A Restless Bandit Marginal Productivity Index for Opportunistic Spectrum Access with Sensing Errors

José Niño-Mora

Department of Statistics

Carlos III University of Madrid

NET-COOP 2009, November 23–25, 2009, Eindhoven

Background: OSA

- Opportunistic spectrum access (OSA) in cognitive radio networks: using efficiently the radio frequency spectrum

- Traditional static spectrum allocation to licensed primary users (PUs): severe inefficiencies, measurements indicate low spectrum utilization (under 30 %)

- Dramatic demand surge for spectrum access by unlicensed secondary users (SUs), due to emerging technologies (wireless Internet, sensor networks, RFID tags, . . .)

- OSA: dynamically exploit transmission opportunities created by PUs' bursty usage patterns, granting SUs access to licensed channels when these are perceived to be unused

Background: OSA policies & sensing errors

- OSA: dynamically exploit transmission opportunities created by PUs' bursty usage patterns, granting SUs access to licensed channels when these are perceived to be unused

- Surge of research on design of OSA policies for dynamic spectrum allocation to SUs: tractable & near-optimal

- Design of OSA policies: complicated by error-prone identification of transmission opportunities in PUs' channels, being based on SUs' spectrum sensors

Background: OSA policies & sensing errors

- Two types of sensing errors: (H_0 : channel idle vs. H_1 : channel busy)
  - Type I error (false alarms: idle channel sensed to be busy): overlooked transmission opportunities
  - Type II error (misdetections: busy channel sensed to be idle): collisions with PUs
- PUs require collision probabilities below acceptable levels
- Some issues to consider for designing OSA policies:
  - distinction: spectrum sensing and spectrum access
  - synchronization: channel status info via ACKs of receipt
  - cost of sensor usage (battery power)

Multichanel OSA model

- M unlicensed SUs $m \in \mathsf{M} \triangleq \{1, \ldots, M\}$
- $N \geq M$ channels licensed to PUs, $n \in \mathsf{N} \triangleq \{1, \ldots, N\}$
- Time slots $t \geq 0$ (synchronized)
- Channel n's bandwidth: $b^{(n)}$ Mbs per slot
- At t: channel n is ($s_t^{(n)} = 1$) or is not ($s_t^{(n)} = 0$) available
- Channel-state transition probabilities:

$p^{(n)} = \mathsf{P}\{s_{t+1}^{(n)} = 0 \,|\, s_t^{(n)} = 1\}$ and $q^{(n)} = \mathsf{P}\{s_{t+1}^{(n)} = 1 \,|\, s_t^{(n)} = 0\}$

- Positive autocorrelation: $\rho^{(n)} \triangleq 1 - p^{(n)} - q^{(n)} > 0$
- Sensor usage cost: $c^{(n)}$

Multichanel OSA model (cont.)

- Channel belief state (availability probab.): $X_t^{(n)} \in \mathsf{X} \triangleq [0,1]$
- Belief-state vector $\mathbf{X}_t = \left(X_t^{(n)}\right)$ known to SUs at start of t
- Actions: $a_t^{(n)} = 1$ (SU senses channel n); $a_t^{(n)} = 0$
- Sensor outcome: $o_t^{(n)} \in \{0,1\}$
- Sensor's detection-quality (ROC curve) on channel n:
  - false alarm probability: $\epsilon^{(n)} \triangleq \mathsf{P}\{o_t^{(n)} = 0 \,|\, s_t^{(n)} = 1\}$
  - misdetection probability: $\delta^{(n)} \triangleq \mathsf{P}\{o_t^{(n)} = 1 \,|\, s_t^{(n)} = 0\}$
- Assume $\delta^{(n)} + \epsilon^{(n)} < 1$
- Channel n's cond. collision probab. must not exceed $\zeta^{(n)}$

