ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Efficient Query Optimization for Easy Retrieval of Crowd Resources

G.Archana, Dr.P.Srinivasan

ME, Department of CSE, Muthayammal Engineering College, Rasipuram, Namakkal, India

Professor, Department of CSE., Muthayammal Engineering College, Rasipuarm, Namakkal, India

ABSTRACT: Declarative crowdsourcing is intended to cover the complexities and relieve the user of the burden of addressing the group. The user is simply needed to submit an SQL-like question and therefore the system takes the responsibility of assembling the question, generating the execution set up and evaluating within the crowdsourcing marketplace. A given question will have many different execution plans and therefore the distinction in crowdsourcing value between the most effective and therefore the worst plans could also be many orders of magnitude. Previously, we tend to proposed CROWDOP, a cost-based question optimization approach for declarative crowdsourcing systems. We incline to develop efficient algorithms within the CROWDOP for optimizing 3 forms of queries: choice queries, join queries, and complicated selection-join queries. In this paper we propose, extract the information from the CROWDOP using preprocessing. Suggestions given by the statistical analysis of car buyers. This suggestions includes the various car's attributes like, price, model, colour, speed, fuel type.It consists of system performance and query suggestions for consumer liking.

KEYWORDS: Declarative crowdsourcing, Query Optimization, CROWDOP

I.INTRODUCTION

Data mining is the process of evaluating data from different perspectives and summarizing it into useful information. It is the process of collecting, searching through, and analysing a large amount of data in a database, as to discover patterns or relationships. Data preprocessing describes any type of processing performed on raw data to prepare it for another processing procedure. Normally used as ainitial data mining practice,data preprocessing converts the data into a format that will be more easily and effectively processed for the purpose of the user. Data mining is also known as Knowledge Discovery in Data. The Knowledge discovers from the CROWDOP.

CROWDSOURCING has attracted increasing interest in recent years as an efficient tool for harnessing human intelligence to solve issues that computers cannot perform well, like translation, handwriting recognition, audio transcription and picture tagging. Consequently, for a given question, a declarative system should first compile the question, generate an in an execution plan, post human intelligence tasks (HITs) to the team according to the arrange, collect the answers, handle errors and decide the inconsistencies within the answers.

This paper involves extract the information from the CROWDOP using preprocessing technique. Data preprocessing methods are divided into following categories, Data Cleaning,Data Integration, Data Transformation and Data Reduction The system used the information and extract the required information based on the parameters like car's model, cost, company, speed, fuel type etc. using crowd optimization algorithms, named Optimization Framework, Optimization Select, Optimization Join, Generate Parse tree, Latency Bound Optimization. Then implement the values into all algorithms, and analyse the customer requirements. The evaluation process implemented based on the customer analysis.Suggestions given by the statistical analysis of car buyers. Query suggestions will be given to the user based on the analysis of the customer requirements.This is used to the customer to purchase the best product.

II. RELATED WORK

In order to improve the performance of the system, various query suggestions methods with different features have been proposed. The problem of evaluating top-k and group by queries using crowd to answer either type or value

