**Incentivizing information explorers (when they’d really rather exploit)**

**By Microsoft blog editor**

Everyone is familiar by now with recommendation systems such as on Netflix for movies, Amazon for products, Yelp for restaurants and TripAdvisor for travel. Indeed, quality recommendations are a crucial part of the value provided by these businesses. Recommendation systems encourage users to share feedback on their experiences and aggregates the feedback in order to provide users with higher quality recommendations – and, more generally, higher quality experiences – moving forward.

From the point of view of online recommendation systems in which users both consume and contribute information, it could be said that users play a dual role; they are information explorers *and* information exploiters. We’ve each of us been one or the other at various moments in our decision-based sorties online, shopping for a specific item, deciding on a service provider, and so on. The tradeoff between exploration and exploitation is a well-known subject in machine learning and economics. Information exploiters seek out the best choice given the information available to date. Information explorers on the other hand would seem willing to try a lesser known – or even unknown – option, for the sake of gathering more information. But an explorer does this at the real risk of a negative experience.

In an ideal world, a recommendation system would not only control the information flow to the users in a beneficial way in the short-term, it would also strive to *incentivize exploration* in the longer term via what we call *information asymmetry:* the fact that the system knows more than any one agent. It would do this despite users’ incentives often tilting this delicate balance in favor of exploitation.

This tension, between exploration, exploitation and users’ incentives, provides an interesting opportunity for an entity – say, a social planner – who might manipulate what information is seen by users for the sake of a common good, balancing exploration of insufficiently known alternatives with the exploitation of the information amassed. And indeed, designing algorithms to trade off these two objectives is a well-researched area in machine learning and operations research.

Again, there can be disadvantages to choosing to be an explorer. You risk all the downsides of your strategy (say, buying a product that you wish you hadn’t, or staying in a hotel you can’t wait to check out of). The upsides (your cautionary feedback) on the other hand benefit numerous users in a future that has yet to arrive. This is to reiterate that users’ incentives are naturally skewed in favor of being an exploiter.

Enter the problem of incentivizing information exploration (a.k.a., Bayesian Exploration), where the recommendation system (here called the “principal”) controls the information flow to the users (the “agents”) and strives to incentivize exploration via information asymmetry

**“Think of Bayesian exploration as a protection from *selection bias* – a phenomenon when the population that participates in the experiment differs from the target population. People who’ll rate a new Steven Seagal movie mainly come from a fairly small (and weird) population of fans, such as myself.” – Alex Slivkins, Senior Researcher, Microsoft Research New York City**

In “[Bayesian Exploration with Heterogenous Agents](https://nam06.safelinks.protection.outlook.com/?url=https%3A%2F%2Farxiv.org%2Fabs%2F1902.07119&data=02%7C01%7Cv-gesole%40microsoft.com%7Ca2c19e84a4964f49095f08d6c354490a%7C72f988bf86f141af91ab2d7cd011db47%7C1%7C0%7C636911164097567895&sdata=iSSYzrTnNeofgmUjLalS1PFWWpfFkk72LzHf7lrzVfY%3D&reserved=0)” to be presented at [The Web Conference 2019](https://www2019.thewebconf.org/) in San Francisco May 13-17, Nicole Immorlica and Aleksandrs Slivkins of Microsoft Research New York City, along with Jieming Mao of the University of Pennsylvania and Zhiwei Steven Wu of the University of Minnesota[[1]](#footnote-1) propose a simple model that takes as its embarkation point Bayesian exploration. The researchers allow heterogeneous users, relaxing a major assumption from prior work that users have the same preferences from one moment in time to another. The goal is to learn the best personalized recommendations. One particular challenge was how to incentivize some of the user types to take some of the actions, no matter what the principal does or how much time the principal has. The researchers considered several versions of the model, depending on whether and when the user types are reported to the principal, and they designed a near-optimal “recommendation policy” for each version. They also investigated how the model choice and the diversity of user types impact the set of actions that can possibly be “explored” by each type.

Earlier work in Bayesian exploration – much of which came from Microsoft coauthors -- relies on the inherent information asymmetrybetween the recommendation system (the principal) and a stream of users – the self-interested agents – arriving one at a time. Each agent needs to take an action from a given set of alternatives. The principal issues a recommendation and observes the outcome. But it can’t tell the agent what to do. The problem is to design a recommendation policy for the principal that learns over time to make good recommendations *and* thatalsoensures that the agents are incentivized to follow this recommendation. A single round of this model is a version of a well-known Bayesian persuasion game from theoretical economics

In contrast, the researchers used Bayesian exploration with the heterogeneous preferences of each agent being known up front and assigned using type. For example, pets allowed versus no pets allowed in accommodations bookings. Each time an agent takes an action, the outcome depends on the action itself (for example, the selection of hotel), the “state” of the world (for example, the qualities of the hotels), and the type of the agent. The state is persistent; it does not change over time. However, the state is not known initially. A Bayesian prior on the state is common knowledge. In each round, the agent type is drawn independently from a fixed and known distribution. The principal strives to learn the best possible recommendation for each agent type.

