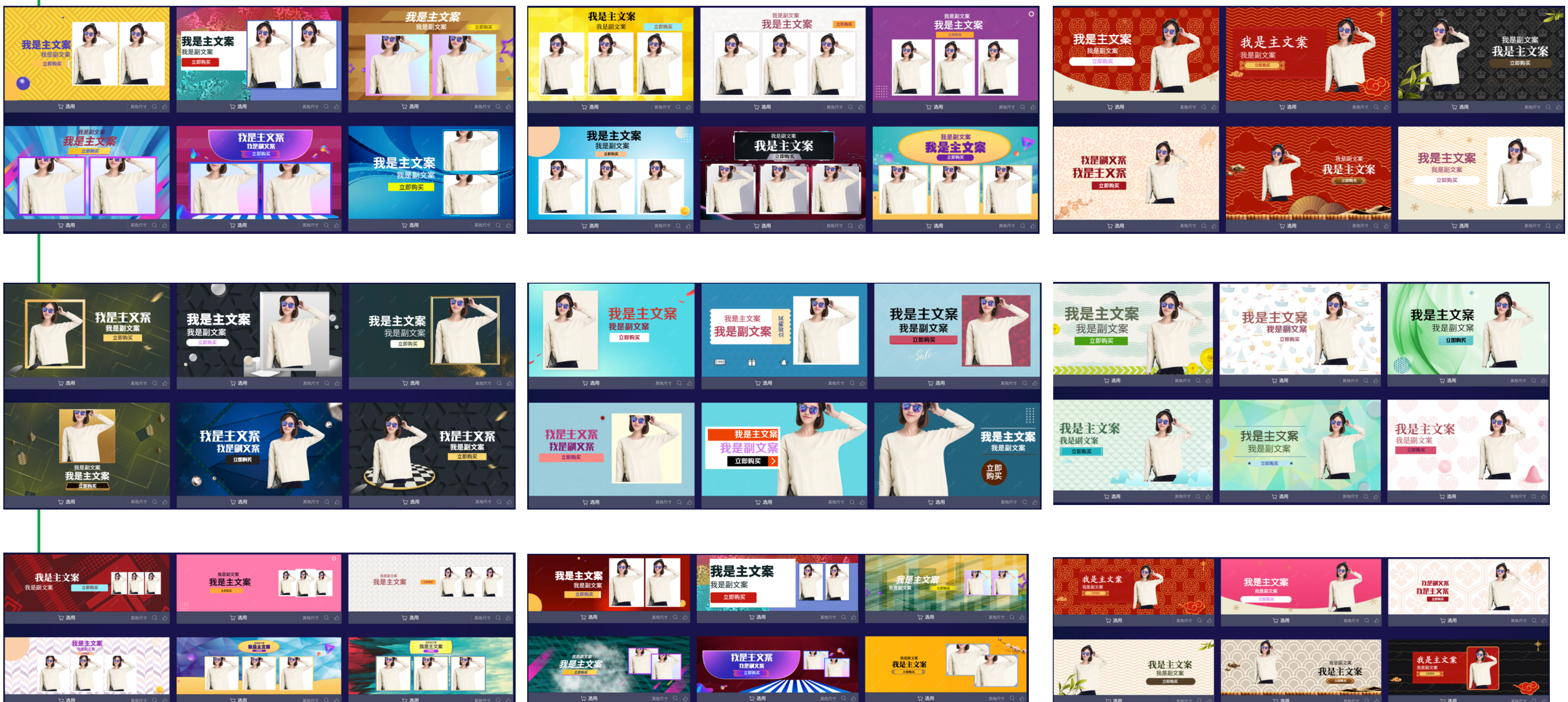


Result: Generated 5.5M+ templates. Manual evaluation from sampling shows 73% achieve designer grade

