Thompson Sampling for Contextual Bandits with Linear Payoffs

Akshayaa Magesh

May 2019

Abstract

Multi-armed bandit problems provide a framework to model sequential decision problems with an inherent exploration and exploitation trade-off. One of the earliest algorithms for solving the multi-armed bandit problem is Thompson Sampling. It is a randomized algorithm based on Bayesian ideas. A generalization of Thompson Sampling for the stochastic contextual multi-armed bandit problem with linear payoff functions with contexts provided by an adaptive adversary has been studied in [5]. A high probability bound of $O(\frac{d^2}{\epsilon}\sqrt{T^{1+\epsilon}})$ for a time horizon T for any $0 < \epsilon < 1$ has been provided in [5]. This bound is close to the theoretical bound of $O(d\sqrt{T})$ provided for this problem.

1 Introduction

Multi-armed bandit problems model the exploration-exploitation trade-off inherent in sequential decision problems. Many versions of the multi-armed bandit framework have emerged over the decades to solve a variety of problems. The general setting of the multi-armed problem is that there are N arms available to a player and the player at each time upon choosing an arm receives a reward. The player doesn't have any prior knowledge about the reward distributions. The goal of the policy employed by the user is to maximise the rewards accumulated over time. One of the most commonly used frameworks is the stochastic multi-armed bandit. In this setting the reward of each arm is in $[0, 1]$ with unknown expectation values and reward distributions. Let μ_i denote the mean reward of arm i and $\mu^* = \arg\max_{i\in[N]} \mu_i$. The regret accumulated until time horizon T is given by:

$$R(T) = T\mu^* - \sum_{i=1}^{T} \mathbb{E}\left[\mu_{a(t)}\right] \tag{1}$$

where $a(t)$ is the action taken by the player at time t. The player typically faces an "exploration versus exploitation dilemma" : at time t, she can take advantage of the information she has gathered, by choosing the so-far best

performing arm, but she has to consider the possibility that the other arms are actually under-rated and she must play sufficiently often all of them. In the seminal work of Lai and Robbins [1], a lower bound of $O(\log T)$ was proved for the stochastic multi-armed bandit setting and a set of algorithms achieving this lower bound for parametric families was proposed. However these algorithms weren't computationally efficient.

Two classes of algorithms has emerged over the decades to solve the multi-armed bandit. The first one is a frequentist approach that usually involves calculating an upper confidence bound (UCB) for the mean rewards of the arms in addition to their estimates. At each time t the player plays the arm with highest UCB. The kl-UCB algorithm proposed by Garivier and Cappe [2] has been shown to be order optimal. The other class of algorithms takes a Bayesian approach to solving the multi-armed bandit problem. Thompson Sampling is one of the oldest heuristics in the class of randomized probability matching algorithms. It was proposed by W.R. Thompson and dates back to 1933. The basic idea is to assume a simple prior on the parameters of the rewards of every arm and at every time step, play an arm according to its posterior probability of being the best arm. Although Thompson Sampling algorithms is a Bayesian approach, the algorithm and the analysis provided by Agarwal and Goyal [3] apply to the prior-free stochastic multi-armed bandit model where parameters of the reward distribution of every arm are fixed, though unknown. The assumed Bayesian priors could be interpreted as the current knowledge the algorithm has about the arms.

Several studies have shown the efficacy of Thompson Sampling and that Thompson Sampling is more robust to delayed or batched feedback than other methods. Though TS has been empirically shown to perform optimally, the first theoretical guarantee of $O(\log T)$ for the stochastic multi-armed bandit case was provided by Agarwal and Goyal [3].

In the contextual multi-armed bandit problem, in each round, the player is presented with the choice of N arms. Before making the choice of which arm to play, the learner sees d-dimensional feature vectors b_i for each arm i, referred to as context. This context vector is provided by an adaptive adversary that can observe the previous plays and rewards obtained by the user. The player uses these feature vectors and rewards of the arms played in the past to make the choice of arm to play in the current round. Over time, the learner's aim is to gather enough information about how the feature vectors and rewards relate to each other, so that the player can predict with some certainty, which arm is likely to give the best reward in relation to the feature vectors.

