Proceedings of the of the 17th Conference on Computer Science and Intelligence Systems pp. 301–304
DOI: 10.15439/2022F235
ISSN 2300-5963 ACSIS, Vol. 30

A chance-constraint approach for optimizing Social Engagement-based services

Michel Bierlaire*, Edoardo Fadda†, Lohic Fotio Tiotsop‡, and Daniele Manerba§

*Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne - CH-1015 EPFL, Lausanne (Switzerland)

†Dept. of Mathematical Sciences, Politecnico di Torino - corso Duca degli Abruzzi 24, Torino (Italy)

‡Dept. of Control and Computer Engineering, Politecnico di Torino - corso Duca degli Abruzzi 24, Torino (Italy)

§Dept. of Information Engineering, Università degli Studi di Brescia, via Branze 38, Brescia (Italy)

Emails: *michel.bierlaire@epfl.ch, edoardo.fadda@polito.it, lohic.fotiotiotsop@polito.it, daniele.manerba@unibs.it*

***Abstract*—Social Engagement is a novel business model transforming final users of a service from passive into active components. In this framework, people are contacted by a company and they are asked to perform tasks in exchange for a reward. This arises the complicated optimization problem of allocating the different types of workforce so as to minimize costs. We address this problem by explicitly modeling the behavior of contacted candidates through consolidated concepts from utility theory and proposing a chance-constrained optimization model aiming at optimally deciding which user to contact, the amount of the reward proposed, and how many employees to use in order to minimize the total expected costs of the operations. A solution approach is proposed and its computational efficiency is investigated through experiments.**

I. Introduction and Related Works

SOCIAL Engagement (SE) is a new business paradigm involving the customers of a company in its operations. More precisely, people agree to perform specific services in exchange for a reward. This model has been enabled by the increase of the number of users connected on the web and technologies able to get people information [1]. This gives to the companies the possibility to easily communicate with *candidates* and then to propose *tasks* in exchange for a reward.

A concrete application of the SE paradigm is the so called *crowd-shipping* logistics, in which the companies ask the people to collect the packages to a certain location and deliver it to the final user [2][3]. By doing this, companies do not only decrease the costs, but also the environmental impact since people accepting the delivery usually would take advantage of travels that they have to do anyhow for other activities. Another interesting application of SE occurs in an evolution of the Internet of Things (IoT) concept called opportunistic IoT (oIoT) [4]. Since the IoT development is considerably slowed down by the difficulty and costs involved in building telecommunication networks capable of continuously transmitting large amounts of data collected by sensors, through oIoT the citizens share (in exchange for a reward) the internet of their devices (mobile phones, modems) so that the nearby sensors can exploit it to communicate the gathered data. In this work, we do not want to concentrate on a specific application rather on a very general SE-based setting in order to embrace all the basic characteristics of such a business model.

An effective planning of operations under the SE paradigm yields an interesting optimization problem. The decision-maker must decide how much he is willing to pay to a candidate for each task, when and where to rely on employees and on candidates, which tasks to assign to the employees and for which tasks the candidates must be contacted, in order to minimize the total operational costs. It is important to note that the reward paid to a candidate is generally lower on average than the cost that the company bears for an employee. However, while an employee is obliged to accept and carry out the tasks assigned to him, there is no certainty that a candidate will accept a proposed task.

Little attention has been devoted to the development of optimization models aimed at effectively scheduling companies operations that exploit SE. Just few works [5][6][7] have tried to tackle the problem and, therefore, there is a large room for improvement of existing approaches as well as for the design of more innovative and complete ones (as claimed regarding crowd-shipping in [3]). In particular, to the best of our knowledge, there is no published optimization model that explicitly accounts for individual candidate behaviour when planning SE-based operations. As already mentioned, one characteristic that makes challenging the optimization problems deriving from the implementation of the SE paradigm is the fact that candidates are not constrained a priori to respect a contract. This means that, once contacted, the candidate may not accept the task and, if we assume a pure rational profit-maximization behavior of the candidate, the reject can happen because the proposed reward is lower than the candidate expectation. It is therefore important to integrate tools in the decision-making process that allow monitoring the individual behavior of potential candidates.