Product Retrieval based on Learned User and Product Representations



Demo of Generated Templates



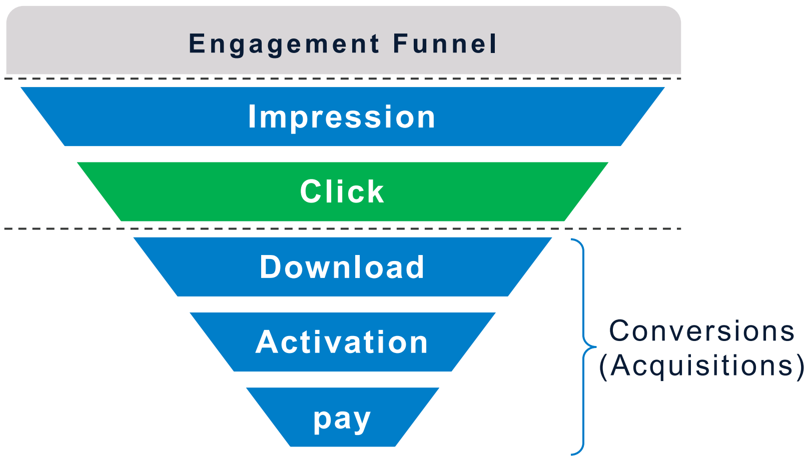
Machine Learning Applications

Smart Targeting

Smart Creative

Smart Bidding

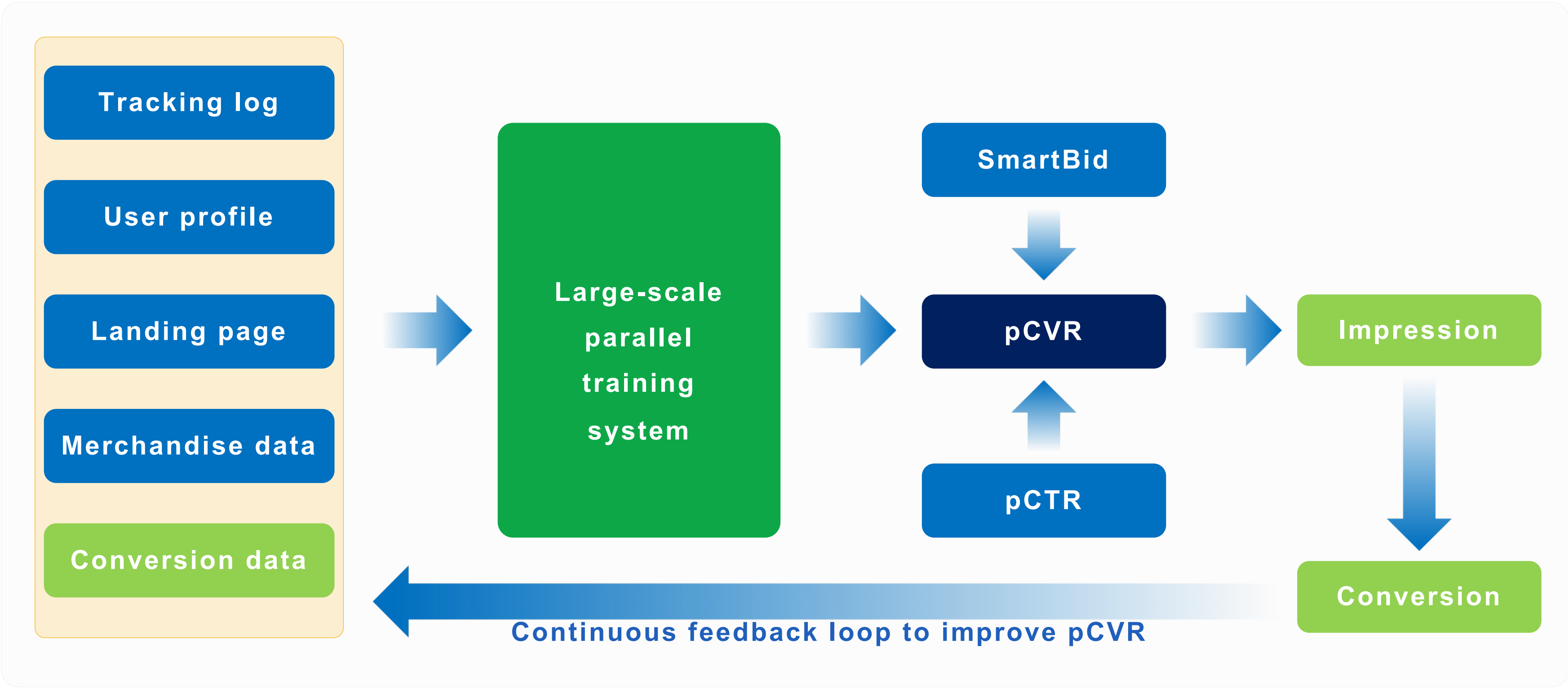
Optimized Cost Per Acquisition (oCPA) Bidding