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

questions developed in Using the crowd for top-k and group by queries by S.B Davidson[2] Also includes Objective of minimizing the number of comparisons performed by the crowd to find the difficult exact top-k elements or the exact clusters.J.Fan proposed, A Hybrid Machine-Crowd sourcing System forMatching Web Tables used toTwo-pronged approach for web table matching that effectively addresses the incompletenessin web tables.We made a simplification that the crowd was assumed to produce perfect answer, which is not always the case.It occurs Low crowd accuracy into account in the model[3]. J Gao proposed the method name isAn Online Cost Sensitive Decision-Making Method inCrowd sourcing Systems. He introduce a linear model for online decision making. We first estimate the accuracy of the answers according to the question status and prior distribution. Based on the estimation of the answer accuracy, we obtain the marginal income and the profit of each status. This enables us to make decisions at each status according to the economic profit [4]. CrowdDB: Answering Queries with Crowd sourcingdeveloped by M. J. Franklin,CrowdDB uses human input via crowd sourcing to process queries that neither database systems nor search engines can adequately answer[5].JU Fan propose the Crowdop, a cost-based question optimization approach for declarative crowdsourcing systems. We incline to develop efficient algorithms within the CROWDOP for optimizing 3 forms of queries: choice queries, be a part of queries, and complicated selection-join queries[1].
N. Polyzotis, and J.Widom, proposed the Deco: Declarative crowdsourcing, for enablesprogrammers to incorporate "human computation" as a building block in algorithms that cannot be fullyautomated, such as text analysis and image recognition.[7].H. Park and J. Widom, declares the Query optimization over crowdsourced datafor answering declarative queries posed over stored relational data together with data obtained on demand from the crowd[8].J.M.Hellerstein and M.Stonebraker suggested the system for predicate migration: optimizing queries with expensive predicates ,involves the theory for moving exclusive builds in a query plan.The total cost of the plan counting the costs of both joins and boundaries is minimal [9]. C.J.Ho,Jabbari develops the adaptive task assignment for crowd sourced classification which is the problem of task project and label inference to various classification tasks. In applying online primal-dual techniques[10].Counting with the crowd established by the A.Marcus,D.R. Karger and R.C Miler for discernment estimation for crowdsourced catalogue, used to spammer detection technique for label and count based method.It reduces the downstream monetary cost and latency [11]. Also developed the crowdsourced databases: query processing with people,includes the number of query completing and optimization experiments and propose the novel system for managing the challenges. For managing and writing the SQL queries are too complex. There are several decisions and opportunities in the future [12]. A.G. Parameswaran and H. Park established the crowd screen: algorithm for filtering data with humans comprises deterministic andprobabilistic algorithms to optimize the likely cost and expected error. Applied in a variety of crowdsourcing scenarios. In future work includes integratinghuman accurate, behaviour correlations among filters and the multiple filters, outspreading the techniques in the classification and collecting problems, and attempting to determination the question of shapes are ideal for deterministic approaches [13].P. Venetis and J. Feng developed Max algorithms in crowdsourcing environments explores the problem of recovering the determined item from a set in crowdsourcing environments[14].Question selection for crowd entity selection developed by S.E.Whang,to evaluate our best question algorithms on real and imitation datasets [15].

III. SYSTEM ARCHITECTURE

The new user wants to buy a car in show room open the appropriate web page for that car show room. Select the various option for car make, model, price, color, etc. and submit.Analyse the system for purchase the product. There are lot of brands in crowd. The user select the query for demand products. The gatherings all requirements for query optimizer select query based generate the query optimized plan. Fill the query to complete their expectations. Join query to combine all the results. The user select option based retrieves the result. They also consider the latency rating for product. The latency rating has been calculated for collect the optimize user comments and parse tree based split positive and negative comments. If less negative comments product for high latency rating. The rating based display the user result. Easily acquires the result from the crowd based on the ratings of the products like, brand, style, make, price etc.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