In the contextual bandits setting with linear payoff functions, the player competes with the class of all linear predictors on the feature vectors. The predictor is defined by a d-dimensional parameter $\overline{\mu} \in \mathbb{R}^d$ and the predictor ranks the arms according to $b_i^T \overline{\mu}$. In [5], the stochastic contextual bandit problem is considered under linear realizability assumption, *i.e.*, there is an underlying unknown parameter $\mu \in \mathbb{R}^d$ such that the expected reward for each arm i, given the context b_i is $b_i^T \mu$. Under this assumption, the player's aim is to learn this underlying parameter.

The contextual bandit problem with linear payoffs is a widely studied problem in statistics and machine learning often under different names as mentioned by Chu et al. (2011): bandit problems with co-variates (Woodroofe, 1979; Sarkar, 1991), associative reinforcement learning (Kaelbling, 1994), associative bandit problems (Auer, 2002; Strehl et al., 2006), bandit problems with expert advice (Auer et al., 2002), and linear bandits (Dani et al., 2008; Abbasi-Yadkori et al., 2011; Bubeck et al., 2012). The name contextual bandits was coined in Langford & Zhang (2007).

A lower bound of $O(d\sqrt{T})$ for this problem was given by Dani et al. [4], when the number of arms is allowed to be infinite. Abbasi-Yadkori et al. (2011) analyze a UCB-style algorithm and provide a regret upper bound of $O(d \log T\sqrt{T} + \sqrt{dT \log T\delta})$. Apart from the dependence on ϵ, the bounds presented in [5] are essentially away by a factor of d from the lower bound.

In [5], a natural generalization of TS for contextual bandits using Gaussian priors for the parameters of the reward functions and Gaussian likelihood function for the reward distribution are used. A novel martingale-based analysis technique is used to demonstrate that Thompson Sampling achieves high probability, near optimal regret bounds for stochastic contextual bandits with linear payoffs.

While the regret bounds provided in [5] do not match or better the best available regret bounds for the extensively studied problem of linear contextual bandits, the results demonstrate that the natural and efficient heuristic of Thompson Sampling can achieve theoretical bounds that are close to the best bounds. The main contribution of [5] is to provide new tools for analysis of Thompson Sampling algorithm for contextual bandits, which despite being popular and empirically attractive, has eluded theoretical analysis.

2 Thompson Sampling

The priors of the Thompson Sampling algorithm are updated according to the Bayes rule. If the likelihood function of reward r with parameter μ is denoted by $p_\mu(r)$ and the prior for parameter μ given by $q(\mu)$, then the posterior update according to the Bayes rule is given by:

$$q(\mu|r) = \frac{p_\mu(r)q(\mu)}{\int p_\mu(r)q(\mu)d\nu(\mu)} \tag{2}$$

The Thompson Sampling approach gives an intuitive algorithm for the case of Bernoulli bandits, where the rewards can be either 0 or 1 and for arm i the probability of success is μ_i. The algorithm maintains Bayesian priors on the Bernoulli means $\mu_i's$. It employs the class of Beta distributions as the priors for the $\mu_i's$. Beta distribution turns out to be a very convenient choice of priors for Bernoulli rewards. The beta distributions form a family of continuous probability distributions on the interval $(0, 1)$. The pdf of Beta (α, β), the beta distribution with parameters $\alpha > 0$, $\beta > 0$, is given by $f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$. The mean of Beta$(\alpha, \beta)$ is $\alpha/(\alpha + \beta)$;

and as is apparent from the pdf, higher the α, β, tighter is the concentration of Beta(α, β) around the mean. Beta distribution is useful for Bernoulli rewards because if the prior is a Beta(α, β) distribution, then after observing a Bernoulli trial, the posterior distribution is simply Beta$(\alpha + 1, \beta)$ or Beta$(\alpha, \beta + 1)$, depending on whether the trial resulted in a success or failure, respectively.

The Thompson Sampling algorithm initially assumes arm i to have prior Beta$(1, 1)$ on μ_i, which is natural because Beta$(1, 1)$ is the uniform distribution on $(0, 1)$. At time t, having observed $S_i(t)$ successes (reward = 1) and $F_i(t)$ failures (reward = 0) in $k_i(t) = S_i(t) + F_i(t)$ plays of arm i, the algorithm updates the distribution on μ_i as Beta$(S_i(t) + 1, F_i(t) + 1)$. The algorithm then samples from these posterior distributions of the μ_i's, and plays an arm according to the probability of its mean being the largest.