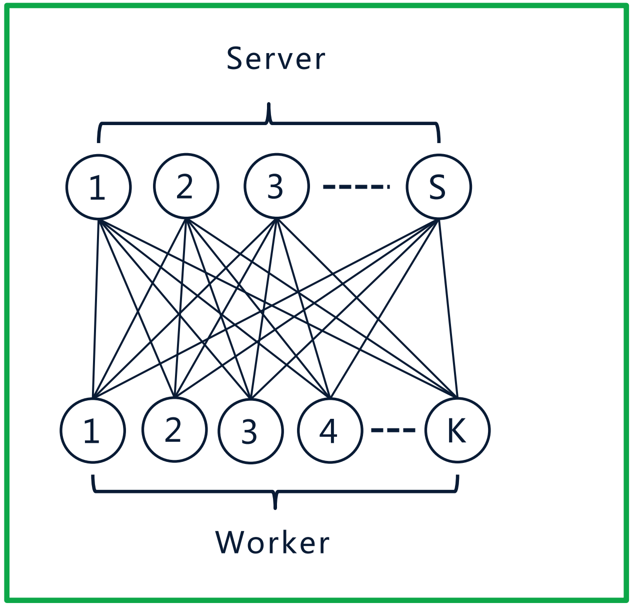
- oCPA lets the advertisers specify a target conversion type to optimize
- Advertisers specify the target cost per conversion
- The system adjust the bid on behalf of advertisers based on the predicted conversion rate (pCVR)

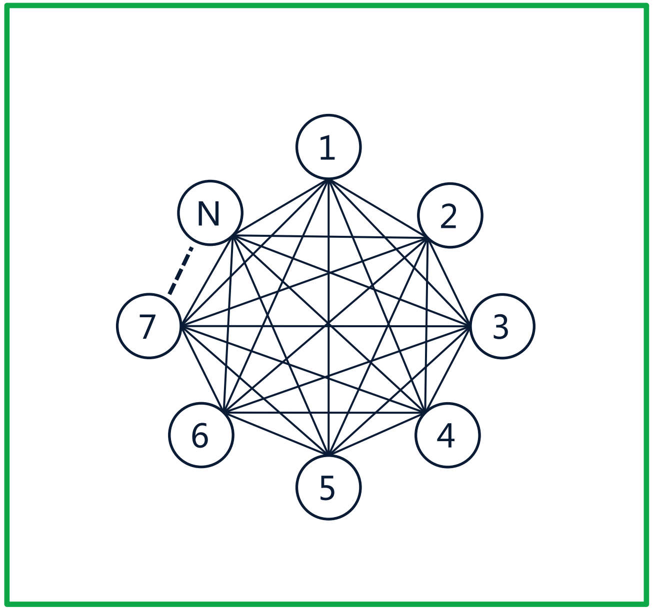
Accurate prediction of pCVR is the key to acquiring the most conversions within cost and budget

