

# Propuestas de I+D en ML para NSF

Manel Martínez Ramón

1 de julio de 2014

La Fundación Americana para la Ciencia (NSF) es la agencia nacional de financiación del sistema de investigación y ciencia de los EE.UU. Está radicada en Arlington, Virginia.

Datos, datos!!!

- Fue creada en 1950.
- Recibe aproximadamente 40.000 propuestas al año, de las que financia 11.000.
- En 2014 obtuvo 7172 millones de dólares del congreso para investigación (287,8 más que el año anterior).
- Ha financiado a 212 premios Nobel.

- Liderado por dos componentes principales
  - Un director que supervisa al equipo
  - Un responsable de gestión que se encarga de la creación y administración de programas, revisión de méritos, planificación, presupuesto y otros.
- Funciona con un comité de 24 miembros (National Science Board) que se reúne seis veces al año.
- Todos son nombrados por el presidente y ratificados en el senado, y se renuevan cada seis años.
- Existe una plantilla administrativa de 2100 empleados.

Existen siete directorios:

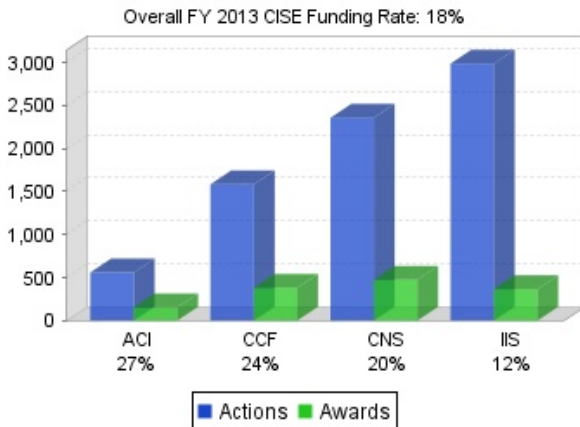
- Ciencias biológicas,
- Ingeniería y ciencia de la computación y la información,
- Ingeniería,
- Geociencias,
- Ciencias matemáticas y físicas,
- Ciencias sociales, conductivas y económicas,
- Educación y recursos humanos.

Cada sección tiene diversas subdivisiones. Da soporte cada año a cerca de 200.000 investigadores.

Nótese que no hay ciencias médicas, ya que dependen de otro organismo, los Institutos Nacionales de Salud (NIH, 30.000 M\$).

- Se pueden enviar propuestas a programas anunciados o propuestas libres que no se vean reflejadas en ningún programa.
- Son evaluadas por tres o más expertos independientes de la NSF.
- Hay un pull de más de 40.000 expertos.
- Hay un sistema de revisión por méritos establecido.
- Las solicitudes rechazadas son informadas y hay derecho a respuesta e iteración.
- ~~Se puede hablar con~~ Hay que llamar a los responsables de directorios o divisiones para preguntar acerca de las posibilidades de la propuesta antes de enviarla.

# The truth...



ACI: Advanced Cyberinfrastructure  
CCF: Comp. and Comm. Found.

CNS: Computer and Network Systems  
IIS: Information & Intelligent Systems

# OK, pero tú qué has hecho...

He presentado cuatro propuestas hasta hoy.

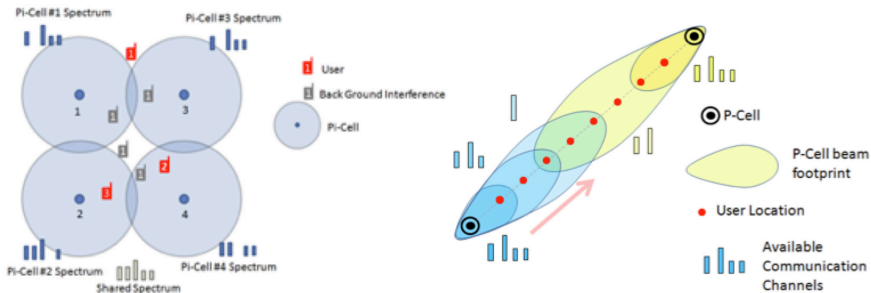
- Una “dear friend letter” en Smart Grid.  
Respuesta: “Dear friend. Your proposal has nothing to do with the call. Please refrain from sending a full proposal...” De buen rollito, cuidao.
- Una propuesta al programa Enhancing Access to the Radio Spectrum (EARS) en radio cognitiva.
- Una propuesta complementaria Small Business Innovation Research (SBIR).
- Una propuesta a CyberSEES en Smart Grid.

## A User Centric Predictive Intelligent (PI-cell) Cooperative Spectrum Sharing Network Approach

- Nuevas aplicaciones móviles demandan computación en línea y transmisión de grandes cantidades de datos.
- Necesidad de transmitir 20 Mbps en entornos competitivos, bandas sin licencia.
- La potencia transmitida está limitada por motivos de seguridad y de autonomía.
- Se dispone de pico celdas y machine learning.