Algorithm 1 Thompson Sampling for Bernoulli bandits

For each arm $i = 1, \ldots, N$, set $S_i = 0, F_i = 0$
for $t = 1, 2, \ldots,$ **do**
  For each arm $i = 1, \ldots, N$, sample $\theta_i(t)$ from the Beta$(S_i + 1, F_i + 1)$ distribution.
  Play arm $i(t) := \arg\max_i \theta_i(t)$ and observe reward r_t.
  If $r = 1$, then $S_i = S_i + 1$, else $F_i = F_i + 1$.
end

The Bernoulli Thompson sampling algorithm is adapted to the general stochastic bandits case, i.e. when the rewards for arm i are generated from an arbitrary unknown distribution with support $[0, 1]$ and mean μ_i, in a way that allows to reuse the analysis of the Bernoulli case. The TS is adapted so that after observing the reward $\tilde{r}_t \in [0, 1]$ at time t, it performs a Bernoulli trial with success probability $\tilde{r}_t$. Let random variable r_t denote the outcome of this Bernoulli trial, and let $\{S_i(t), F_i(t)\}$ denote the number of successes and failures in the Bernoulli trials until time t. The remaining algorithm is the same as for Bernoulli bandits.

3 Contextual Bandit Setup

There are N arms. At time $t = 1, 2, \ldots$, a context vector $b_i(t) \in \mathbb{R}^d$, is revealed for every arm i. These context vectors are chosen by an adversary in an adaptive manner after observing the arms played and their rewards up to time $t - 1$, i.e. history $\mathcal{H}_{t-1}$,

$$\mathcal{H}_{t-1} = \{a(\tau), r_{a(\tau)}(\tau), b_i(\tau), i = 1, \ldots, N, \tau = 1, \ldots, t - 1\},$$

where $a(\tau)$ denotes the arm played at time τ. Given $b_i(t)$, the reward for arm i at time t is generated from an (unknown) distribution with mean $b_i(t)^T \mu$, where $\mu \in \mathbb{R}^d$ is a fixed but unknown parameter.

$$\mathbb{E}\left[r_i(t) | \{b_i(t)\}_{i=1}^N, \mathcal{H}_{t-1}\right] = \mathbb{E}\left[r_i(t) | b_i(t)\right] = b_i(t)^T \mu.$$

An algorithm for the *contextual bandit problem* needs to choose, at every time t, an arm $a(t)$ to play, using history $\mathcal{H}_{t-1}$ and current contexts $b_i(t), i = 1, \ldots, N$. Let $a^*(t)$ denote the optimal arm at time t, i.e. $a^*(t) = \arg\max_i b_i(t)^T \mu$. And let $\Delta_i(t)$ be the difference between the mean rewards of the optimal arm and of arm i at time t, i.e.,

$$\Delta_i(t) = b_{a^*(t)}(t)^T \mu - b_i(t)^T \mu.$$

Then, the regret at time t is defined as

$$\text{regret}(t) = \Delta_{a(t)}(t).$$

The objective is to minimize the total regret $\mathcal{R}(T) = \sum_{t=1}^{T} \text{regret}(t)$ in time T. The time horizon T is finite but possibly unknown.

It is assumed that $\eta_{i,t} = r_i(t) - b_i(t)^T \mu$ is conditionally R-sub-Gaussian for a constant $R \geq 0$, i.e.,

$$\forall \lambda \in \mathbb{R}, \mathbb{E}\left[[e^{\lambda \eta_{i,t}} | \{b_i(t)\}_{i=1}^N, \mathcal{H}_{t-1}]\right] \leq \exp\left(\frac{\lambda^2 R^2}{2}\right).$$

This assumption is satisfied whenever $r_i(t) \in [b_i(t)^T \mu - R, b_i(t)^T \mu + R]$