Separation of channel sensing and access

- Suppose a SU senses channel n at time t, gets outcome $o_t^{(n)}$
- Access decision: randomized by access probability $y^{(n)}(o)$ given sensor outcome $o \in \{0,1\}$
- Separation principle (Chen et al. '08): choose $y^{(n)}(o)$ to max. the cond. probab. of accessing the channel when available, subject to the cond. collision probab. when unavailable not exceeding $\zeta^{(n)}$
- i.e., solve the linear program (LP)

$$\kappa^{(n)} = \max \epsilon^{(n)} y^{(n)}(0) + (1-\epsilon^{(n)}) y^{(n)}(1)$$

subject to

$$(1-\delta^{(n)}) y^{(n)}(0) + \delta^{(n)} y^{(n)}(1) \le \zeta^{(n)}$$

$$0 \le y^{(n)}(o) \le 1, \quad o \in \{0,1\}.$$

ACKs and belief-state dynamics

- After a transmission on channel n: $K_t^{(n)} = 1$ (positive ACK, successful) or $K_t^{(n)} = 0$
- **Belief-state dynamics (Bayesian update):**
  - Under action $a_t^{(n)} = 1$ (sense channel n),

$$X_{t+1}^{(n)} = \begin{cases} q^{(n)} + \rho^{(n)} & \text{w.p. } \kappa^{(n)} X_t^{(n)} \\ q^{(n)} + \rho^{(n)} \dfrac{1 - \kappa^{(n)}}{1 - \kappa^{(n)} X_t^{(n)}} X_t^{(n)} & \text{w.p. } 1 - \kappa^{(n)} X_t^{(n)} \end{cases}$$

  - Under action $a_t^{(n)} = 0$, $X_{t+1}^{(n)} = q^{(n)} + \rho^{(n)} X_t^{(n)}$
- Sensing allocation decisions: based on **scheduling policy** $\pi \in \Pi$ (nonanticipative, $\sum_n a_t^{(n)} \leq M$)

Rewards and costs for channel n

- Reward per successful transmission: $b^{(n)}$

- Sensing cost: $c^{(n)}$ (energy cost)

- One-slot reward function: $R^{(n)}(x, a) \triangleq \big(b^{(n)}\kappa^{(n)}x - c^{(n)}\big)a$

Scheduling problem: discounted criterion

- **Performance optimization problem:**

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \mathsf{E}_{\mathbf{x}_0}^{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \sum_{n \in \mathsf{N}} \beta^t R^{(n)}\big(X_t^{(n)}, a_t^{(n)}\big) \right]$$

- Intractable **Partially Observed Markov Decision Process** (POMDP)

- **Goal: tractable, near-optimal policy** $\widehat{\boldsymbol{\pi}}^{\beta,*}$

Scheduling problem: average criterion

- **Performance optimization problem:**

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \liminf_{T\to\infty} \frac{1}{T} \mathsf{E}^{\boldsymbol{\pi}}_{\mathbf{x}_0} \left[\sum_{t=0}^{T} \sum_{n\in\mathsf{N}} R^{(n)}\left(X_t^{(n)}, a_t^{(n)}\right) \right]$$

- Intractable **Partially Observed Markov Decision Process** (POMDP)

- **Goal:** construct **tractable, near-optimal policy** $\widehat{\boldsymbol{\pi}}^*$

Dynamic index policies

- A practical **class of policies**
- To each channel n: attach **index** $\lambda^{(n)}(x^{(n)})$
- **Index policy:**
  - Use $\lambda^{(n)}(x^{(n)})$ as a sensing-priority index
  - Select channel(s) to sense in nonincreasing order of index values (among those w/ $\lambda^{(n)}(x^{(n)}) > c^{(n)}$, if any)
- **How to design & evaluate a good index $\lambda^{(n)}(x^{(n)})$?**

Some recent work

- $n = 1$ channel, perfect sensing, transmission errors: Johnston & Krishnamurthy '06, optimal threshold policy

- $n = 1$ channel, sensing errors, etc.: Hoang et al. '09

- n-channel case:

  - w/ sensing errors: Zhao et al. '07, heuristic myopic index policy $\lambda^{(n)}(x^{(n)}) = b^{(n)}\kappa^{(n)}x^{(n)}$

  - w/ sensing errors, separation principle: Chen et al. '08

  - w/out sensing errors, multiarmed restless bandits, Whittle index policy: Le Ny et al. '08, Liu & Zhao '08, NM '08 (closed-form index & bound)

  - w/ sensing errors, multiarmed classic bandits,Gittins index policy: Ai & Abouzeid '09