Predicting CVR Accurately is the Key to oCPA Bidding



Large-Scale Parallel Training Framework





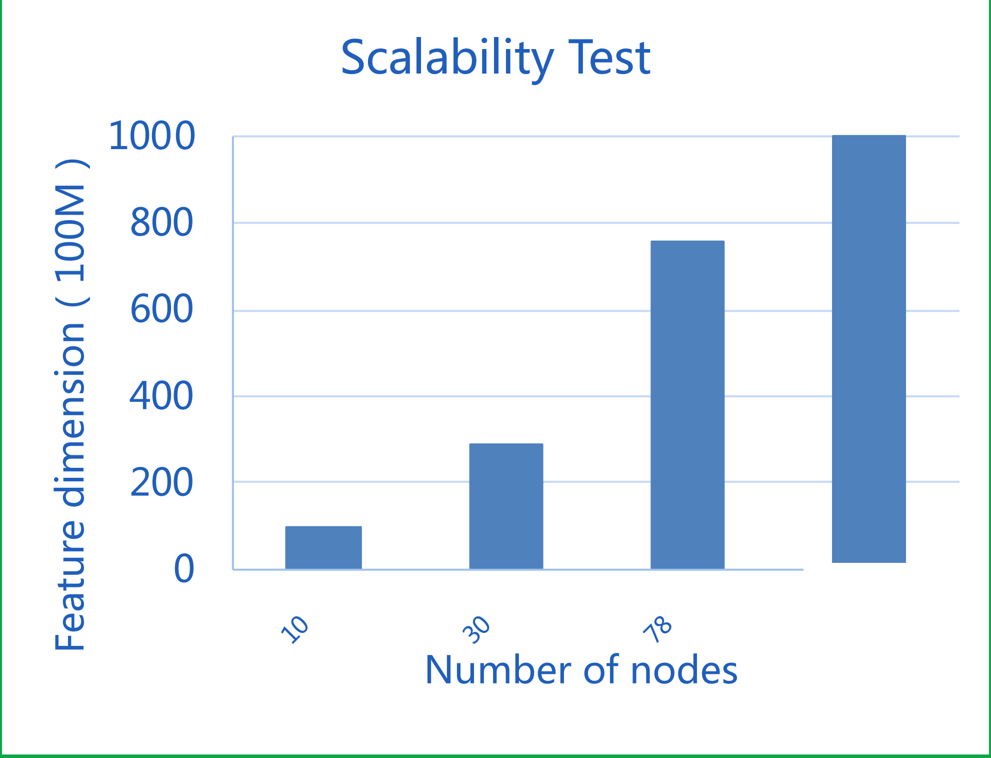
P2P topology

Parameter servers and workers are collocated

Balance different load profiles of the two roles

Scalability Test: Processing Power

Scaling with added nodes



Model configuration

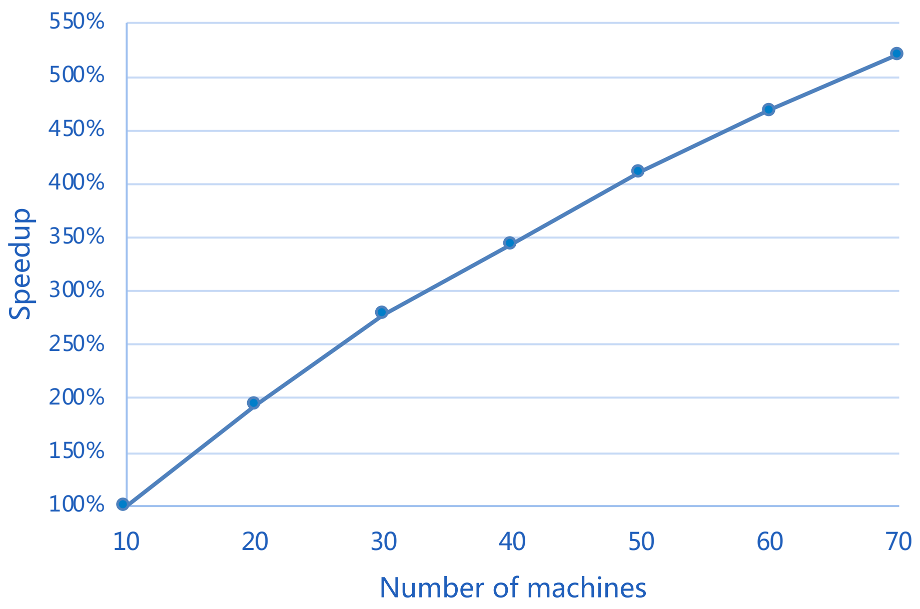
- ❑ Feature dimension : 5
- ❑ Input : 890
- ❑ 2 hidden layers : 135、67 nodes
- ❑ output : 1 node

Observation

- ❑ Linear increase in processing power wrt number of machines
- ❑ Successfully trained on 100B features