4 TS for Contextual Bandits

Gaussian priors and Gaussian likelihood function are used to design the TS algorithm. The likelihood of reward $r_i(t)$ at time t, given context $b_i(t)$ and parameter μ are given by the Gaussian pdf $\mathcal{N}(b_i(t)^T \mu, v^2)$ where $v = R\sqrt{\frac{24}{\epsilon} d \ln \frac{1}{\delta}}$ with $\epsilon \in (0, 1)$ which is the parameter in the algorithm. Let

$$B(t) = I_d + \sum_{\tau=1}^{t-1} b_{a(\tau)}(\tau) b_{a(\tau)}(\tau)^T$$

$$\hat{\mu}(t) = B(t)^{-1} \left(\sum_{\tau=1}^{t-1} b_{a(\tau)}(\tau) r_{a(\tau)}(\tau) \right)$$

If the prior for μ at time t is given by $\mathcal{N}(\hat{\mu}(t), v^2 B(t)^{-1})$, it can be computed using the Bayes update rule that the posterior distribution at time $t+1$ is $\mathcal{N}(\hat{\mu}(t+1), v^2 B(t+1)^{-1})$. In the TS algorithm, a sample $\tilde{\mu}(t)$ from the distribution $\mathcal{N}(\hat{\mu}(t), v^2 B(t)^{-1})$ is generated and the arm i that maximises $b_i(t)^T \tilde{\mu}(t)$ is played.

Algorithm 2 Thompson Sampling for Contextual Bandits

Set $B = I_d$, $\hat{\mu} = 0_d$, $f = 0_d$
for $t = 1, 2, \ldots,$ **do**
  Sample $\tilde{\mu}(t)$ from distribution $\mathcal{N}(\hat{\mu}(t), v^2 B(t)^{-1})$.
  Play arm $a(t) := \arg\max_i b_i(t)^T \tilde{\mu}(t)$, and observe reward $r_{a(t)}(t)$.
  Update $B = B + b_{a(t)}(t) b_{a(t)}(t)^T$, $f = f + b_{a(t)}(t) r_t$, $\hat{\mu} = B^{-1} f$
end

The Gaussian likelihood function and prior for rewards are only used to design the Thompson Sampling algorithm for contextual bandits. The analysis of the algorithm allows for the models to be completely unrelated to the actual reward distribution. The assumptions on the reward distribution are only those assumed in Section 3, *i.e.*, the R-sub-Gaussian assumption.

Every step t of Thompson Sampling (both algorithms) consists of generating a d-dimensional sample $\tilde{\mu}(t)$ from a multi-variate Gaussian distribution, and solving the problem $\arg\max_i b_i(t)^T \tilde{\mu}(t)$. Therefore, even if the number of arms N is large (or infinite), the above algorithms are efficient as long as the problem $\arg\max_i b_i(t)^T \tilde{\mu}(t)$ is efficiently solvable. This is the case, for example, when the set of arms at time t is given by a d-dimensional convex set $\mathcal{K}_t$ (every vector in $\mathcal{K}_t$ is a context vector, and thus corresponds to an arm). The problem to be solved at time step t is then $\max_{b \in \mathcal{K}_t} b^T \tilde{\mu}(t)$, where $\mathcal{K}_t$.

5 Results

Theorem 1. *[5] For the stochastic contextual bandit problem with linear payoff functions, with probability $1-\delta$, the total regret in time T for Thompson Sampling is bounded by $O(\frac{d^2}{\epsilon}\sqrt{T^{1+\epsilon}}(\ln T d \ln \frac{1}{\delta}))$, for any $0 < \epsilon < 1$, $0 < \delta < 1$. Here, ϵ is a parameter used by Thompson Sampling.*

Remark. *[5] The parameter ϵ can be chosen to be any constant in $(0, 1)$. If T is known, one could choose $\epsilon = \frac{1}{\ln T}$, to get $O(d^2 T)$ regret bound.*

Remark. *[5] The regret bound in Theorem 1 does not depend on N, and is applicable to the case of infinite arms, with only notational changes required in the analysis.*

Consider the setting where each of the N arms is associated with a different d-dimensional parameter $\mu_i \in \mathbb{R}^d$, so that the mean reward for arm i at time t is $b_i(t)^T \mu_i$. This setting is a direct generalization of the basic MAB problem to d-dimensions. Thompson Sampling for this setting will maintain a separate posterior distribution for each arm i which would be updated only at the time instances when i is played. And, at every time step t, instead of a single sample $\tilde{\mu}(t)$, N independent samples will have to be generated: $\tilde{\mu}_i(t)$ for each arm i.

