

# Aprendizaje no supervisado

## TC02: Despedida de la asignatura

Félix José Fuentes Hurtado

[felixjose.fuentes@campusviu.es](mailto:felixjose.fuentes@campusviu.es)

Universidad Internacional de Valencia

18 de marzo de 2021

# Índice

1. Resumen asignatura
2. Fechas importantes

# Resumen de la asignatura

## 1. Introducción

- Introducción a medidas de distancia y métricas de evaluación

## 2. Clustering

- Basado en particiones (k-means, k-medoids)
- Jerárquico (aglomerativo y divisivo)
- Espectral (grafos, laplaciana, eigenvectors, k-means)
- Basado en densidad (DBSCAN, mean-shift, Affinity Propagation)
- Basado en modelos probabilísticos (Gaussian Mixture Models)

## 3. Aprendizaje semi-supervisado

- Naive Bayes con E-M

## 4. Análisis de componentes

- Análisis de Componentes Principales
- Análisis de Componentes Independientes

## 5. Análisis de grafos y transacciones

- PageRank
- APRIORI

# Evaluación

## Evaluación continua

- E1-E4: notebooks de los bloques (10% cada uno)
- Cuestionarios (20% en total)
- Foro (5%)

## Evaluación final

- E5: Trabajo final (35%)

**¡Es necesario entregar y aprobar todas las actividades para aprobar!**

Esto incluye las actividades E1, E2, E3, E4 y el trabajo final (E5).

- Si no se entregan más de 2 actividades E1-5 → NP
- Si se entregan más de 2 actividades E1-4 pero no el trabajo final → min(4, nota)
- Si se entrega todo pero hay alguna E1-5 suspendida y aún así media > 5 → 4

# Trabajo final

- Trabajo de investigación a desarrollar de forma individual
- Implementar y mejorar 1 algoritmo a vuestra elección del bloque de clustering
- Validarlo con 2 datasets distintos:
  - Uno con agrupamiento conocido (calcular métricas extrínsecas e intrínsecas)
  - Otro sin agrupamiento conocido (calcular métricas intrínsecas)

Para aprobar será necesario realizar todo lo mencionado

Entrega: memoria y código (notebook)

Deadline: **11 de abril de 2021, 23:59 CET.**

# Calendario

FEBRERO						
Lunes	Martes	Miércoles	Jueves	Viernes	S	D
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
MARZO						
1	2	3	4	5	6	7
	VC5		VC6		E1	
8	9	10	11	12	13	14
	VC7		VC8		E2	
15	16	17	18	19	20	21
	VC9		TC2	FIESTA	E3	
22	23	24	25	26	27	28
29	30	31				
ABRIL						
			1	2	3	4
5	6	7	8	9	10	11
						E1,E2,E3,E4,EF