In this work, to account for individual behaviour, we rely on the candidate's *willingness to accept* (wta) a task, i.e., the minimum reward expected by a candidate to accept a task. The wta is a well consolidated concept in utility theory and has been used since long to explain human subject preferences in economics [8]. From the decision-maker point of view, the candidate's wta is not deterministically known, since it depends on some factors that are intrinsic of the candidates. Therefore, we consider the candidate wta as a random variable.

Thus, the probability of acceptance for a candidate will be equal to the probability that the offered reward is greater or equal to the wta of the candidate. The adopted perspective is similar to [6]. However, instead of relying on a single random variable describing the number of candidates, we model each single candidate behavior through a Bernoulli random variable. The parameter of such a Bernoulli random variable, i.e. the probability that the candidate accepts the task, is not fixed but depends by the proposed reward.

This paper's contribution is twofold. First, we propose a novel mathematical model for SE-based services optimization. The formulation, which includes chance constraints [9], results to be the first one that explicitly accounts for each individual candidates behaviour. Second, since the complexity of the proposed model and the explicit consideration of stochastic parameters do not allow to obtain a simple solution, we derive a mixed-integer quadratic programming model that approximates the original model. This is done by making some reasonable hypothesis on the probability distribution of the wta of each candidate, and by exploiting the Markov inequality. Several computational experiments validate the suitability of our proposed model and solution approach.

The rest of the paper is organized as follows. The optimization problem is defined and modeled in Section II. Our solution approach is described in Section III. Section IV presents the experimental results, while Section V concludes the paper.

II. The social engagement optimization problem

The social engagement optimization problem that we want to study considers a decision-maker (in general a company) whose goal is to use people, in the following called *candidate*, in addition to employees in order to perform a set of tasks. In particular, we consider a urban environment divided in several geographical areas such as mobile phone cells, neighborhoods of different markets or just geographical areas. Each of these areas is characterized by a number of tasks to perform and each tasks is characterized by different workloads, thus a single task may require more candidates to be done. For example, in the crowd shipping setting these tasks are the delivery required by customers out of the store, while in the oIoT application these tasks consist in sharing the internet connection with smart sensors in the city.

Each task can either be performed by using employees or candidates. Employee are more expensive, are available in a small number but they execute the tasks assigned. Instead, candidates are less expensive, their quantity is virtually unlimited (since the number of people considered for SE is far greater than the number of tasks) but they can refuse to perform a task with a given probability. We assume that the acceptance probability increases as the offered reward increase. Please note that, in practice, am employee has greater productivity than a candidate. The goal of the decision-maker is to minimize the total operative costs while enforcing that with high probability all the tasks must be performed.

Let us consider a set $\mathcal{I}$ of tasks and a set $\mathcal{M}$ of candidates. For each task i, let W_i be the workload required, α_i be the required probability for its accomplishment, Δ_i^m be a random variable representing the wta of candidate m, and c_i be the cost of using an employee. Moreover, let B be the number of available employees and $r > 1$ be the ratio between the productivity of an employee and that of a candidate, i.e., the workload that a single employee can afford as compared to a candidate in the same time frame.

We define the decision variables $Q_i^m \in \mathbb{R}^+$ as the reward offered to candidate m to accept task i and $z_i \in \mathbb{N}$ as the number of employee assigned to tasks i. Moreover, we consider the probability for candidate m to accept task i called $x_i^m \in [0,1]$ and the random variables Y_i^m distributed according to a Bernoulli distribution of probability x_i^m which assume value 1 if candidate m accepts to perform task i. Then, the Social Engagement Optimization Problem ($SEOP$) can be formulated as follows:

$$\min \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}} Q_i^m x_i^m + \sum_{i\in\mathcal{I}} c_i z_i \tag{1}$$

$$\text{s.t.} \quad x_i^m = \mathbb{P}[Q_i^m \geq \Delta_i^m], i \in \mathcal{I}, m \in \mathcal{M} \tag{2}$$

