Virtual agents, learning how to reach a goal by making appropriate compromises

Dilyana Budakova, Veselka Petrova-Dimitrova, Lyudmil Dakovski
Technical University of Sofia, Branch Plovdiv, Bulgaria, e-mail: dilyana_budakova@yahoo.com
Technical University of Sofia, Branch Plovdiv, Bulgaria, e-mail: vesi_s_petrova@yahoo.com
European Polytechnic University, Pernik, Bulgaria, e-mail: l.dakovski@gmail.com

***Abstract**: This paper proposes a modification to the model-free Reinforcement learning algorithm Q-learning. It is implemented to train smart gift shopping-cart learning agents (SGSCLA). The aim of the modification is to empower the learning agent to reaching a goal by making appropriate compromises only. That is the way in which the measure models and emotional models, represented as new agent memory matrixes are introduced. This models show how the user perceives and evaluates the environment. The Shopping Center is represented by a multigraph in which the nodes represent three groups of shops. The edges illustrate the connections between the shops; the primary (major) and the secondary (minor) paths between them and the emotions, evoked in the customer under consideration by a visit to a particular shop. The user can be see the route suggested by the virtual agent and the made compromises to the goal.The emotion types chosen for the purpose of the experiment are boredom, joy and worry. The environment model allow for exploring and predicting the change in the customer's mood as he/she moves from one shop to another.*

***Keywords:** Intelligent system, Reinforcement learning, Control the way of reaching a goal.*

1. INTRODUCTION

The task of smart gift shopping modeling and learning proves to be appropriate for the aims of the experiment as it allows for studying: people's social behavior; the way of building relationships between them; the processes of modeling emotional and motivational models; the application of the Theory of Mind; people's emotions in relation to others. Therefore the proposed modified algorithm has been applied to training smart gift shopping-cart learning agents (SGSCLA). Concrete implementation of the shopping agents can be running on each gift shopping cart in the shopping centers or on virtual or holographic displays. That's why this task allows for building and exploring connections between computer science, robotics and psychology.

Choosing and exchanging gifts can show what a person thinks of others, what is valuable, or what fun about it is. It can show how people build relationships between themselves [1] [2]. A number of publications [3] [4] discuss some of the therapeutic benefits of shopping. For example, when people shop, they might want to get prepared at mental level to starting something new. They naturally visualize how they use the products they view or purchase. By making small gifts to themselves, people can recharge with energy and can naturally reduce their anxiety and stress, [3] look at ways to tell if a shopping habit is becoming a problem. Therefore an intelligent system for smart gift shopping modeling allows, on the one hand, for studying people's social behavior. On the other hand, it allows for applying and improving the learning algorithms. In [5][6] smart shopping system and smart shopping cart learning agents development are describe.

This paper proposes a modification to the model-free Reinforcement learning algorithm. It is implemented to train SGSCLA. The aim of the modification is to empower the learning agent to reaching a goal by making appropriate compromises. That is the way

the measure models and emotional models, represented as new agent memory matrixes are introduced. This models show how the user perceives and evaluates the environment.

The learning process in the reinforcement learning algorithm is characterized by maximizing, reward signal to reach the goal [7][8]. Getting to the cash-points in the shopping center or getting back home after the shopping has been done can be set as such a goal in the considered example for training SGSCL agents. Thus the reward signal will guarantee the convergence of the algorithm. Additional abstract goals are set to manage the way of reaching the goal. For example: shopping therapy to be implemented by the end of the shopping; to recharge the customer with energy and happy emotions; the customer to rejoice a friend of his/hers by joint visiting his/her favorite shops; the customer to shop safely, following the prescriptions for limiting the infection with coronavirus, etc. Achieving these abstract goals requires construction and use of multiple models of how the user perceives the environment and what the user think about desires and preferences of others. New rules are also being introduced for selecting a model and for selecting criteria for forming a gift shopping path.

The SGSCL agents are trained as follow: measure models and emotional criteria are learnt and represented in the memory of the agent. The following models can be learnt, for example: emotional model according to the user's preferences; an emotional model according to what the customer thinks of his friend's preferences; model of major (busiest) and secondary (quiet detours) to the shops; motivational models according to Maslow's theory of human needs and motivation [9]. Then, all of the built models are used as particular complex criteria in order to make the Q-learning agent find the optimal path for gift shopping.