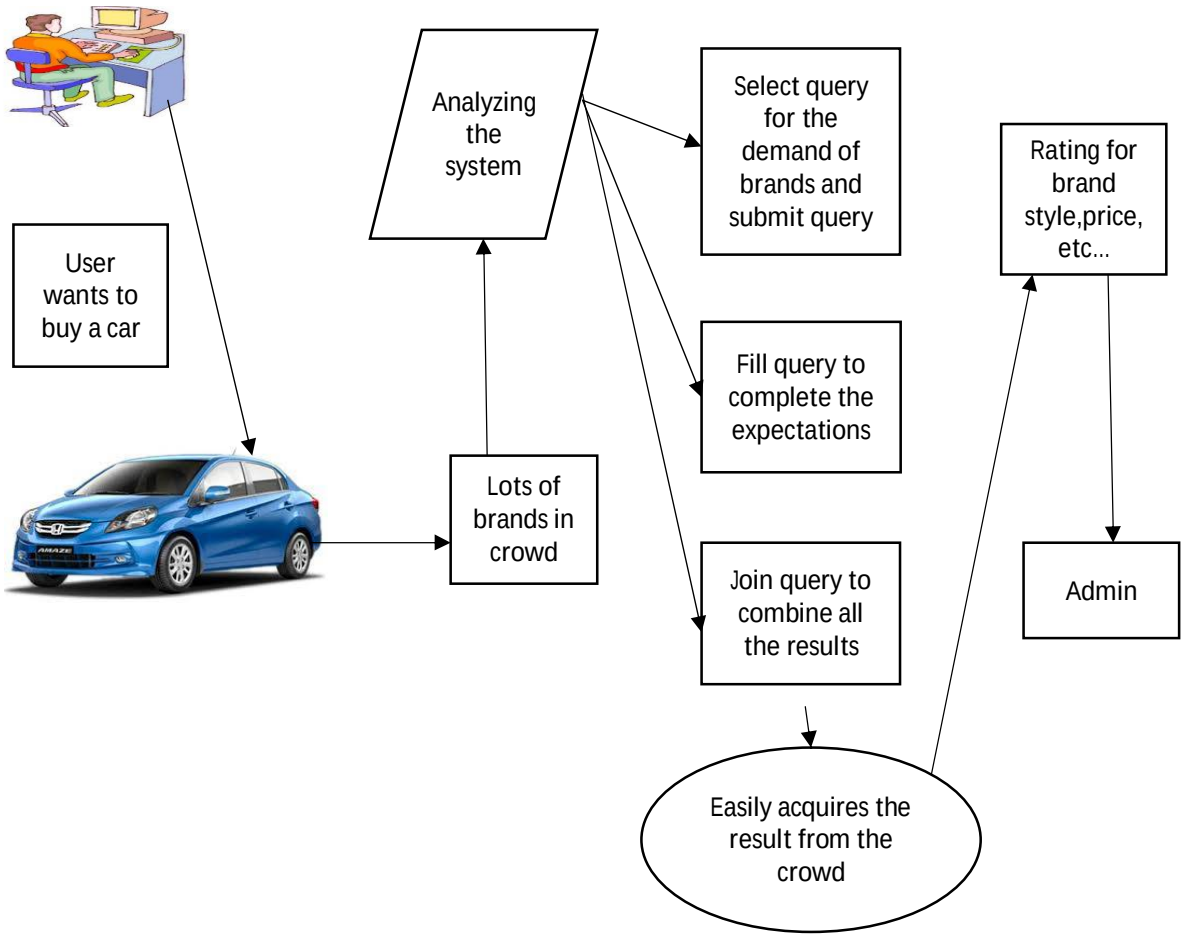

Fig 1.System Architecture

IV. CROWDOP DATA MODEL AND QUERY LANGUAGE

Data model CROWDOP employs relative information model, like previous work on crowdsourcing systems [3], [12], [15]. In CROWDOP, the info is nominal as a schema that consists of a set of relations R= {R1, R2, R|R|}. These relations area unit designated by schema designers and may be queried by crowdsourcing users. Fig. one provides an example schema with 3 relations. Different from ancient databases, some attributes of tuples are unknown before execution crowdsourcing.

1) Optimal query: A variety question applies one or additional human-recognized choice conditions over the tuples in a very single relation. Choice question has several applications in real crowdsourcing situations, such as filtering information [14] and finding bound things [18].

2) Connection query: A connection question leverages human intelligence to combine tuples from two or additional relations in keeping with certain be a part of conditions. One typical application of be a part of question is crowdsourcing entity resolution, that identifies pairs of records representing constant real-world entity. Other applications embody subjective classification (e.g., sentimental analysis) and schema matching.

3) Advanced (selection-join (SJ)) question: CROWDOP supports more general queries containing each choices and joins. These queries will facilitate user's categorical additional advanced crowdsourcing requirements. The CROWDOPsystem includes,

- Query optimized Plan

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

- Identifying Non-filling Option
- Latency Bound Optimization
- Optimized work for crowd sourcing executor

A. *Query optimized plan*

The new user buy a car in show room open the web page select the various option for car make, model, price, colour, etc and submit. The gatherings all requirements for query optimizer select query based generate the query optimized plan.