$$\mathbb{P}[Y_i^m = a] = (x_i^m)^a (1 - x_i^m)^{(1-a)}, i \in \mathcal{I}, m \in \mathcal{M} \tag{3}$$

$$a \in \{0, 1\} \tag{4}$$

$$\mathbb{P}\left[\sum_{m\in\mathcal{M}} Y_i^m + r z_i \geq W_i\right] \geq \alpha_i, i \in \mathcal{I} \tag{5}$$

$$\sum_{i\in\mathcal{I}} z_i \leq B \tag{6}$$

$$z_i \in \mathbb{N}, i \in \mathcal{I}, \tag{7}$$

$$Q_i^m \in \mathbb{R}^+, x_i^m, Y_i^m \in [0,1], i \in \mathcal{I}, m \in \mathcal{M}. \tag{8}$$

The total cost in (1) is expressed as the summation between the total expected cost offered as rewards (the reward Q_i^m is paid with probability x_i^m), and the sum of the costs paid for employees. Constraints (2) define the variables x_i^m as the acceptance probability, while constraints (3) and (4) ensure Y_i^m to follow a Bernoulli distribution. Constraints (5) are chance constraints enforcing a minimum probability of doing a given task either by using employees or candidates. It is worth noting that ensuring that each task is performed with a given probability is less strict than requiring that all the tasks will be performed with a given probability. Nevertheless enforcing this second condition would lead to too conservative solutions. Finally, constraint (6) accounts for the limited number of employees.

III. Solution Approach

The optimization problem in (1)-(6) is difficult to solve due to the definition of x_i^m in constraints (2), of Y_i^m in constraints (3) and (4), and the chance constraints in (5). Hence, we approximate these constraints in order to get a model which can be readily solved with off-the-shelf solvers.

Constraints (2) involve the cdf of the random variable Δ_i^m. We approximate it by means of a piece-wise linear function with J breakpoints. In particular, instead of constraints (2) we add a set of constraints of the form

$$x_i^m \leq k_j Q_i^m + q_j, \quad j = 1, \ldots, J, i \in \mathcal{I}, m \in \mathcal{M}, \tag{9}$$

where k_j and q_j are obtained by imposing proper conditions (e.g. the passage in J points of the cdf). This choice is equivalent to enforce $x_i^m \leq \min[1, m_1 Q_i^m + q_1, \ldots, m_J Q_i^m + q_J]$, where the first term of the minimum comes from the definition of x_i^m. Since the approximation proposed in (10) just lead to concave functions (being the pointwise minimum of affine functions) and since the a general cdf may be convex in some portion of the domain, the proposed approximation is not guarantee to converge to the cdf for all the distributions. In the following, for the sake of simplicity, we consider just $J = 1$ and we impose the passage for the point $(0, 0)$ meaning that with 0 reward the probability that the candidate will perform the task is 0, and for the point $(\bar{Q}_i^m, 1)$ where $\bar{Q}_i^m$ is a reward for which the candidate m is willing to perform the task i with a probability that we may approximate to be 1. By making this choice, the obtained final approximation of Constraints (2) is:

$$x_i^m \leq Q_i^m / \bar{Q}_i^m, \quad i \in \mathcal{I}, m \in \mathcal{M}. \tag{10}$$

Now let us consider the constraints in (5), note that these constraints can be written as:

$$\mathbb{P}\left[\sum_{m \in \mathcal{M}} Y_i^m \geq W_i - r z_i\right] \geq \alpha_i, \quad i \in \mathcal{I}. \tag{11}$$

By using the Markov inequality, for each $i \in \mathcal{I}$, it holds:

$$\frac{\mathbb{E}[\sum_{m \in \mathcal{M}} Y_i^m]}{W_i - r z_i} \geq \mathbb{P}\left[\sum_{m \in \mathcal{M}} Y_i^m \geq W_i - r z_i\right] \geq \alpha_i. \tag{12}$$

Hence, since $\mathbb{E}\left[\sum_{m \in \mathcal{M}} Y_i^m\right] = \sum_{m \in \mathcal{M}} \mathbb{E}[Y_i^m] = \sum_{m \in \mathcal{M}} x_i^m$,, Eq. (12) leads to the following constraint:

$$\sum_{m \in \mathcal{M}} x_i^m \geq \alpha_i (W_i - r z_i), \quad i \in \mathcal{I}. \tag{13}$$

Eq. (13) is enforcing that the expected workload form the candidates must be greater than the α_i percent of the people needed. Moreover, by considering the bound provided by Eq. (13), we are reducing the feasible set, thus the condition in (11) will be satisfied for greater value of α_i.