**“Users are both consumers and producers of information. How do we coordinate a crowd of diverse users to efficiently gather information for the population? Information asymmetry is the key.” – Steven Wu, Assistant Professor, University of Minnesota**

The researchers considered three models, depending on whether and when the agent type would be revealed to the principal. They designed a near-optimal recommendation policy for each modeling choice. The three models envisioned were:

* Public Types – The type is revealed immediately after the agent arrives.
* Reported Types – The type is revealed only after the principal issues a recommendation.
* Private Types – The type is never revealed.

The researchers took pains to examine how the set of all explorable type-action pairs – the explorable set – might be influenced by the model choice and the diversity of types. For starters, it soon became evident that for each problem instance the explorable set stays the same when transitioning from public types to reported types, and can only become smaller if transitioned from reported types to private types. Concrete examples of when the reported-to-private transitions make a significant difference are provided.

They also played with the distribution of agent types. For public types (and therefore also for reported types), they observed that the explorable set is determined by the support set of the distribution. Further, if the support set is larger, then the explorable set can only become larger. In other words, they found that *diversity of agent types helps exploration*.

The paper provides a solid example of the explorable set increasing very substantially even when the support set increases by a single type. For private types, however, the reality is very different—*diversity can hurt.* Intuitively, with private types, diversity muddles the information available to the principal making it harder to learn about the state of the world, whereas for public types diversity helps the principal hone its beliefs about the state.

**Explore this**

In staking out their methodology, the researchers first embarked on developing a recommendation policy for public types. In the long run, the public type policy matched the benchmark of “best explorable action”. And though it’s a simple matter to demonstrate that such a policy exists, the challenge lay in providing it as an explicit procedure. They opted for a policy that focused on exploring all explorable type-action pairs. Exploration needs to proceed gradually, whereby exploring one action can enable the policy to explore another. In fact, they discovered that exploring some action for one type could enable the policy to explore some action for another type.

The policy proceeded in phases. In each phase, they explored all actions for each type that could be explored using information available at the start of the phase. Agents of different types learn separately, in per-type “threads"; the threads exchange information after each phase.

An important building block was the analysis of the single-round game. They used information theory to  
characterize how much state-relevant information the principal has. In particular, they proved a version of *information-monotonicity*: the set of all explorable type-action pairs can only increase if the principal has more information.

The special pieceis a policy for private types. Recommending one particular action to the current agent is not very meaningful because the agents’ type is not known to the principal. Instead, one can recommend a *menu*: a mapping from agent types to actions. As in the case with public types, the focus was on explorable menus and the steady exploration all such menus, eventually matching the Bayesian-expected reward of the best explorable menu. The tricky bit was that exploring a given menu does not immediately reveal the reward of a particular type-action pair (because multiple types obviously might map to the same action). As a result, even keeping track of what the policy knew became non-trivial. The analysis of the single-round game got involved, with the need to argue about “approximate information-monotonicity”. To handle these issues, the researchers went for a recommendation policy that satisfied a relaxed version of incentive-compatibility.

In the reported types model, the researchers faced a similar issue, but got a much stronger result. They designed a policy that matched their public-types benchmark in the long run. This might strike one as counterintuitive; remember, “reported types” are completely useless to the principal in the single-round game—and public types are useful as can be. Essentially, they reduced the problem to the public types case, at the cost of a much longer exploration.

**Further exploration**

These are interesting and relevant problems to solve. “There’s so much to be done,” said Slivkins, “both in terms of improving the machine learning performance and in terms of relaxing the assumptions on self-interested human behavior.” Slivkins believes possible applications in the medical and drug trials arena are especially promising. “You can think of a medical trial as an algorithm which assigns drugs to patients, trading off exploration and exploitation. Every time I take a medicine I could be in a medical trial, contributing valuable data to science. How do we improve participation incentives to make it happen?”

The researchers look forward to discussing these ideas with you in further detail and take the conversation in perhaps unexplored directions next month in San Francisco at The Web Conference 2019.

Jieming Mao and Steven Wu were affiliated with MSR-NYC at the time of this research, respectively as an intern and a postdoc.

**Social**

**Twitter**

Are you an explorer or an exploiter when it comes to information? Well, we're one or the other or both, depending. Learn how incentivizing exploration via information asymmetry can lead to better recommendation systems – with potential life enhancing applications beyond mere shopping. LINK

**To explore or exploit: what should we recommend?**

**Facebook**

Are you an explorer or an exploiter when it comes to information? The answer is all of us are one or the either, depending. And that’s OK. But we’re all better off when each of us does just a bit more info-exploring. Learn how incentivizing exploration via information asymmetry can be a better system – with potential life enhancing applications beyond mere shopping.

1. [↑](#footnote-ref-1)