B. *Identifying non-filling option*

The query optimizer gathering all query analyze the all the filling and non- filling option. Already set each option set default value, so if some option has not filling set default value. The join query used joins all select queries.

C. *Latency bound optimization*

The user select option based retrieves the result. They also consider the latency rating for product. The latency rating has been calculated for collect the optimize user comments and parse tree based split positive and negative comments. If less negative comments product for high latency rating. The rating based display the user result.

D. *Optimized work for crowd sourcing executor*

Finally crowd sourcing executer join and fill query and also latency rating based retrieve the optimized result display from user.

Algorithm for optimization framework

OPTFRAMEWORK(Q,Cost)
{
if Cost=Nil then
P← CostOpt(Q)
else
L_{min} ←COMPUTEMINLATENCY(Q)
L_{max} ←COMPUTEMAXLATENCY(Q)
While $L_{min} \geq L_{max}$do
L← (L_{max}+ L_{min})/2
P←LATENCYBOUNDOPT (Q,L)
If P.cost ≤ Cost then
L_{max} ←L-1
else
L_{min} ←L+1
Return P
}

Algorithm for optimization selection

OPTSELECT(Cost,L)
{
Sort c in increasing order of selectivity;
For i=1…|Cost| do
G(i,1)← cost (i,|C|)
for j=2…L do G(i,j)
G(|C|,j)←cost(|C|,|C|)
For i=1…|C|- 1 do
G(I,j)← $min_{K=i}$ {cost(I,k)+G(k+1,j-1)}
P[i][j]←k with the minimum cost
Generate query plan P w.r.t G(1,L)from p[][]
Return p
}

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

V. PERFORMANCE MATRICS

Monetary price. The monetary price of query strategy Q, represented by cost (Q), is that the overall rewards obtained for executing all crowdsourcing operators in the query plan P_Q.The cost of an operator depends on the price given to crowd for each query produced by the operator.
Latency. As crowdsourcing takes time, latency is obviously introduced to enumerate the quickness of question analysis. However, it is non-trivial to calculate and enhance latency. On one hand, the crowdsourcing tasks arefinished by a pool of theemployees in parallel and also the size of the pool is usually moving based on the employee. In different words, employees could enter or leave the pool at any time. On the opposite hand, the latency may be affected by the different crowdsourcing tasks. There are a restricted number of crowdsourcing employees on public crowdsourcing platforms comparable to AMT. The employees are able to take the HITs that interest them, comparable to those with high reward. In that sense, printed crowdsourcing tasks vie against one another for the employees. As such, the latency in tangible crowdsourcing development has several uncertainties.

Accuracy. Crowdsourcing could yield comparatively low-quality results or maybe noise, if there are spammers or cruel workers. Thus, accuracy is occupied as another necessary performance metric to live the standard of crowdsourcing results. In our CROWDOP system, we tend to address the accuracy issue by using our previous work on internal control as a building block. Specifically, the standard management model consists of a analyst and verifier. Given a required accuracy, the predictor estimates the amount of workers that area unit required to realize the necessity based mostly on the worker's average accuracy. If no average accuracy is available within the system, a default price three is employed. The admirer is to resolve the inconsistencies within the results came back by totally different employees and choose the simplest answer. A probability- based verification model is adopted if every worker's historical performance is monitored, otherwise a straightforward voting- based strategy is employed. During this paper, we tend to target finding out the cost-latency optimization issues whereas presuming the accuracy issue has been adequately self-addressed. Table1 describes the different query suggestions for the product like car. It includes Based on the cost, Based on the rating, and Based on the model and based on the colour. First table describes the rating of the product. Second table describes model of the product. Next two table describes the product cost and model. Fig 2.a describes the query suggestions for the product based on the cost. Tata cars takes more place in the graph. Fig 2.b denotes the query suggestions for the product based on the rating. In this graph swift cars are placed in a high views and buy. Fig 2.c and fig 2.d describes the query suggestions for the product based on the colour and model.