Then, the resulting approximation of the SEOP ($SEOP_{ap}$) is the following mixed integer quadratic model:

$$\min \sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}} Q_i^m x_i^m + \sum_{i \in \mathcal{I}} c_i z_i \tag{14}$$

$$\text{s.t. } x_i^m \leq \frac{Q_i^m}{\bar{Q}_i^m}, \quad i \in \mathcal{I}, m \in \mathcal{M} \tag{15}$$

$$\sum_{m \in \mathcal{M}} x_i^m \geq \alpha_i (W_i - r z_i), \quad i \in \mathcal{I} \tag{16}$$

$$\sum_{i \in \mathcal{I}} z_i \leq B \tag{17}$$

$$z_i \in \mathbb{N}, \quad i \in \mathcal{I}, \tag{18}$$

$$Q_i^m \in \mathbb{R}^+, x_i^m \in [0, 1], \quad i \in \mathcal{I}, m \in \mathcal{M}. \tag{19}$$

IV. Experimental results and analysis

We now present CPU experiments validating the proposed solution approach. All the experiments were performed on a *Intel(R) Core(TM)* i7-5500U CPU@2.40GHz computer with 16GB of RAM and running *Ubuntu* v20.04. The exact solver used was Gurobi v9.1.1 via its Python3 APIs. The instances, according to realistic crowd-shipping scenarios, were generated considering $|\mathcal{I}| = \{5, 10, 20, 50, 100\}$, $|\mathcal{M}| = 4|\mathcal{I}|$, $w_i = 7 * (0.8 + 0.4 \cdot \mathcal{U}(0, 1))$, $B = 2$, and $\alpha_i = \bar{\alpha}, \forall i \in \mathcal{I}$ with $\bar{\alpha}$ drawn from $\mathcal{U}(0.6, 0.9)$. Finally, the random variables Δ_i^m were drawn from a Gumbel distribution, while $\bar{Q}_i^m$ was set to be equal to the 99 percentile of Δ_i^m.

A. CPU results

We first study the CPU solving time with respect to the dimension of the $SEOP_{ap}$. In particular, versus the growth of $|\mathcal{I}|$ and $|\mathcal{M}|$, we evaluate the CPU time (sec), the time-to-best (sec) (the number of seconds from the start of the execution of Gurobi to the time in which it founds the best solution of the run), and the MIP gap (%) (computed as the percentage difference between the lower and upper objective bound. In particular, we consider the least gap value that Gurobi has to reach before stopping its execution). The average and standard deviation on 10 instances are shown in Table I. In all the runs we set the solver time limit to 1 hour.

TABLE I
Average $[\mu]$ and standard deviations $[\sigma]$ of the CPU time, time to best, and MIP gap for different values of $\mathcal{I}$ and $\mathcal{M}$.

Instance		CPU time(sec)		time-to-best (sec)		MIP gap (%)	
$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{M}\rvert$	μ	σ	μ	σ	μ	σ
5	20	0.09	0.01	0.09	0.01	0	0
10	40	1569.56	175.83	614.31	235.43	0	0
20	80	2780.84	573.26	2634.94	897.98	0	0
50	200	3600.00	0.00	1054.31	562.25	56	47
100	400	3600.00	0.00	1679.57	720.77	100	0