Scalability Test: Speedup

Speedup vs number of nodes

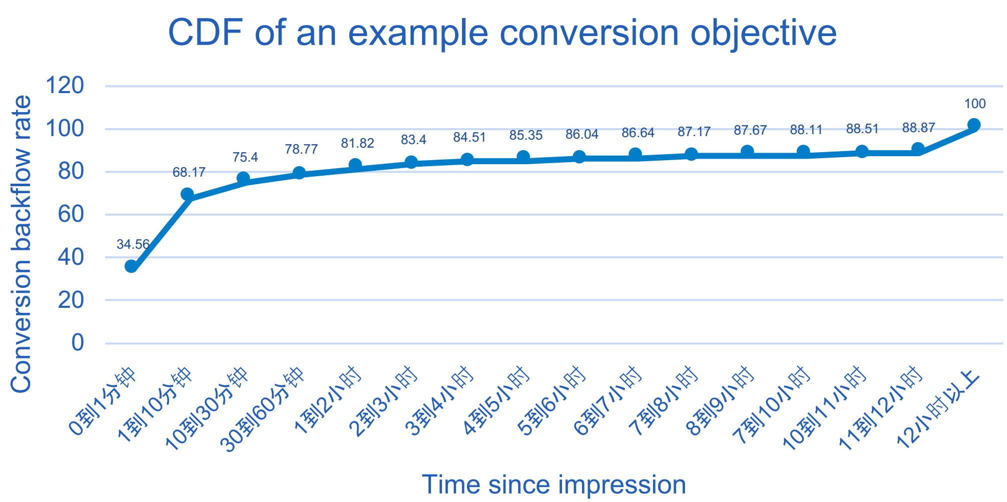


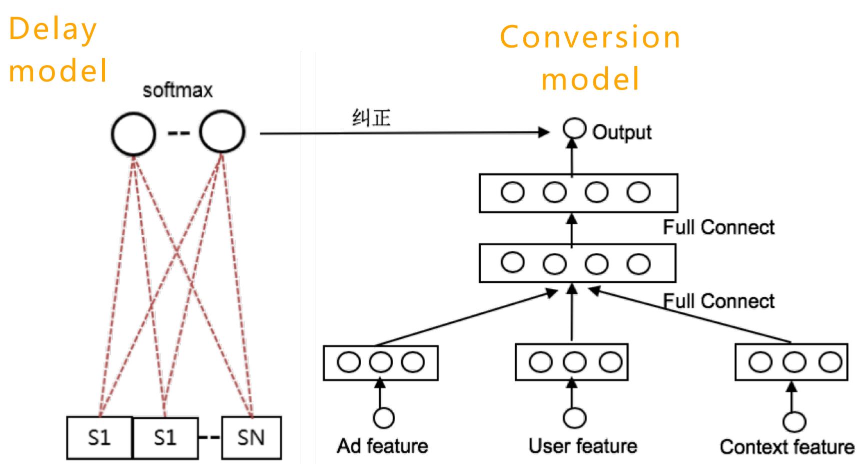
Model structure :
- ❑ Feature vector dim : 11
- ❑ Input : 110
- ❑ 5 hidden layers : 400、400、300、200、100
- ❑ Output : 1

Data :
- ❑ Training example : 6B
- ❑ Feature dimension : 4.8B

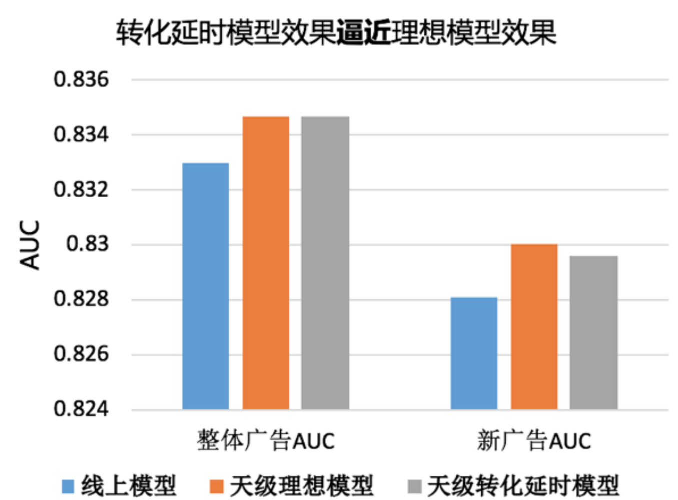
Observation : Linear acceleration wrt the number of machines

Deep conversion objectives have a very long delay





Joint training of conversion and delay model



腾讯社交广告
Tencent Social Ads

The Power to
Connect Businesses and People
赋能商业｜始终于人

Thank you!