Research goal and methodology

- Recall: without sensing errors, multiarmed restless bandits, Whittle index policy: Le Ny et al. '08, Liu & Zhao '08, NM '08 (closed-form index policy with near-optimal performance, & bound)

- Main goal: obtain and test Whittle's index policy in a model incorporating sensing errors (& sensing costs)

- Difficulty: lack of prior general methods of analysis for continuous-state restless bandit indexation

- Methodology: will deploy recent results (NM) on general sufficient indexability conditions & index evaluation for continuous-state restless bandits

Index policies & bandits

- **Bandit**: MDP project model, active/passive-action
- **Multi-armed bandit problem** (MABP): Which project to engage at each time to max. exp. total discounted reward?
- Gittins & Jones '74: optimal index policy for **classic MABP** (passive projects do not change state)
- Whittle '88: heuristic index policy for **restless MABP** (passive projects can change state); Weber & Weiss '90, '91
- NM (last decade): theory, algorithms and applications for restless bandit indexation, based on **partial conservation laws** (PCLs) and **marginal productivity** (MP) index

Towards the Whittle index

- general Markovian (single) project of restless bandit type
- π: activation policy; Π: admissible policies (nonanticip.)
- x_0: initial state
- λ: activity charge (price of work per period)
- **Reward measure:** $f(x_0, \pi) \triangleq \mathsf{E}_{x_0}^{\pi}\left[\sum_{t=0}^{\infty} \beta^t R(X_t, a_t)\right]$
- **Work measure:** $g(x_0, \pi) \triangleq \mathsf{E}_{x_0}^{\pi}\left[\sum_{t=0}^{\infty} \beta^t a_t\right]$
- **λ-charge problem:** $\max_{\pi \in \Pi} f(\cdot, \pi) - \lambda\, g(\cdot, \pi)$

The Whittle index

- **λ-charge problem:** $\max_{\pi \in \Pi} f(\cdot, \pi) - \lambda\, g(\cdot, \pi)$

- Find a project-activation policy maximizing the value of rewards minus activity charges

- **Def:** project is **indexable** if $\exists$ an **index** $\lambda^*(x)$ such that $\forall$ charge $\lambda \in \mathbb{R}$ and state x,

it is optimal to engage the project in state $X_t = x \iff \lambda^*(x) \geq \lambda$

Towards sufficient indexability conditions

- Given a subset of states S, policy $\langle a, S \rangle$ takes action a at time $t = 0$, and then follows the S-**active policy**

- **Marginal work measur.**: $w(x, S) \triangleq g(x, \langle 1, S \rangle) - g(x, \langle 0, S \rangle)$

- **Marginal reward**: $r(x, S) \triangleq f(x, \langle 1, S \rangle) - f(x, \langle 0, S \rangle)$

- **Marginal productivity measure**: $p(x, S) \triangleq \dfrac{r(x, S)}{w(x, S}$

- Focus on **threshold policies:** S is of the form $S = (z, \to)$ for some z (project is active when its state X_t lies above z)

Suff. indexab. cond. w.r.t. threshold policies

- **Def:** The project is **PCL-indexable** if:

(i) **Positive marginal work**: $w(x, z) > 0, \quad \forall\, x, z$

(ii) The following MP index is nondecreasing:

$$p^*(x) \triangleq \frac{r(x, x)}{w(x, x)}$$

- **Th.:** PCL-indexable $\Longrightarrow$ indexable, w/ Whittle index $p^*(x)$

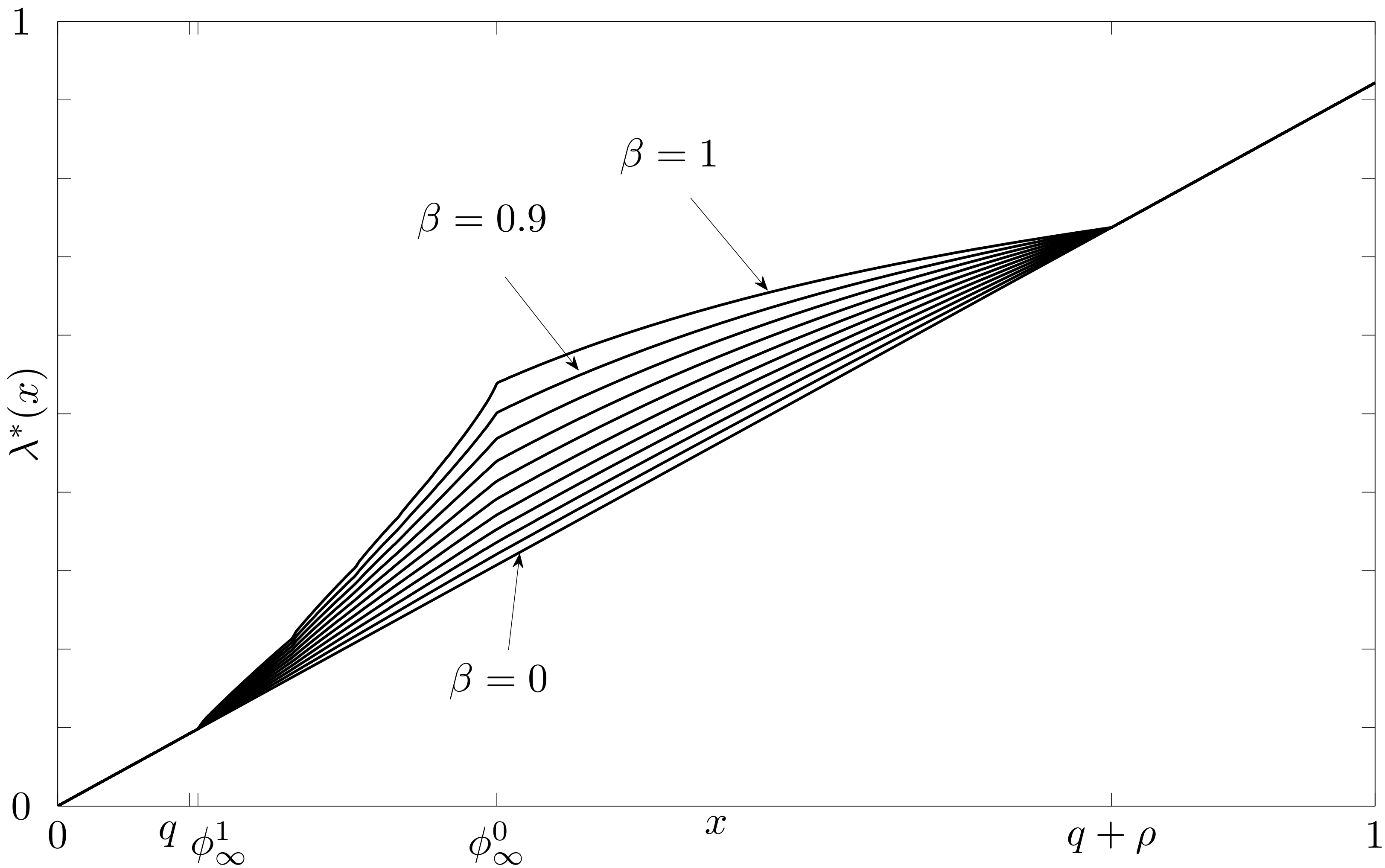
- "**PCL**": **partial conservation laws**

- Extends previous results of NM for discrete-state restless bandits to continuous-state case

Results of indexability analysis

- The model satisfies the sufficient indexability conditions
- Obtain closed-form index expressions
- Piecewise smooth index formulae:
  - cases # I, II: $0 \le x \le \phi^1_\infty$, $p^*(x) = b\kappa x$ (myopic)
  - case # III: $\phi^1_\infty < x < \phi^0_\infty$, closed-formula for $p^*(x)$ (involving inf. series)
  - case # IV: $\phi^0_\infty \le x \le q + \rho$, closed-formula for $p^*(x)$ (involving inf. series)
  - case # V: $q + \rho < x \le 1$, $p^*(x) = b\kappa x$ (myopic)