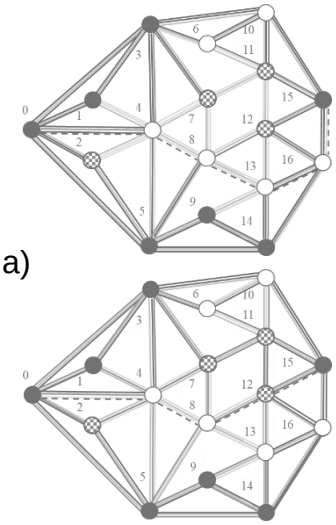
If the goal cannot be reached by following the learnt criteria, the agent can compromise appropriate criterion. The agent makes appropriate tradeoffs only in order to reach the goal. Agent can choose more in number and more acceptable compromises, rather than make fewer but unacceptable ones.

In [8][9][10] the authors considering real-life problems that involve multiple objectives, that giving rise to a conflict of interests. They propose multi-objective reinforcement learning algorithms that provide one or more Pareto optimal balances of the problem's original objectives. Single-policy techniques such as secularization functions can be provide to guide the search toward a particular compromise solution [10][11][12] or it might be appropriate to provide a set of Pareto optimal compromise solutions to the decision maker, each comprising a different balance of objectives. Simultaneously learn a set of compromise solutions is another idea [10][11][12]. In other words, criteria are to be set, characterizing a goal, and solutions proposed that lead to a tradeoff goal, optimally balancing between the pre-set requirements.

2. Q-LEARNING ALGORITHM MODIFICATION REALIZATION. USE OF CRITERIA MODEL REPRESENTED AS K MATRIX AND EMOTIONAL MODELS, PRESENTED AS EMOTIONAL MATRIX E_{USER_FRIEND}

The task considered here is related to smart shopping realization. Besides, the goal is different every time. The shops and stands in the Shopping Center that customers want to reach are different. So not only reaching the goal is important in this task. The way of reaching it, together with the criteria, which a certain path meets, are of the same level of importance.

In order to make the Q-learning agent find the optimal sequence of lobbies or hallways by meeting specific criteria, the use of environment criteria model represented as K matrix and the emotional criteria models represented as E_{user_friend} matrix are introduced.



a)

b)

```
//E user_friend – emotional user-friend matrix
public static string[,] q_emotion = new string[,]{
{ -1, bb, bb, bb, bb, bb, -1, -1,  -1, -1, -1, -1, -1, -1, -1, -1, -1 },
{  bb, -1, -1, bb, bb, -1, -1, -1,  -1, -1, -1, -1, -1, -1, -1, -1, -1 },
{  wj, -1, -1, -1, wj, wj, -1, -1,  -1, -1, -1, -1, -1, -1, -1, -1, -1 },
{  bb, bb, -1, -1, bb, -1, bb, bb, -1, -1, bb, -1, -1, -1, -1, -1, -1 },
{  jj,  jj,  jj,  jj, -1,  jj, -1,  jj,   jj, -1, -1, -1, -1, -1, -1, -1, -1 },
{  bb, -1, bb, -1, bb, -1, -1, -1, bb, bb, bb, -1, -1, -1, bb, -1, -1 },
{ -1, -1, -1,  jj, -1, -1, -1, -1,  -1, -1,  jj,  jj, -1, -1, -1, -1, -1 },
{ -1, -1, -1, wj, wj, -1, -1, -1, wj, -1, -1, wj, -1, -1, -1, -1, -1 },
{ -1, -1, -1, -1, jj, jj, -1, jj,  -1, jj, -1, -1,  jj,  jj, -1, -1, -1 },
{ -1, -1, -1, -1, -1, bb, -1, -1, bb, -1, -1, -1, -1, bb, bb, -1, -1 },
{ -1, -1, -1,  jj, -1, -1, jj, -1,  -1, -1, -1,  jj, -1, -1, -1,  jj, -1 },
{ -1, -1, -1, -1, -1, -1, wj, wj, -1, -1, wj, -1, wj, -1, -1, wj, -1 },
{ -1, -1, -1, -1, -1, -1, -1, -1, wj, -1, -1, wj, -1, wj, -1, wj, wj },
{ -1, -1, -1, -1, -1, -1, -1, -1, -1,  jj,  jj, -1, -1, jj, -1, jj, -1, jj },
{ -1, -1, -1, -1, -1, -1,  bb, -1, -1, -1, bb, -1, -1, -1, bb, -1, -1, bb },
{ -1, -1, -1, -1, -1, -1, -1, -1, -1,  -1, -1, bb, bb, bb, -1, -1, -1, bb },
{ -1, -1, -1, -1, -1, -1, -1, -1, -1,  -1, -1, -1, -1, -1,  jj,  jj,  jj,  jj, -1 }};
```

c)

