ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

Global Feedback Based Online Learning through Distributed Multi Agent

Radha D [1], S Karthi Prem[2]

P.G. Student, Department of Computer Science Engineering, AMC Engineering College, Bangalore, India[1]

Assistant Professor, Department of Computer Science Engineering, AMC Engineering College, Bangalore, India[2]

ABSTRACT: In this paper, we develop online learning algorithmsthat enable the agents to cooperatively learn how to maximize the overall prize in scenarios where only noisy worldwide criticism is accessible without trading any data among themselves. We demonstrate that our calculations learning regrets—the misfortunes incurred by the algorithms due to uncertainty—are logarithmically expanding in time and thus the time average prize converges to the optimal average prize. Besides, we additionally delineate how the regret relies on upon the size of the action space, and we demonstrate that this relationship is impacted by the informativeness of the reward structure with respect to every agent's individual action. At the point when the overall reward is completely instructive, regret is shown to be straight in the aggregate number of actions of all the agents. Our logical and numerical results demonstrate that the proposed learning calculations fundamentally outperform existing internet learning arrangements as far as regret and learning speed. We represent how our hypothetical structure can be utilized as a part of practice by applying it to online Big Data mining using distributed classifiers.

KEYWORDS: Big Data mining, distributed cooperative learning, multiagent learning, multiarmed bandits, online learning, reward informativeness.

I. INTRODUCTION

In this paper, we consider a multi-agent decision making and learning problem, in which an arrangement of disseminated agents select activities from their own particular activity sets keeping in mind the end goal to augment the general framework reward which relies on upon the joint action of all agents. In the considered situation, agents don't have the foggiest idea about from the earlier how their activities impact the general framework prize, or how their impact might change progressively after some time. In this manner, keeping in mind the end goal to amplify the general framework reward, agents should progressively figure out how to choose their best activities after some time. Be that as it may, agents can just watch/measure the general framework execution and henceforth, they just get worldwide criticism that relies on upon the joint activities of all agents. Since individualized criticism about individual activities is missing, it is inconceivable for the agents to figure out how their activities alone influence the general execution without coordinating with each other. Be that as it may, on the grounds that agents are appropriated they can't convey and facilitate their activity decisions. Besides, agents' perceptions of the worldwide input might be liable to individual blunders, and along these lines it might be greatly troublesome for an agent to guess other agents' activities construct exclusively with respect to its own particular watched reward history.

Be that as it may, on the grounds that agents are appropriated they can't convey and facilitate their activity decisions. Besides, agents' perceptions of the worldwide input might be liable to individual blunders, and along these lines it might be greatly troublesome for an agent to guess other agents' activities construct exclusively with respect to its own particular watched reward history. The way that individualized input is missing, correspondence is impractical, and the worldwide criticism is loud makes the advancement of productive learning calculations which amplify the joint compensate extremely difficult. Essentially, the considered multi-agent learning situation varies fundamentally from the current arrangements in which agents get individualized prizes.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

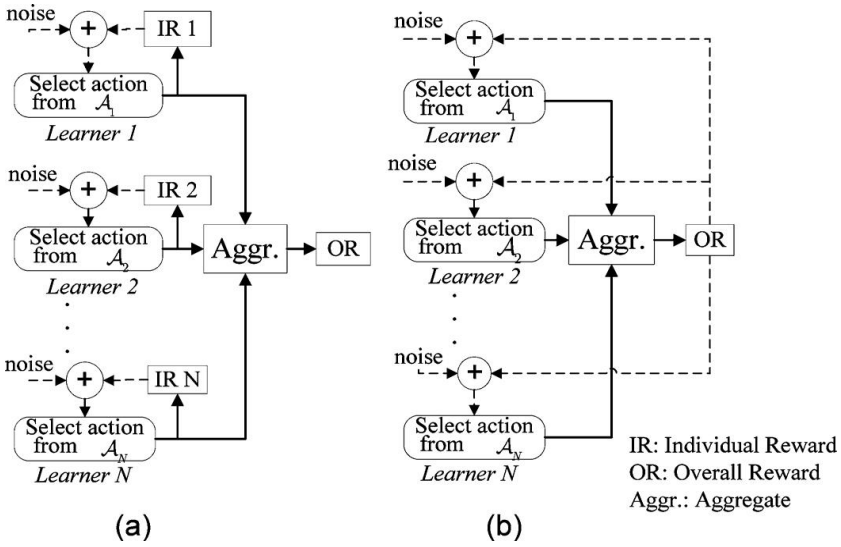

Fig.1. Learning in multi-agent systems with (a) individualized feedback; (b) global feedback

To represent the distinctions, Fig. 1(a) and (b) depict ordinary learning in multi-agent frameworks in light of individualized input and the considered learning in multi-agent frameworks taking into account worldwide criticism with individual clamor, separately.