Fechas recomendadas de entrega de cada notebook

Fecha límite de entrega de notebooks y trabajo final

# Encuestas

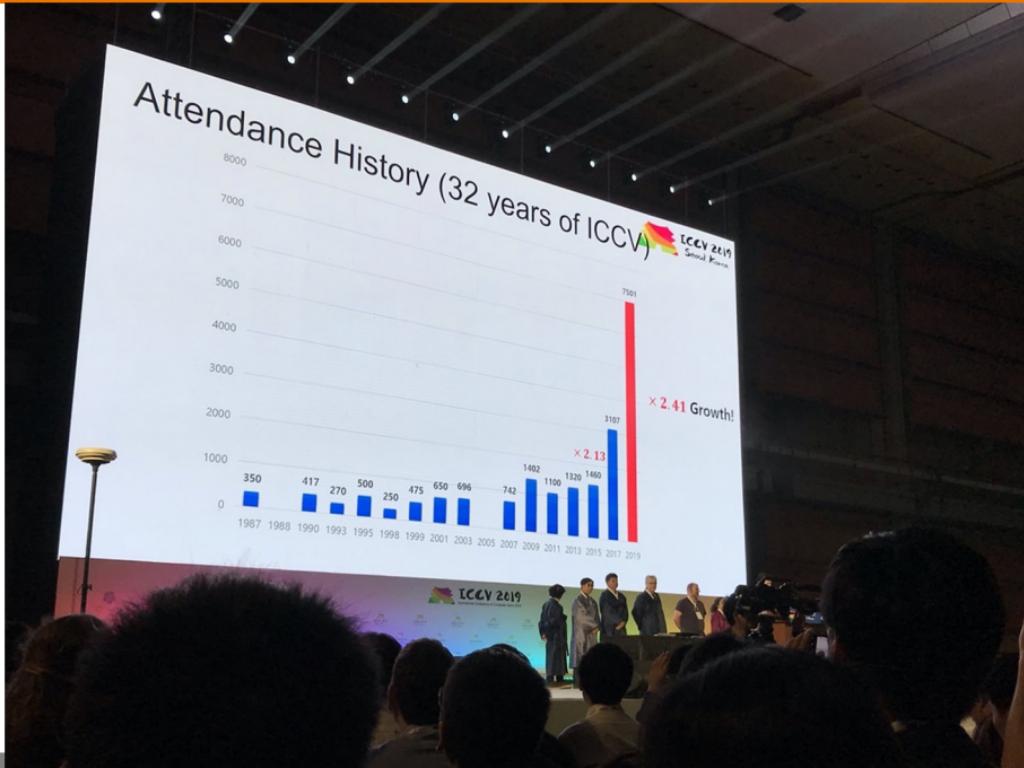
Es muy importante responder a las encuestas, no penséis que no sirven para nada.

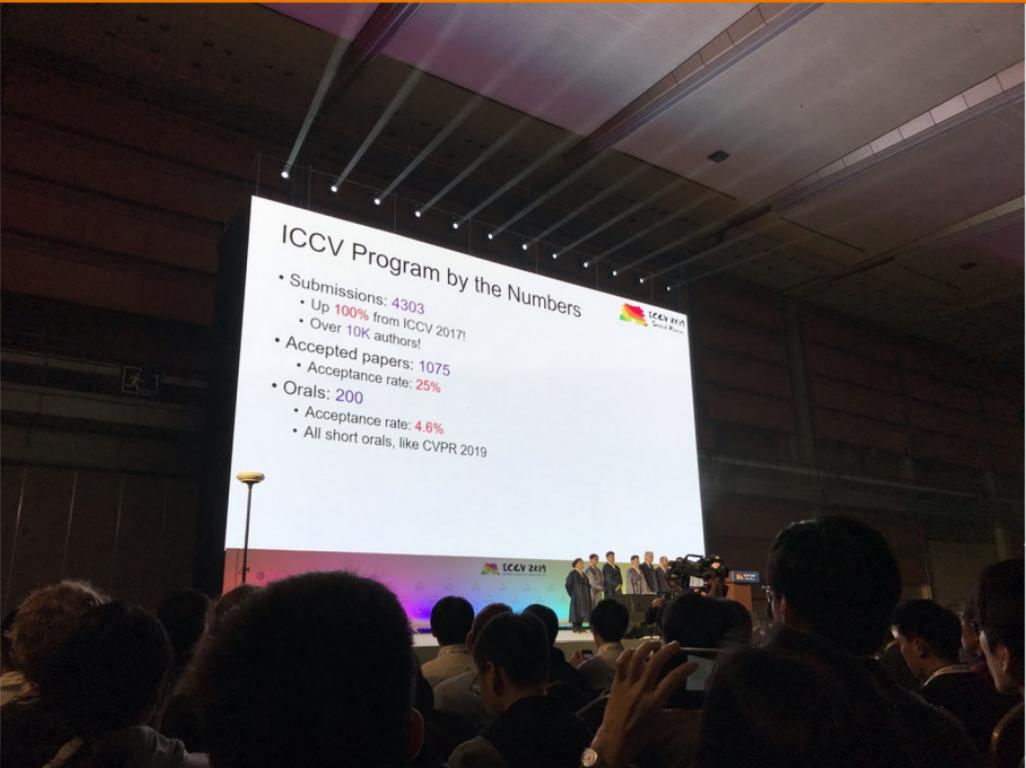
Las tomamos muy en serio y nos ayudan a mejorar.

