Selecting Multiple Website Elements
James Edwards, supervised by David Leslie
STOR-i CDT, Lancaster University

THE PROBLEM
Select a set of website elements to display to a user.
- Adverts on search results.
- Movie suggestions.
- Recommendations on a retail site.

Clicks depend on the set of elements not individual elements. This is a bandit problem whose arms are the elements. The objective is to maximise expected click-through rate (CTR) over time.

Context and Uncertainty
The context is extra information specific to each time that affects the CTR e.g. user information (location, preferences) or search terms entered. This is rarely known exactly:
- A search term has multiple meanings e.g. jaguar or flash.
- A travel based search could be for business or a holiday.
- Movie preferences change when watching alone or with friends.

This uncertainty induces diversity in solution sets.

Click Models
At each time the user is summarised by a latent state $x \in \{1, 2, \ldots, n\}$. Only its distribution is known. Each arm $a$ has $n$ corresponding weights $w_{a,x} \in [0, 1]$ representing its quality for state $x$.

For a single arm the CTR is $w_{a,x}$. For a set of arms $A$ either one or none are clicked. Two possible models for set CTR are:

- **Probabilistic Click Model (PCM)**: An independent Bern($w_{a,x}$) trial is run for each arm. There is a click if any is a success.
- **Threshold Click Model (TCM)**: Each user has a threshold $u \sim U(0, 1)$. There is a click if $\exists a \in A$ such that $w_{a,x} > u$.

The set CTR is then:

\[
Pr(\text{Click}) = \begin{cases} 
1 - \prod_{a \in A} (1 - w_{a,x}) & \text{for PCM} \\
\int_{0}^{1} - \prod_{a \in A} (1 - w_{a,x}) \, du & \text{for TCM} 
\end{cases}
\]

Comparison. PCM is simpler but TCM induces greater diversity. With two identical arms the second arm increases CTR under PCM but not under TCM. This suggests PCM undervalues diversity.

The solution method takes three parts:
- A Bayesian scheme that converges to the true weights with sufficient observations (updating beliefs).
- An algorithm that chooses a good arm set when weights are known (exploitation).
- A bandit algorithm that learns the weights without neglecting short term rewards (exploration).

SOLUTION METHOD
The unobserved state $x$ makes exact updating impractical. A form of online expectation maximisation can be used:
1. Obtain $X$, the posterior distribution for $x$ given user action.
2. Sample an $\tilde{x}$ from $X$.
3. Update weight beliefs assuming $x = \tilde{x}$.

Exploitation
The CTR for both PCM and TCM is submodular. For these problems a greedy algorithm is known to perform well. Arms are added sequentially, choosing at each step the arm that maximises the increase in expected CTR.

Exploration
The true weights are not known so instead a proxy $\tilde{w}$ is used. This is usually some function of the weight beliefs and can be implemented easily by adapting an existing bandit index policy e.g. Thompson Sampling, UCB.

Acknowledgements
This work was funded by the Google Faculty Research Award.