IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Future Research Directions in Crowd Computing: A Novel Perspective

Manas Kumar Yogi[1], Sirisha Ratnala[2]

Asst. Prof., CSE Department, Pragati Engineering College, Surampalem, A.P.,India [1]

B.Tech. III year Student, CSE Department, Pragati Engineering College, Surampalem, A.P.,India [2]

Abstract: This paper discusses areas where crowd computing needs attention to optimize crowd resources as well as increase the efficiency of crowd computing. We explore only three horizons namely user modeling,labels integration, sample selection. We observe that each one is a research area in its own where work has been progressing at appreciable rate.But they are open issues, so scope of improvement exists in all the areas of crowd computing. Our paper is a sincere attempt to bring out the shortcomings of current strategies used in development of crowd computing applications. This paper will serve as a readymade guide for researchers who attempt to treadin the path of crowd computing research.

Keywords: Crowd sourcing, Crowd computing, model selection, label classification, uncertainty.

I. INTRODUCTION

Human beings are capable of solving non-algorithmic issues. Crowd computing approach exploits this strength where computers cannot perfectly solve the problem. Human group can efficiently solve few areas of computing where machine intelligence cannot outperform humans. Humans perception in certain issues have better approach which cannot be installed in computer systems due to natural behaviour of human beings. The term crowd sourcing was advocated by Jeff Home in the year 2006. The operating principle is simple. A group of people are asked to perform a task to contribute to a complex task which cannot be finished by a single person. For instance, Wikipedia which is one of the most popular crowd-computing system where daily millions of users are contributing to the content on various topics all over the world. There are many benefits of crowd-computing. Business organisations reap innumerable benefits from users feedback to improve their services. Classical AI systems have some inherent shortcomings. For example, the optical character recognition (OCR) perform poorly when it comes to low quality of characters. Using crowd-computing recaptcha system is built to serve the purpose . Two different OCR systems are used along with a reference directory to achieve a valid authentication. According to experimental results by Newyork Times Archive, the Recaptcha system acheived 99% precision against 84% of standard OCR systems. One of the remarkable thing of RecaPtcha system is that the crowd does not charge any money for their effort in completing the task. Yet another significant crowd computing marketplace is Amazon Mechanical Turk(MTurk). It has Provision for API's for developers so that the developers can directly connect to MTurk servers to efficiently finish the computing task. MTurk largely popular due to large number of members ,high-diversity of member's knowledge , locations ,skills along with low-cost labors. The rapid cycle of deployment and testing also matters. Crowd-computing applications should solve problems which have following characteristics. First of all problem divisibility. The problem should be divisible into non dependant sub Problems. They should not change with time. The sub Problems when solved should result into sub-solution which should be in a verifiable state. There should be a strategy that should b efficient enough to integrate sub solutions into solution to the original-Problem. second characteristic is cost of crowd-computing should be reasonable . A crowd size limited to few users who are expert in solving a specific computing problem will increase the cost inevitably. Hence this characteristic should be kept in mind while modelling a crowd-computing scenario . Numerous applications in recent years used this crowed sourcing-approach : Music similarity evaluation , Improvement in text - writing ,Measurement of relevance of results by search engines, Construction of training datasets of audio ,video, images for classic AI systems. There are 3 design steps for designing a crowd-computing system. They are : defining system overall strategy , generation of sub problems designing & optimising process . In crowd-computing, it’s a good Practise that users compete to improve the sub-problems ,thereby leading to generation of optimised sub solutions. For the designers one crucial task is Problem (or) sub Problem assignment depending on two types of computing systems . First is for active systems , each user can be modeled by its history upon which types of sub Problem can be assigned to that users. Whereas passive systems there is no task assignment . Users select task based on their skill set and time available to solve that task . Hence , criteria to select a task plays an important role. The following shows crowd computing model.

Human crowd workers have more knowledge base and consciousness than machines which gives them an edge over machines . Also , it has been experimentally Proved that human beings can in parallel learn algorithms than machine . A machine handles sequential logic as well as the operating system should be changed in a computer which is cost efficient for solving a big computing challenge

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

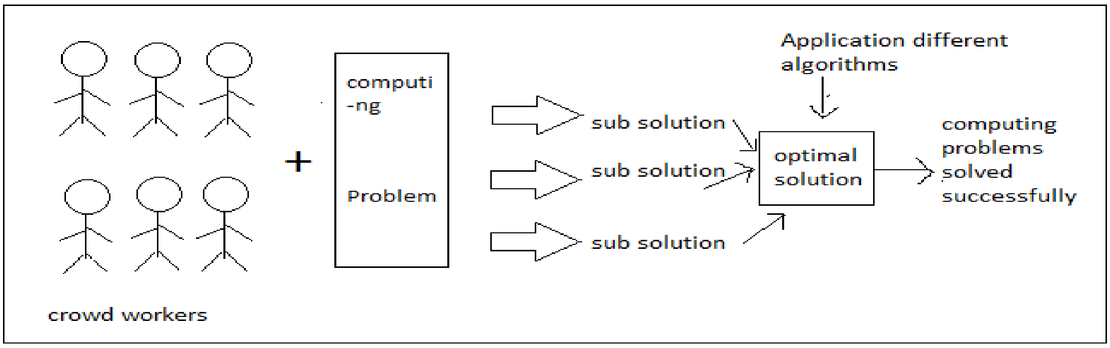

Figure 1.Scemhatic representation of crowd computing

So, a crowd is a best choice in such cases. consciousness in humans helps in accessing original information when compared to a AI system where the knowledge is derived as a reflection of human consciousness , the availability of memory in huge quantity is hindering its viability . Similarly common-sense knowledge requires application of human vision, Perception, feeling which cannot be imprinted in a machine. creating a common sense knowledge -base is difficult due to its huge size in demand , no efficient , method to represent properly the knowledge , updation operation will be expensive , inferencing from knowledge base is more complex . In this paper , we are presenting a clear vision of the research challenges for the researchers and other crowd computing practitioners. We broadly categorize the problems into three groups and present few novel guidelines which can be carried out for development of crowd-computing applications.