II. RELATED WORKS

The writing on multi-furnished crook issues can be followed which ponders a Bayesian plan and requires priors over the obscure appropriations. In our paper, such data is not required. A general strategy in view of upper certainty limits is exhibited which accomplishes asymptotically logarithmic misgiving in time given that the prizes from every arm are drawn from a free and indistinguishably circulated (i.i.d.) process. It additionally demonstrates that no approach can show improvement over Ω (K ln t) (i.e., straight in the quantity of arms and logarithmic in time) and along these lines, this arrangement is request ideal as far as time.

In existing papers upper certainty bound (UCB) calculations are exhibited which are demonstrated to accomplish logarithmic lament consistently after some time, as opposed to just asymptotically. These strategies are appeared to be request ideal when the arm prizes are created freely of each other. At the point when the prizes are created by a Markov process, calculations with logarithmic misgiving concerning the best static strategy are proposed. Nonetheless, these calculations naturally expect that the prize procedure of every arm is free, and subsequently they don't abuse any relationships that may be available between the prizes of various arms. In this paper the prizes might be profoundly related, thus it is essential to plan calculations that consider this.

III. DISTRIBUTED COOPERATIVE LEARNING ALGORITHM

Description of the Algorithm

The DisCo algorithm is divided into phases: exploration and exploitation. Each agent using DisCo will alternate between these two phases, in a way that at any time t, either all agents are exploring or all are exploiting. In the exploration phase, each agent selects an arm only to learn about the effects on the expected reward, without considering reward maximization, and updates the reward estimates of the arm it selected. In the exploitation phase, each agent exploits the best (estimated) arm to maximize the overall reward.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

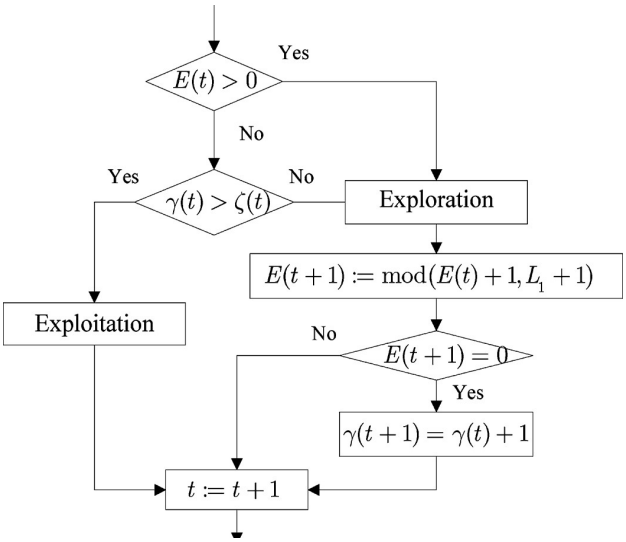

Fig. 2.Flowchart of the Phase Transition.

There is a deterministic control function $\zeta(t)$ of the form $\zeta(t)$= Aln t commonly known by all agents. This function will be designed and determined before run-time, and thus is an input of the algorithm. Each exploration phase has a fixed length of L= $\prod_{n=1}^{N} Kn$ slots, equal to the total number of joint arms. Each agent maintains two counters. The first counter$\gamma(t)$ records the number of exploration phases that they have experienced by time slot. The second counter E(t)€{0,1,…L_1} represents whether the current slot is an exploration slot and, if yes, which relative position it is at. Specifically, E(t)=0 means that the current slot is an exploitation slot; E(t)>0 means that the current slot is the E(t)-th slot in the current exploration phase. Both the counters are initialized to zero. Each Agent maintains sample mean reward estimates one for each relative slot position in an exploration phase. The reward estimates are initialized to be zero and will be updated over time.