# Concepto de comunicación centrada en el usuario



- $L$  picoceldas.
- “Bandwidth budget” (máx BW)  $w_i^{(p)}$ .
- $M$  usuarios, ancho de banda demandado  $w_j^{(u)}$ .
- Parámetros  $\theta$ .

Se define una matriz  $\mathbf{W}$  con elementos  $w_{ij}$ : ancho de banda en la celda  $i$  para el usuario  $j$ .

Ancho de banda utilizado en cada celda:

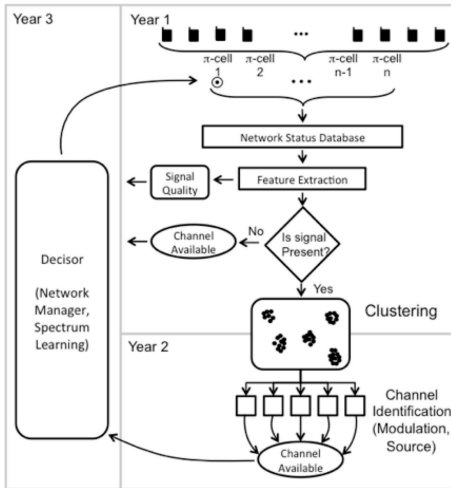
$$W_i^{(p)} = \sum_{j=1}^M w_{ij} \quad (1)$$

Ancho de banda concedido a cada usuario

$$W_j^{(u)} = \sum_{i=1}^L w_{ij} \quad (2)$$

## Elementos de aprendizaje máquina

- 1 “Spectrum Learning” (1st fancy word): capacidad para adquirir **datos** del entorno, procesarlos para obtener **información** (feature extraction) y luego extraer **conocimiento** (machine learning). Objetivo: detectar localmente el espectro utilizable.
- 2 Optimización: ajuste de la matriz **W**.
- 3 “Big data” (3rd fancy word, 2nd is ML).
- 4 Seguimiento de usuarios mediante información de ítem 1 y uso de HMM.
- 5 Other fancy words: “Deep Learning”.



## Personal:

- Un catedrático (IP), un DS (coIP).
- Un profesor titular (coIP).
- Dos estudiantes de doctorado.

## Medios materiales

- Equipos de cómputo.
- Celdas basadas en SDR. Controladas por LabView.
- Terminales de comunicaciones.

## Compañías de comunicaciones:

- National Instruments.
- K&A communications (PYME).

Presupuesto para tres años: 750.000 dólares.

## Cognitive Radio Small Cell for Pervasive Coverage and Sustained Data rates in Mobile Applications

Propuesta complementaria a la anterior: implementación real de una picocelda con las capacidades anteriores, a fin de introducir un producto competitivo en el mercado.

Características	<b>CLX-4G</b>	IP-access	PowerWave	Huawei
4G	<b>LTE</b>	LTE	LTE	LTE
20Mbps	<b>Si</b>	No	No	No
IPV6	<b>Sí</b>	Sí	Sí	Sí
Portadora	<b>Agnóstica</b>	Específica	Específica	Específica
Gestión de espectro	<b>Sí</b>	No	No	No

Presupuesto en la fsse 1: 350.000 dólares.

## Machine-Learning-Based Cyber-Physical and Augmented Virtual Reality Environments for Smart Home Area Networks and micro-grids

- Obj. 1 Toma de decisiones y optimización de micro-redes basado en ML.
- Obj. 2 Aplicaciones sociales cooperativas basadas en ML y fusión de datos. Análisis demográfico y social para el estudio del comportamiento individual con respecto al uso energético.
- Obj 3 Entornos virtuales de realidad aumentada y apps interactivas para el conocimiento, educación y entrenamiento del usuario.
- Obj. 4 Visión común de la conciencia y educación para uso sostenible de la energía en un entorno de medios cibernéticos y sociales conectados con las redes inteligentes.

- Se dispone de TB o PB: Big data.
- Se necesita estimar y predecir el comportamiento del usuario.
- Hay que “optimizar” al usuario, pero no se puede contar excesivamente con su colaboración.
- Diferentes criterios de optimización.
- Se propone utilizar:
  - DBN distribuido para el análisis de grandes cantidades de datos.
  - Predicción con modelos lineales para carga y generación.
  - Herramientas de filtrado colaborativo y apps para la recomendación de uso energético.
  - Heurística para optimización.



Actualmente trabajando en la propuesta de un proyecto al NIH en estudios de neuroimagen multicéntricos.

Problema: se dispone de datos de resonancia de diferentes centros, los cuales no son totalmente consistentes.

Idea: Uso de manifold learning para integrarlos.

Resultados esperados:

- Mejora de la tasa de detección y caracterización de patologías al integrar miles de estudios.
- Detección de las diferencias entre centros.

Gracias :)