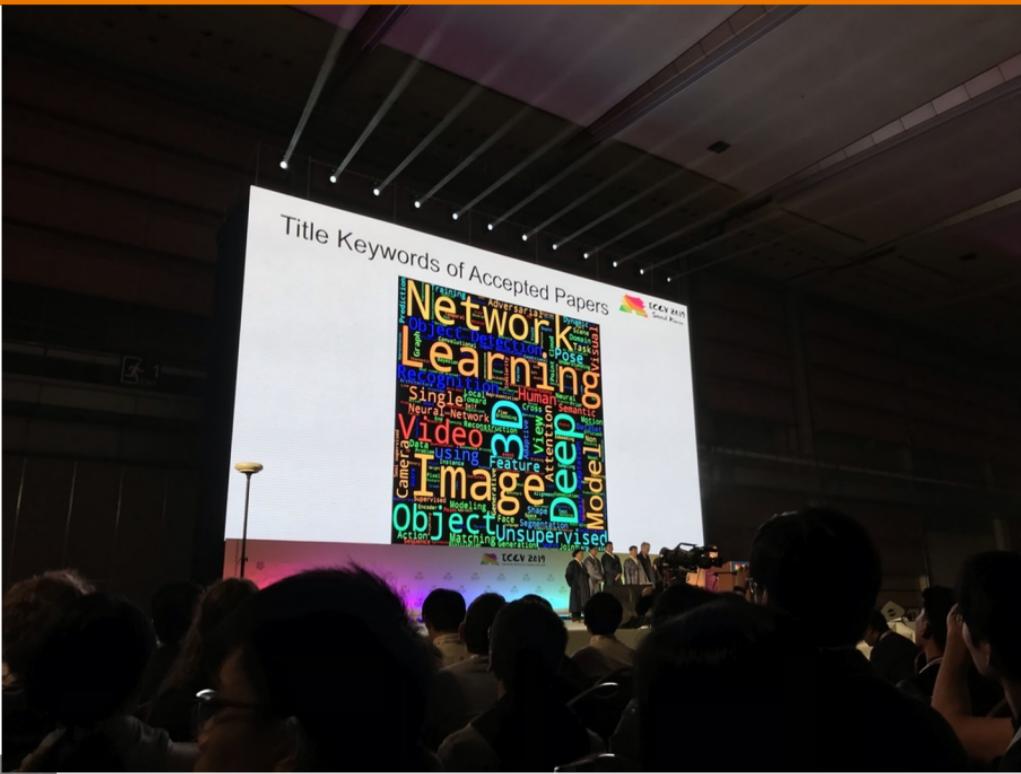
¿Respondéis ahora? ☺

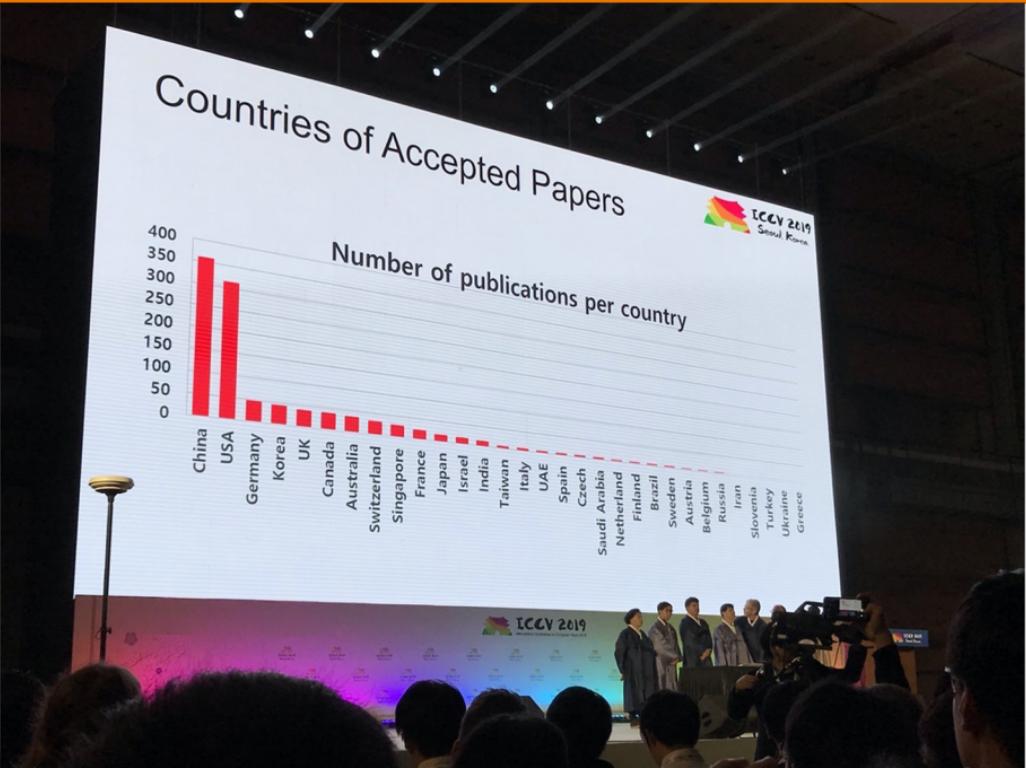
# ICCV'19











Summary

- **SinGAN** - Single Image GAN
- Controllable generation
- Image manipulation applications

Code Available!



Thursday morning, Hall D1



**Video Face Clustering with Unknown Number of Clusters**

Makarand Tapaswi<sup>1,2,3</sup>, Marc T. Law<sup>2,3,4</sup>, Sanja Fidler<sup>2,3,4</sup>  
 {makarand, law, fidler}@cs.toronto.edu  
<sup>1</sup>Inria, <sup>2</sup>University of Toronto, <sup>3</sup>Vector Institute, <sup>4</sup>NVIDIA

**ICCV 2019 Seoul, Korea**  
[github.com/makarandtapaswi/  
 BallClustering\\_ICCV2019](https://github.com/makarandtapaswi/BallClustering_ICCV2019)

**Introduction**

**GOAL:** Face clustering in TV series and movies

- known #characters
- ignore tracks for background characters

**Contributions**

- cluster **all** tracks
- **unknown** #characters
- loss function to learn embedding and stopping criterion for hierarchical clustering

**Other Metric Learning Approaches**

- Contrastive Loss**  
optimizes distance between positive/negative pair
- Triplet Loss**  
optimizes distance between anchor/positive/negative
- Logistic Discriminant Metric Learning**  
converts distance to probability, optimizes BCE
- Prototypical Networks**  
minimize distance to centroids, push class centroids apart

**Ball Cluster Learning**

**Constraints and Loss Functions**

I. Samples from same cluster are in a ball of radius  $b = r^2$

$$\forall x_i \in C_k, d^2(f_i, \mu_k) \leq b$$

$$L_{\text{sim}} = \frac{1}{N} \sum_{k \in C_k} (d^2(f_i, \mu_k) - b)_+$$

Similar case  $d(f_i, \mu_k) \leq r$   
 Dissimilar case  $d(f_i, \mu_k) > 3r$

II. Distance from sample to other cluster centroid is at least  $3r$

$$\forall x_i \in C_k \neq C_\ell, d^2(f_i, \mu_\ell) \geq 9b + \epsilon$$

$$L_{\text{dist}} = \frac{1}{N} \sum_{\ell \in C_k} \max [9b + \epsilon - d^2(f_i, \mu_\ell)]_+$$

Why  $3r$ ?