IV. RESULTS

The proposed DisCo algorithm can also learn the optimal joint action. However, since the joint arm space is large, the classifiers have to stay in the exploration phases for a relatively long time in the initial periods to gain sufficiently high confidence in reward estimates while the exploitation phases are rare and short. Thus, the classification accuracy is low initially. After the initial exploration phases, the classifiers begin to exploit and hence the average accuracy increasesrapidly. Since the reward structure satisfies the Fully Informative condition, DisCo-FI rapidly learns the optimal joint action and performs the best among all schemes.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

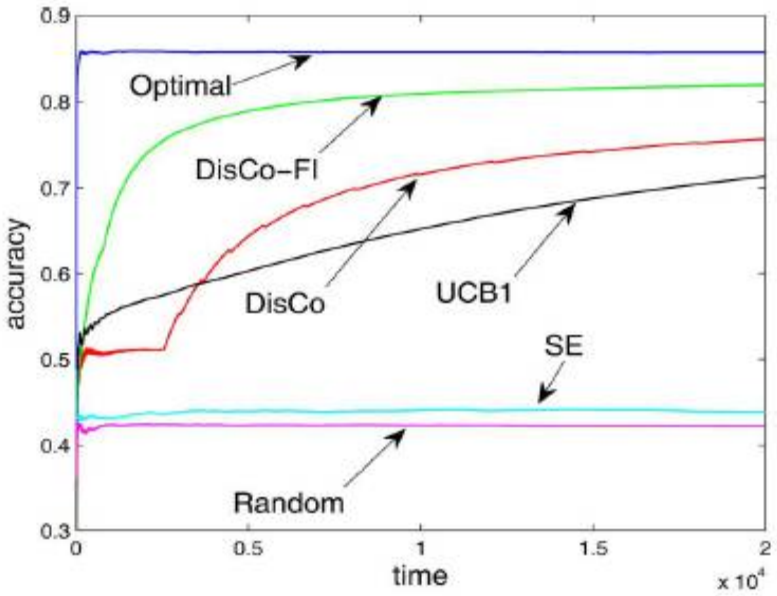

Fig.3. Performance comparison for various algorithms.

V. CONCLUSION

In this paper, we studied a general multi-agent decision making problem in which decentralized agents learn their best actions to maximize the system reward using only noisy observations of the overall reward. The challenging part is that individualized feedback is missing, communication among agents is impossible and the global feedback is subject to individual observation errors. We proposed distributed cooperative learning algorithm that addresses all these problems. Importantly, our theoretical framework can also be applied to learning in other types of multi-agent systems where communication between agents is not possible and agents observe only noisy global feedback. A more optimized approach can be provided with the multi agent system. An agent will be having several other multi agent which will work under one agent. This will help the user to get a wide range of feedback from the online sites as well as the workload on a single agent will be reduced.

REFERENCES

[1] M. Shah, J. Hellerstein, and M. Franklin, "Flux: An adaptive partitioning operator for continuous query systems," in Proc. Int. Conf. Data Eng. (ICDE), 2003.
[2] Y. Jiang, S. Bhattacharya, S. Chang, and M. Shah, "High-level event recognition in unconstrained videos," Int. J. Multimed. Inf. Retr., Nov. 2013.
[3] F. Fu, D. Turaga, O. Verscheure, and M. van der Schaar, "Configuring competing classifier chains in distributed stream mining systems," IEEE J. Sel. Topics Signal Process., vol. 1, no. 4, Dec. 2007.
[4] J. Xu, C. Tekin, and M. van der Schaar, "Learning optimal classifier chains for real-time big data mining," in Proc. 51st Ann. Allerton Conf. Commun., Contr., Comput., 2013.
[5] B. Foo and M. van der Schaar, "A distributed approach for optimizing cascaded classifier topologies in real-time stream mining systems," IEEETrans.ImageProcess., vol. 19, no. 11, pp. 3035–3048, Nov. 2010.
[6] B. Foo and M. van der Schaar, "A rules-based approach for configuring chains of classifiers in real-time stream mining systems," EURASIP J. Adv. Signal Process., vol. 2009, 2009.
[7] R. Ducasse, D. Turaga, and M. van der Schaar, "Adaptive topologic optimization for large-scale stream mining," IEEE J. Sel. Topics Signal Process., vol. 4, no. 3, Jun. 2010.
[8] J. C. Gittins, "Bandit processes and dynamic allocation indices," J. Royal Statist. Soc. Ser. B (Methodolog.), pp. 148–177, 1979.
[9] P. Whittle, "Multi-armed bandits and the Gittins index," J. Royal Statist. Soc. Ser. B (Methodolog.), 1980.
[10] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," Adv. Appl. Math., vol. 6, no. 1, 1985.