TABLE 1
DATASET DESCRIPTION

SESSION	MODEL	VIEW	BUY	RATING
1	TATA	100	40	3
2	SKODA	100	20	4
3	HYNDAI	100	50	2
4	BMW	100	10	5
5	SWIFT	100	60	1
6	DUSTER	100	10	6

SESSION	COMPANY	MODEL	VIEW	BUY
1	VOLVO	S80	100	40
2	TOYOTA	AVALON	100	20
3	VOLVO	XC60	100	50
4	TOYOTA	COROLLA	100	10
5	BMW	X5	100	60
6	TOYOTA	CAMRY	100	10

SESSION	MODEL	VIEW	BUY	COST
1	TOYOTO	50	20	4 LAK
2	SKODA	70	40	2 LAK
3	BMW	20	10	7 LAK
4	TATA	80	50	3 LAK
5	HYNDAI	30	20	5 LAK

SESSION	MODEL	COLOUR	BUY
1	TOYOTO	BLACK	20
2	SKODA	WHITE	40
3	BMW	RED	10
4	TATA	GREEN	50
5	HYNDAI	BLUE	20

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

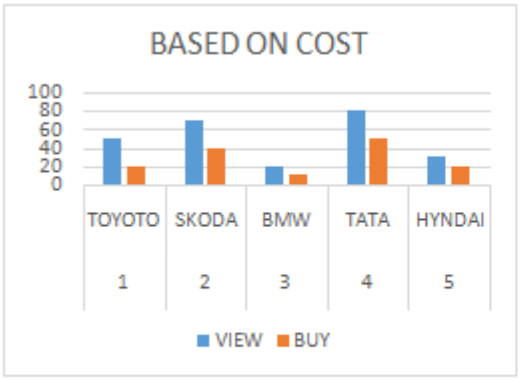

Fig 2.a. Based on cost

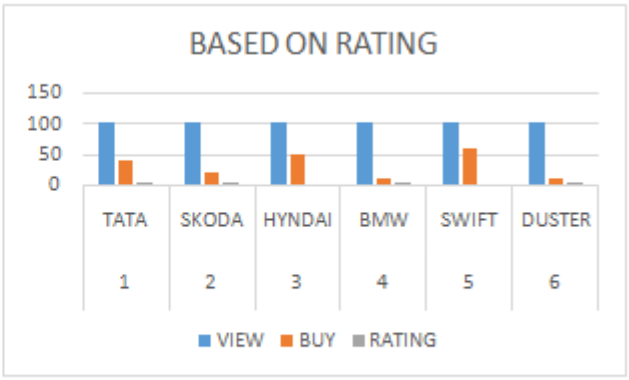

Fig 2.b. Based on Rating

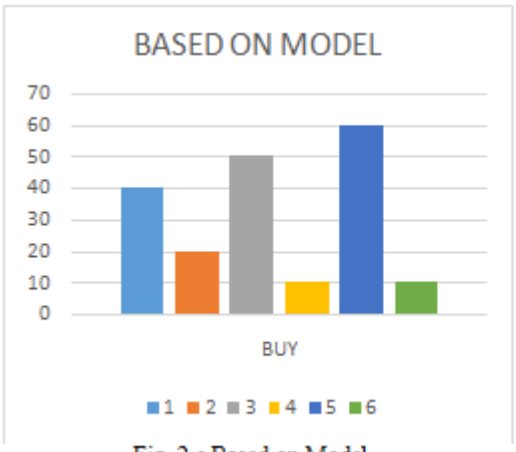

Fig. 2.c Based on Model

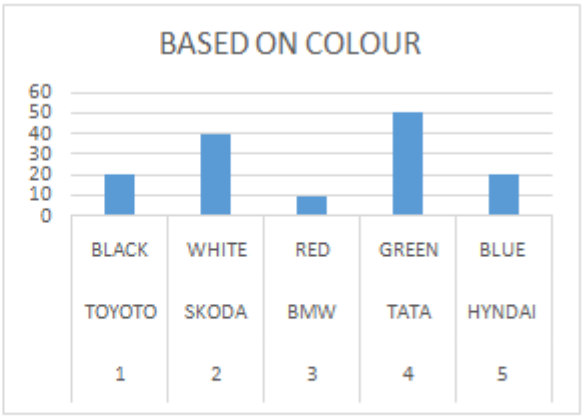

Fig.2.d Based on Colour

VI. QUERY OPTIMIZATION

We outline two improvement objectives thought-about during this paper. The primary one specially takes into consideration the financial cost and aims to seek out the foremost economical question arrange.

Objective one (Cost minimization). Given a question Q, it aims to find a question arrange PQ that minimizes the financial price,

i.e., P_Q= argP_Q minimum cost(P_Q)

Objective two (Cost delimited latency minimization). Given a query Q and a value budget C, it finds a question arrange $P_Q$with bounded price C (P_Q) C and therefore the minimum latency PQ=arg PQ min latency (PQ). If there are a multiple plans with the minimum latency, it finds the one with lowest price.

P_Q= argP_Q minimum latency (P_Q)

VII. CONCLUSIONAND FUTURE WORK

In query optimization, we have a tendency to propose a cost-based question optimization that considers the cost-latency exchange and supports multiple crowdsourcing operators. A declarative crowdsourcing query can be evaluated in many ways, the user have choice of query execution.we have a tendency to develop economical and effective optimization algorithms for choose, be a part of and complicated queries. Our experiments on each simulated and real crowd demonstrate the effectiveness of our question optimizer and validate our value model and latency model. In the future