II. FUTURE RESEARCH DIRECTIONS

Till now, not much is done on adaptation scenarios in crowd-computing. The sample selection criteria are not effective when user models are considered. With impression monitory budget choice of user model does not have an impact nut considering low budget, cost models have to be designed effectively. Erogenous results also have to be bounded and this area needs realistic approaches to calculate the margin of error which can be gracefully accepted. The subsequent sub-sections of the Paper envision major open issues currently lurking in the world of crowd-computing.

2.1 User-modeling:

User modeling should be based on diverse factors. The most difficult being a realistic time-varying user model. Most of the current user model have applied Bayesian model for efficient label classification. But time factor is not considered in such models. The labels are static and do not evolve until task is finished . Labels cannot be modified during generation of sub-solutions. We need more efficient user models which learn the fact that has time passes ,the labels should also be modified by tagging them with sub-tags. If labels are modified then it will produce more sub-solutions thereby generating scope foe even more better computing solution. thereby generating scope for even more better computing solutions . Work is to be directed towards investigation of application of community-based aggregation approach foe dynamic label creation as well as ranking of the labels. There is scope of improvement in Performance of time-varying user models of upto 8-10% maximum when compared to the best modelling technique used now.

2.2 Labels Integration :

The second challenge is for the researchers to detect and filter out low-quality labels from user-models .Existing techniques use majority voting, entropy , accuracy , uncertainty , Again no such efficient technique is present where labels integration approach is decided . Current mechanisms are able to differentiate the labels per accuracy is considered but we need even more better techniques like label integration to exploit the robustness of labels. If we consider the entropy model where designers of crowd sourcing system have placed selection criteria according to computation-time to finish a task, then the crowd workers will Pick a task according to the skills they posses and they would not consider labels (or) integration of labels . Thus, in a way current approaches lack the amount of dynamism given by designers of the system as they restrict themselves to the skill set of themselves . Label integration is challenging because users have a definite feeling about certain issues . For instance few users of crowed sourcing may always Pick a negative selection which reduces a label can be created . This will be a first step towards label integration. But detection of label community is not so easier . This is due to the fact that prediction-models which group labels into communities are nit robust and cost-effective. so, in an essence we infer that Prediction - models should have less uncertainty and more accuracy while label community formation begins. Researchers have their task easy if this is done effortlessly.

2.3 Sample Selection:

Depending on the budget, it was found after experiments that higher budget resulted in better sample selection. The criteria for sample selection is again depends on average number of labels per each sample as well as uncertainty.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Most of user models consider correctness probability of estimated labels but these user models suffer from short sightedness. They do not consider the overall performance in individual step. So ,currently it is an open issue to improve this shortcoming of sample selection. In addition to these more research needs to be carried out on importance and impact of factors like exploration, exploitation, deterministic or proportional random sample selection. Exploration, exploitation refers to amalgamation of user models where objective function is to enhance the selection strength by considering hybrid factors like socio-economic status of the user along with degree of skills to solve the task.

III. CONCLUSION

This paper utilizes the current open issues in crowd computing to form a novel perspective which can be further developed to enhance the efficiency of crowd computing. But it is not a easy task due to varying amount of scenarios which a crowd encounters .Due to this uncertain factors other suite of metrics have been developed but their suitability has still to be proven right. For that case our paper presents a vision to crowd computing researchers. This paper illustrates how a shortcoming in user model selection or label classification may lead to decrease in the desired performance of a crowd sourcing application. We conclude this paper by underlying the design principles where root level changes have to be used not only in limited scenarios but also in areas where AI is expensive to invest in. Researchers are currently working in similar directions.

REFERENCES

[1] Little, G., Chilton, L. B., Goldman, M. and Miller, R. C. (2009), TurKit: tools for iterative tasks on mechanical turk, Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP 09, ACM, Paris, France, pp 29–30.
[2]von Ahn, L. and Dabbish, L. (2004), Labeling images with a computer game, Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '04, ACM, Vienna, Austria, pp 319–326.
[3] Donmez, P., Carbonell, J. G. and Schneider, J. (2010), A probabilistic framework to learn from multiple annotators with time-varying accuracy, SDM, SIAM, pp 826–837.
[4] Frank, E. and Hall, M. (2001), A simple approach to ordinal classification, ECML 01: Proceedings of the 12th European Conference on Machine Learning, Springer-Verlag, London, UK, pp 145–156.
[5] Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L. and Moy, L. (2010), Learning from crowds, Journal of Machine Learning Resources, 99, MIT Press, Cambridge, MA, USA,pp 1297–1322.