**Experiments**

**Datasets**

- TRAIN on MovieGraphs: 65K face tracks from 1280 actors
- TEST on TV: The Big Bang Theory, Buffy the Vampire Slayer

**Results**

Video	#Cams	Net Unk.	WCP	NM	HC	WCP	NM	BCL	WCP (ours)	NM	
BBT-1	6	2	9	98.5	88.1	12	96.7	82.3	9	99.2	97.3
BBT-2	12	10	24	97.8	64.2	32	80.2	71.0	27	88.1	82.6
BBT-3	12	25	23	77.7	67.9	38	88.7	76.6	26	90.2	81.7
BBT-1.6	11	26	51	80.0	67.9	87	90.8	70.9	69	94.3	88.3
BBT-1.6	26	83	—	68.7	59.7	123	80.5	65.0	78	86.6	77.0
ALL	37	175	111	71.3	64.0	197	82.8	70.2	126	85.8	80.4

**#Clusters vs. WCP and NMI**

**Buffy SSE2**

**Qualitative result**

BBT S1E1  
 8 characters,  
 7 predicted clusters (BCL),  
 WCP: 98.6, NM: 95.8

The figure shows two plots. The top plot shows SSE2 vs. Number of Clusters (1 to 100). The bottom plot shows SSE2 vs. Number of Clusters (1 to 100). A legend indicates: circle, operating threshold; square, threshold.

**Invariant Information Clustering  
for Unsupervised Image Classification and Segmentation**

Xu Ji, Joao Henriques, Andrea Vedaldi – VGG @ University of Oxford

**Key Features**

- Learn a classifier without labels
- Novel application of information theory to deep clustering
- No need for post-processing or fine-tuning
- Automatically prevents degenerate solutions
- 9 state-of-the-art classification and segmentation results

**Theory**

Mutual information breaks down into entropy and conditional entropy.

$$I(x, x') = I(P) = \sum_{i=1}^C \sum_{j=1}^{C'} P_{ij} \ln \frac{P_{ij}}{P_i P_j} = H(x) - H(x|x')$$

Thus IIC optimises for classifications that are invariant to nuisance perturbations, with even spread, and certain prediction.

**Method**

Given an image (or patch)  $x_i$  and randomly transformed version  $g_i x_i$ , maximise mutual information between their predictions, across the batch.

$$\max_{\Phi} I(\Phi(x_i), \Phi(g_i x_i)) = \max_{\Phi} I(x, x')$$

**Results**

	STL10	CIFAR10	CIFAR100	SIM1024
K-means	19.2	22.4	13.0	37.2
Tropics	24.4	26.5	14.7	52.5
DRW 2015	24.4	26.5	14.7	52.5
GAN 2015	29.8	31.5	14.7	52.5
DRC 2016	29.8	31.5	14.7	52.5
DeepCluster 2017	49.0	30.1	15.1	62.8
ADC 2018	13.4	27.4	23.8	84.3
HC 2018	32.5	32.5	16.9	69.6
<b>IIC</b>	<b>89.4</b>	<b>43.7*</b>	<b>28.7</b>	<b>99.2</b>

Dundar 2015      STL10  
Cutout 2017      74.1  
Oyallon 2017      87.3  
DeepCluster 2018      87.6  
ADC 2018      73.4  
DeepINFOMAX 2018      56.7  
**IIC plus finetune**      77.0  
**88.8**

Results on fully unsupervised classification and segmentation (left), and semi-supervised classification and segmentation (right).

# ¿Preguntas?



**¡Muchas gracias a todos!**

**Ha sido un placer,  
¡hasta pronto!**

# Aprendizaje no supervisado

## TC02: Despedida de la asignatura

Félix José Fuentes Hurtado

[felixjose.fuentes@campusviu.es](mailto:felixjose.fuentes@campusviu.es)

Universidad Internacional de Valencia

18 de marzo de 2021