A Change-Detection based Framework for Piecewise-stationary Multi-Armed Bandit Problem

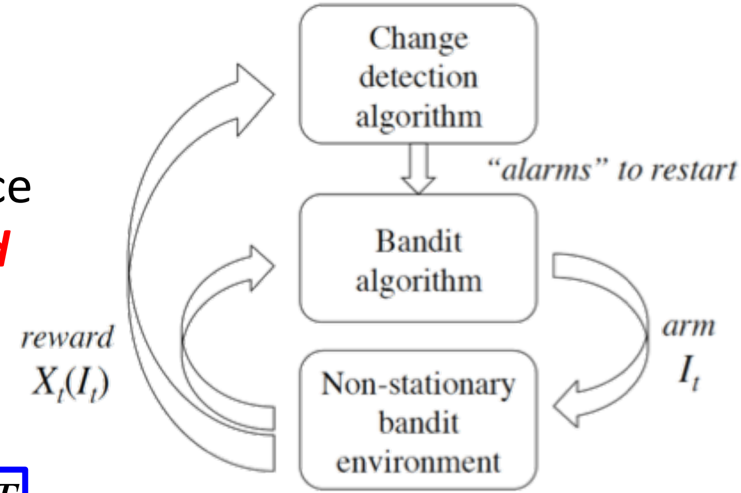
Fang Liu, Joohyun Lee and Ness Shroff

- Multi-Armed Bandits = Slot Machines with Unknown Rewards
  - Example: Ad selection problem of a social media

If P_is are ***stationary****, bandit algorithms achieve* $O(\log(T))$ *regret*

Change Detection based Approach for Non-stationary Bandit

- **In real-life problems, reward distributions are *non-stationary***
- **Non-stationary Multi-Armed Bandit Problem**
  - Developed a Change Detection Algorithm based on CUSUM (CUmulative SUM)
  - Developed CUSUM-UCB (Upper Confidence Bound) with the ***best known regret bound***
  - Evaluated over Big Data (Yahoo click-through rates)



γ_T = number of changes up to time T

	Passively adaptive			*Actively adaptive*		
Policy	D-UCB (Kocsis and Szepesvári 2006)	SW-UCB (Garivier and Moulines 2008)	Rexp3 (Besbes, Gur, and Zeevi 2014)	Adapt-EvE (Hartland et al. 2007)	CUSUM-UCB	*lower bound* (Garivier and Moulines 2008)
Regret	$O(\sqrt{T\gamma_T}\log T)$	$O(\sqrt{T\gamma_T\log T})$	$O(V_T^{1/3}T^{2/3})$	Unknown	$O(\sqrt{T\gamma_T \log \frac{T}{\gamma_T}})$	$\Omega(\sqrt{T})$

  - Generally applicable to many sequential learning problems
    - E.g., Choose a song/channel depending on the moods/situations (Users will give feedbacks)
    - E.g., Choose an angle (control) of drones for a specific mission