Instances with $|\mathcal{I}| = 5$, and $|\mathcal{M}| = 20$ are solved almost instantaneously with 0 gap. The time-to-best is equal to the CPU time since the difference are below the hundredths of a second. For instances with $|\mathcal{I}| = 10, |\mathcal{M}| = 40$, and $|\mathcal{I}| = 20, |\mathcal{M}| = 80$, the CPU time increases, but the solver is still able to find the optimal solution inside the time limit. For the instances with $|\mathcal{I}| = 10, |\mathcal{M}| = 40$ the time to best is near one half of the total CPU time but solutions with gap below the 5% are found by the solver already in the first minutes of the run. Instead, for the instance with $|\mathcal{I}| = 20, |\mathcal{M}| = 80$, the time-to-best is close to the whole computation time and no solution with gap below the 5% is found in the first minute of the run. For instances of greater dimensions, the solver is not able to find the optimal solution in the given time limit, for this reason the CPU time is equal to 3600 seconds with a standard deviation of 0. Nevertheless, for instances with $|\mathcal{I}| = 50$, and $|\mathcal{M}| = 200$, several times, the final gaps are are smaller than 10%, while for instances with $|\mathcal{I}| = 100$, and $|\mathcal{M}| = 400$, the solver is not able to find a good bound in the allocated computational time, hence, a 100% MIP gap with 0 standard deviation is reported.

B. Approximation analysis

We now analyze the goodness of the $SEOP_{ap}$ approximation. Since Δ_i^m is distributed as a Gumbel distribution with concave cdf, the approximation proposed converges to the exact function and several techniques for developing good piece-wise approximation are available [10]. Thus, we are interested in quantifying how much conservative is the Markov inequality with respect to Eq. (5). Hence, we compute the optimal solution of $SEOP$ and we use it to calculate $\hat{\alpha} := \mathbb{P}[\sum_{m\in\mathcal{M}} Y_i^m + rz_i \geq W_i]$. This can be done easily by noting that the Y_i^m are independent with respect to the index m since the knowledge about candidate m performing a task does not provide any information related to the execution of the same task by other candidates. Thus, $\sum_m Y_i^m$ is a sum of independent random variable distributed according to Bernoulli distribution of parameter x_i^m. Central Limit Theorems for non identically distributed random variables are available and, in particular, by applying the Lyapunov Central Limit Theorem it is possible to prove [11] that for large values of $|\mathcal{M}|$ (in practice $|\mathcal{M}| \geq 30$), it holds that:

$$\sum_{m\in\mathcal{M}} Y_i^m \sim \mathcal{N}\left(\sum_{m\in\mathcal{M}} x_i^m, \sum_{m\in\mathcal{M}} x_i^m(1-x_i^m)\right), \quad i \in \mathcal{I}. \tag{20}$$

By using (20), we can compute α by solving:

$$\alpha_i = 1 - \Phi\left(\frac{W_i - rz_i - \sum_{m\in\mathcal{M}} x_i^m}{\sqrt{\sum_{m\in\mathcal{M}} x_i^m(1-x_i^m)}}\right), \quad i \in \mathcal{I}, \tag{21}$$

where Φ is the cdf of a standard normal distribution. We report the value of the $\alpha \in [0.5, 1]$ in $SEOP_{ap}$ versus the $\hat{\alpha}$ computed with Eq. (21) in Figure 1. All the results are averaged over 10 runs and the standard deviation of the observation is represented as an uncertain area. We compute the results for $|\mathcal{I}| = 10$ and $|\mathcal{M}| = 40$ since we are able to get the optimal solution in a reasonable amount of time. Moreover, with $|\mathcal{M}| = 40$ there are enough candidates to apply results in (20)–(21).

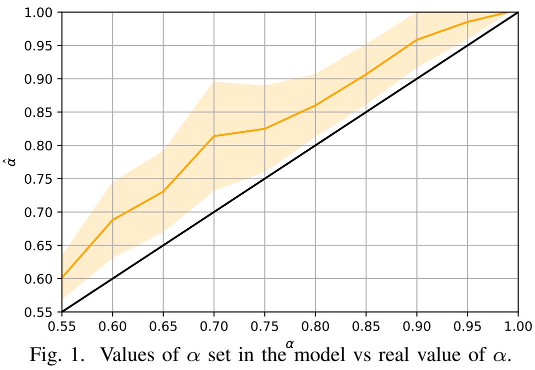

Fig. 1. Values of α set in the model vs real value of α.