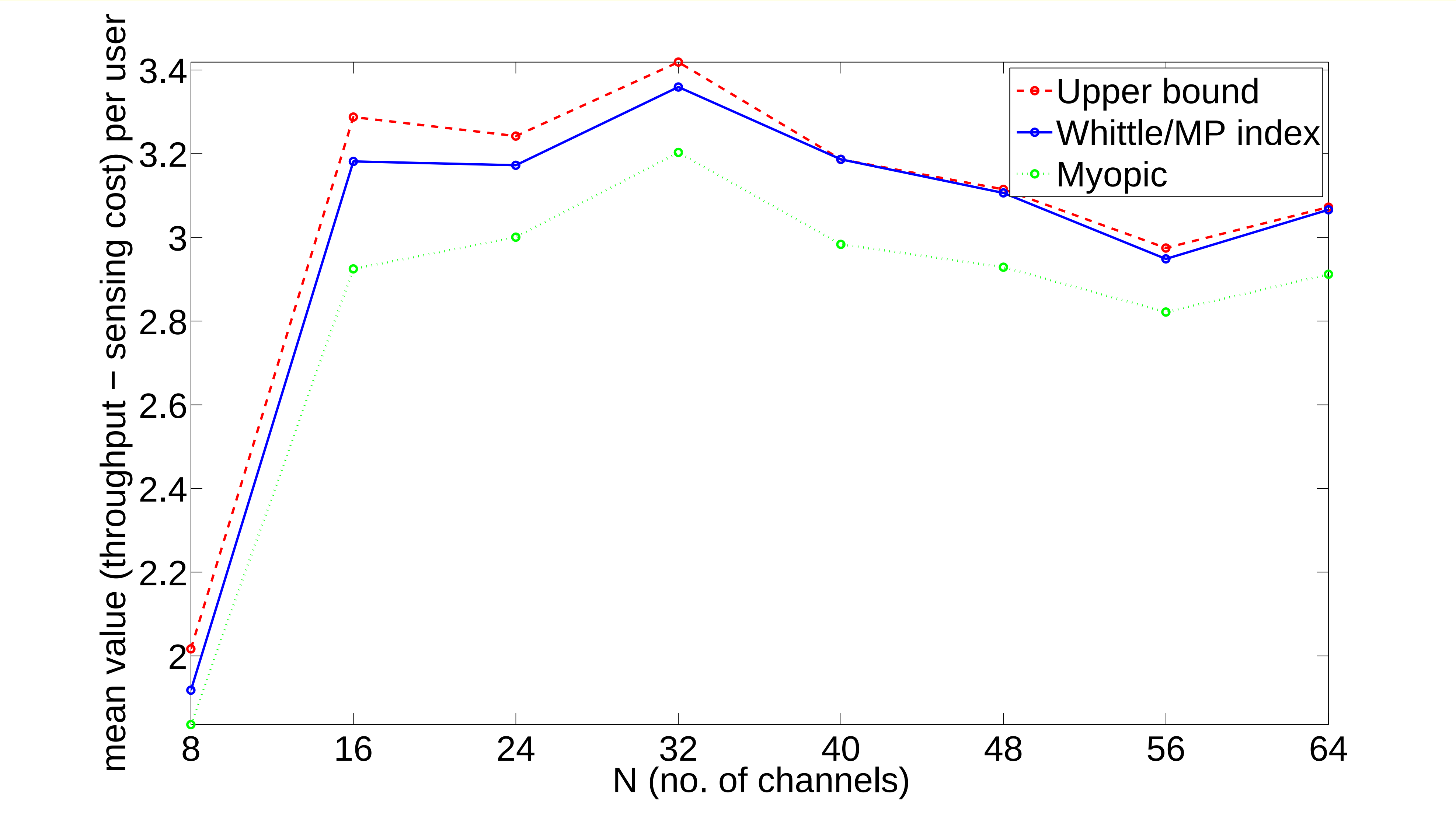
Example: index evaluation



A numerical experiment

- Number of channels: $N = 8, 16, 24, \ldots$
- Number of SUs: $M = N/8$
- $\beta = 0.99$
- sensing cost: $c = 0.4$
- Channel parameters: randomly generated

A numerical experiment: results



Conclusions

- New, well-grounded index policy resulting from a unifying methodology

- Encouraging computational results

- Many possible extensions to address