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

we might wish to study the way to incorporate correlations between select/join conditions into the optimizer for compound queries, and that we additionally arrange to extend CROWDOP to support a lot of advanced SQL operators, such as to sorting and aggregation.

REFERENCES

[1]. Ju Fan, Meihui Zhang, Stanley Kok, Meiyu Lu, and Beng Chin Ooi. CrowdOp: Query Optimization for Declarative Crowdsourcing Systems , in knowledge and data engineering, vol. 27, no.8,august 2015.
[2]. S. B. Davidson, S. Khanna, T. Milo, and S. Roy. Using the crowd for top-k and group-by queries inProc. 16th Int. Conf. DatabaseTheory, 2013, pp. 225–236.
[3]. J. Fan, M. Lu, B. C. Ooi, W.-C. Tan, and M. Zhan. A hybrid machine-crowdsourcing system for matching web tables inProc.IEEE 30th Int. Conf. Data Eng., 2014, pp. 976–987.
[4]. M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin, CrowdDB: Answering queries with crowdsourcing in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 61–72.
[5]. J. Gao, X. Liu, B. C. Ooi, H. Wang, and G. Chen.An online cost sensitive decision-making method in crowdsourcing Systems inProc. ACM SIGMOD Int. Conf. Manage. Data,pp. 217–228, 2013.
[6]. Y. Gao and A. G. Parameswaran, Finish them!: Pricing algorithms for human computation Proc. VLDB Endowment, vol. 7, no. 14, pp. 1965–1976, 2014.
[7]. G. Parameswaran, H. Park, H. Garcia-Molina, N. Polyzotis, and J. Widom, Deco: Declarative crowdsourcing in Proc. 21st ACMInt. Conf. Inf. Knowl. Manage., 2012, pp. 1203–1212.
[8]. H. Park and J. Widom.Query optimization over crowdsourced data Proc. VLDB Endowment, vol. 6, no. 10, pp. 781–792, 2013
[9]. J. M. Hellerstein and M. Stonebraker, Predicate migration: Optimizingqueries with expensive predicates, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 1993, pp. 267–276.
[10]. C.-J. Ho, S. Jabbari, and J. W. Vaughan, “Adaptive task assignment for crowdsourced classification,” in Proc. 30th Int. Conf. Mach. Language, 2013, vol. 1, pp. 534–542.