Theorem 2. *[5] For the setting with N different parameters, with probability $1-\delta$, the total regret in time T for Thompson Sampling is bounded by*
$O(d\sqrt{\frac{NT^{1+\epsilon}\ln N}{\epsilon}}(\ln T \ln \frac{1}{\delta}))$

The proofs of the above theorems are fairly involved. Hence, an outline of the proof is given in the following section.

6 Outline of Proof

In the basic MAB problem there are N arms, with mean reward $\mu_i \in \mathbb{R}$ for arm i, and the regret for playing a suboptimal arm i is $\mu_{a^*} - \mu_i$, where a^* is the arm

with the highest mean. Comparing this to a 1-dimensional contextual MAB problem, where arm i is associated with a parameter $\mu_i \in \mathbb{R}$, but in addition, at every time t, it is associated with a context $b_i(t) \in \mathbb{R}$, so that mean reward is $b_i(t)\mu_i$. The best arm $a^*(t)$ at time t is the arm with the highest mean at time t, and the regret for playing arm i is $b_{a^*(t)}(t)\mu_{a^*(t)} - b_i(t)\mu_i$.

In general, the basis of regret analysis for stochastic MAB is to prove that the variances of empirical estimates for all arms decrease fast enough, so that the regret incurred until the variances become small enough, is small. In the basic MAB, the variance of the empirical mean is inversely proportional to the number of plays $k_i(t)$ of arm i at time t. Thus, every time the suboptimal arm i is played, we know that even though a regret of $\mu_{i^*} - \mu_i \leq 1$ is incurred, there is also an improvement of exactly 1 in the number of plays of that arm, and hence, corresponding decrease in the variance. The techniques for analyzing basic MAB rely on this observation to precisely quantify the exploration-exploitation trade-off. On the other hand, the variance of the empirical mean for the contextual case is given by inverse of $B_i(t) = \sum_{\tau=1:a(\tau)=i}^{t} b_i(\tau)^2$. When a suboptimal arm i is played, if $b_i(t)$ is small, the regret $b_{a^*(t)}(t)\mu_{a^*(t)} - b_i(t)\mu_i$ could be much higher than the improvement $b_i(t)^2$ in $B_i(t)$.

This difficulty is overcome by dividing the arms into two groups at any time: saturated and unsaturated arms, based on whether the standard deviation of the estimates for an arm is smaller or larger compared to the standard deviation for the optimal arm. The optimal arm is included in the group of unsaturated arms. At any time step t, the two groups are given by:

- *saturated arms* defined as those with $g(T)s_i(t) < \ell(T)s_{a^*(t)}(t)$,
- *unsaturated arms* defined as those with $g(T)s_i(t) \geq \ell(T)s_{a^*(t)}(t)$,

where $s_i(t) = \sqrt{b_i(t)^T B(t)^{-1} b_i(t)}$ and $g(T), \ell(T) (g(T) > \ell(T))$ are constants defined as $g(T) = v\sqrt{4d \ln Td} + \ell(T)$ and $\ell(T) = R\sqrt{d \ln T^3 \ln 1/\delta} + 1$. $s_i(t)$ is the standard deviation of the estimate $b_i(t)^T \hat{\mu}(t)$ and $vs_i(t)$ is the standard deviation of the random variable $b_i(t)^T \tilde{\mu}(t)$.

For the saturated arms, standard deviation is small, or in other words, the estimates of the means constructed so far are quite accurate in the direction of the current contexts of these arms, so that the algorithm is able to distinguish between them and the optimal arm. It can be shown that for the unsaturated arms, the regret on playing the arm can be bounded by a factor of the standard deviation, which improves every time the arm is played. Using concentration bounds for $\tilde{\mu}(t)$ and $\hat{\mu}(t)$ to bound the regret at any time by $g(T)(s_{a^*(t)}(t) + s_{a(t)}(t))$. Now, if an unsaturated arm is played at time t, then using the definition of unsaturated arms, the regret is at most $\frac{2g(T)^2}{\ell(T)} s_{a(t)}(t)$. This can be used along with the inequality from Auer at al. (2002), $\sum_{t=1}^{T} s_{a(t)}(t) = O(\sqrt{Td \ln T})$, to bound the regret due to unsaturated arms.