As we expected, the curve is above the line $\hat{\alpha} = \alpha$ since by using the Markov inequality we are considering an upper bound on the probability. Nevertheless, the results are close to the exact value being on average 10% higher than the α set in the model. Thus, in the real field, the decision-maker may lower by 10% the values of the αs and get a solution compliant with the wanted probability of execution.

V. Conclusions and Future Works

We proposed a new probabilistic model for SE-based services optimization encompassing the wta of the candidate involved in the business model. We prove, by means of CPU experiment that, despite the difficult formulation, the model can be approximated into a nice tractable form able to provide timely solution for crowd-shipping applications. However, being the SE a very seminal topic within the optimization field, we believe that a full-fledged experimental design to explore all the solution characteristics is needed. Some questions to answer are related to the performance of the method in the case in which non-concave distributions for the wta are considered or how the solutions of the model are related to the number of breakpoints used by the piece-wise wta approximation.

Acknowledgements

The work has been supported by *ULTRAOPTYMAL - Urban Logistics and sustainable TRAnsportation: OPtimization under uncertainTY and MAchine Learning*, a PRIN2020 project funded by the Italian University and Research Ministry (grant n. 20207C8T9M, website: https://ultraoptymal.unibg.it).

References

[1] F. Corno, L. D. Russis, and T. Montanaro, "Estimate user meaningful places through low-energy mobile sensing," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Oct. 2016. doi: 10.1109/smc.2016.7844703

[2] A. Santini, A. Viana, X. Klimentova, and J. P. Pedroso, "The probabilistic travelling salesman problem with crowdsourcing," *Computers & Operations Research*, vol. 142, p. 105722, Jun. 2022. doi: 10.1016/j.cor.2022.105722

[3] A. Alnaggar, F. Gzara, and J. H. Bookbinder, "Crowdsourced delivery: A review of platforms and academic literature," *Omega*, vol. 98, p. 102139, Jan. 2021. doi: 10.1016/j.omega.2019.102139

[4] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things," *Journal of Network and Computer Applications*, vol. 36, no. 6, pp. 1531–1539, Nov. 2013. doi: 10.1016/j.jnca.2012.12.028

[5] E. Fadda, G. Perboli, and R. Tadei, "A progressive hedging method for the optimization of social engagement and opportunistic IoT problems," *European Journal of Operational Research*, vol. 277, no. 2, pp. 643–652, Sep. 2019. doi: 10.1016/j.ejor.2019.02.052

[6] ——, "Customized multi-period stochastic assignment problem for social engagement and opportunistic IoT," *Computers & Operations Research*, vol. 93, pp. 41–50, May 2018. doi: 10.1016/j.cor.2018.01.010

[7] E. Fadda, D. Mana, G. Perboli, and R. Tadei, "Multi period assignment problem for social engagement and opportunistic IoT," in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*. IEEE, Jul. 2017. doi: 10.1109/compsac.2017.173

[8] W. M. Hanemann, "Willingness to pay and willingness to accept: how much can they differ?" *The American Economic Review*, vol. 81, no. 3, pp. 635–647, 1991.

[9] P. Li, H. Arellano-Garcia, and G. Wozny, "Chance constrained programming approach to process optimization under uncertainty," *Computers & chemical engineering*, vol. 32, no. 1-2, pp. 25–45, 2008.

[10] G. A. Godfrey and W. B. Powell, "An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution," *Management Science*, vol. 47, no. 8, pp. 1101–1112, Aug. 2001. doi: 10.1287/mnsc.47.8.1101.10231

[11] A. Cuzzocrea, E. Fadda, and A. Baldo, "Lyapunov central limit theorem: Theoretical properties and applications in big-data-populated smart city settings," in *2021 5th International Conference on Cloud and Big Data Computing (ICCBDC)*. ACM, Aug. 2021. doi: 10.1145/3481646.3481652