For saturated arms, it can be shown that the probability of playing a saturated arm at any time t is within p of the probability of playing an unsaturated arm, where $p = \frac{1}{4e\sqrt{\pi T^\epsilon}}$. Defining $\mathcal{F}_{t-1}$ as the union of history $\mathcal{H}_{t-1}$ and the

contexts $b_i(t), i = 1, \ldots, N$ at time t, it can be proved that with high probability using the results of the concentration bounds earlier and the definitions of saturated and unsaturated arms,

$$Pr(a(t) \text{ is a saturated arm}|\mathcal{F}_{t-1}) \leq \frac{1}{p} Pr(a(t) \text{ is an unsaturated arm}|\mathcal{F}_{t-1}) + \frac{1}{pT^2}$$

Defining the process $(X_t; t \geq 0)$ as

$$X_t = regret(t) - \frac{g(T)}{p} 1_{\{a(t) \text{ is unsaturated}\}} s_{a^*(t)}(t) - \frac{2g(T)^2}{\ell(T)} s_{a(t)}(t) - \frac{2g(T)}{pT^2}$$

it can be shown that X_t is a super-martingale difference process adapted to the filtration $\mathcal{F}_t$. A sequence of random variables $(Y_t : t \geq 0)$ is called a super-martingale with respect to filtration $\mathcal{F}_t$, if for all t, Y_t is $\mathcal{F}_t$ measurable and for $t \geq 1$;

$$\mathbb{E}\left[Y_t - Y_{t-1}|\mathcal{F}_{t-1}\right] \leq 0$$

The Azuma-Hoeffding inequality for a super-martingale Y_t corresponding to a filtration $\mathcal{F}_t$ satisfying the bounded difference property $|Y_t - Y_{t-1}| \leq c_t$ for constants c_t for $t = 1, ..., T$ states that for any $a \geq 0$

$$Pr(Y_T - Y_0 \geq a) \leq \exp\left(\frac{-a^2}{2\sum_{i=1}^{T} c_t^2}\right)$$

It can be seen that the absolute value of each of the four terms in the definition of the process X_t above is bounded by $\frac{2g(T)^2}{p\ell(T)}$ and hence the super-martingale process defined by $Y_t = \sum_{w=1}^{t} X_w$ has bounded difference property with constants $\frac{8g(T)^2}{p\ell(T)}$. Thus the Azuma-Hoeffding inequality for super-martingales along with the inequality $\sum_{t=1}^{T} s_{a(t)}(t) = O(\sqrt{Td \ln T})$ can be used to obtain the desired high probability bound.

7 Summary

To recapture, Thompson Sampling is one of the oldest heuristics that emerged in 1933. It is a Bayesian approach to solving problems with an inherent exploration-exploitation trade-off and comes under the family of randomized probability matching algorithms. Though it is a Bayesian approach, it works for prior-free cases as well. Though several studies over the decades have empirically shown the efficacy of Thompson Sampling and its robustness to delayed feedback, theoretical understanding of it is limited. It was only recently that the order optimality of TS for stochastic multi-armed bandit was proved. In the case of contextual bandits, the algorithm presented in [5] is order optimal up to the order of $\sqrt{T}$. The algorithm presented is computationally efficient as long as the problem $\arg\max_i b_i(t)^T \tilde{\mu}(t)$ is efficiently solvable. There has been some recent work in solving this problem in a agnostic setting.

References

[1] T.L.Lai and H.Robbins, *Asymptotically Efficient Adaptive Allocation Rules*, Advances in Applied Mathematics 6,4-22 (1985)

[2] A. Garivier and O. Cappe, *The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond*, 24th Annual Conference on Learning Theory, 2011

[3] S. Agarwal and N. Goyal, *Analysis of Thompson Sampling for the Multi-armed Bandit Problem*, 25th Annual Conference on Learning Theory, 2012

[4] V. Dani, T.P. Hayes and S.M. Kakade, *Stochastic Linear Optimization under Bandit Feedback*, 21st Annual Conference on Learning Theory, 2008

[5] S. Agarwal and N. Goyal, *Thompson Sampling for Contextual Bandits with Linear Payoffs*, Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013.