Fig. 1a and Fig. 1b: Multigraph in which the nodes represent three groups of shops. The edges illustrate the connections between the shops; the primary (major) and the secondary (minor) paths between them and the emotions, evoked in the customer under consideration by a visit to a particular shop. The route suggested by the VLA is given too. Fig. 1a: Three compromises are made: three secondary paths followed. Fig. 1b: The route suggested by the VLA when the unacceptability of the decision to visit a shop marked in large checkerboard is mitigated. Fig. 1c: Emotional user-friend matrix E_{user_friend}.

For the purposes of the experiment the Shopping Center is represented by a graph with 17 nodes and 36 edges between them as shown in "Figure 1a" and "Figure 1b". Every shop in the considered Shopping Center is represented as a node. Every lobby or a hallway, connecting the shops, is represented by an edge. The busiest and most wanted to go through lobbies or hallways are marked in *light gray* "Figure 1a" and "Figure 1b" and have a measure of 1 in the K matrix. The secondary, distant, non-desired pathways are marked in *dark gray* "Figure 1a" and "Figure 1b" and have a measure of 2 in the K matrix.

The shops are divided by random law into three groups. It is also randomly assumed that each group of shops creates a particular type of emotion in the customer under consideration and the same or another type of emotion in his/her friend. The emotion types chosen for the purpose of the experiment are *boredom*, *joy* and *worry*. They are denoted in the E_{user_friend} matrix by the corresponding letters: *b* – for *boredom*; *j* – for *joy*; *w* – for *worry* "Figure 1c". In the environment models presented by the graph, the shops from the three groups are marked in different colors and patterns, namely: *black*; *white*; and *large checkerboard* respectively "Figure 1a" and "Figure 1b".

Minus one (-1) is put in the matrix in a place where there is no connection between the number of a shop, set by a number of a row, and the number of a shop, given by a number of a column.

Now, the VL agents are trained in two stages. In the first stage in addition to developing the criteria model represented as K matrix, emotional criteria models represented as E_{user_friend} matrix are learnt as well "Figure 1c". As the name of the matrix

E_{user_friend} itself suggests, this matrix combines two emotional models. The first one concerns the first of the considered customers, while the second concerns his/her friend. The aim for the VLA is to learn the types of emotions these customers experience when they visit the shops under consideration.

In one and the same cell of the E_{user_friend} matrix, the type of emotion that the current shop triggers in the customer under consideration is first recorded and then, in that cell again, the type of emotion that the current shop triggers in the friend of his/hers is stored.

For example, from the shop number 0 you can go to the shop number 1. The type of emotion, triggered by the shop 0 in the customer under consideration, as well and his/her friend, is *boredom*. Therefore, *bb* is written in the cell in row 0 and column 1 of the E_{user_friend} matrix "Figure 1c".

As an example again, from the shop number 2 you can go to the shop number 4. The emotion, caused by the shop number 2 in the customer under question is *worry*, while for his/her friend it is *joy*. Therefore, *wj* is written in row 2 and column 4 of the matrix E_{user_friend} "Figure 1c".

It can be summarized that both for the consumer and for his/her friend the shops marked in *black* cause *boredom*, and the shops, marked in *white*, cause *joy*. The shops, marked in *large checker board*, cause *worry* in our customer, and *joy* in his/her friend, "Figure 1a" and "Figure 1b".

It is interesting to note here that one can explore and predict the change in the customer's mood as he/she moves from one shop to another. For example, when switching from a *white* shop to a *large checker board* one, it is expected that the joyful mood will turn into worry. In contrast, when passing from a shop, marked in *large checker board*, to a shop, marked in *white*, the worry will disappear and will be replaced by joy and relief.

For the purpose of the experiment, it is assumed that if the customer in question and his/her friend visit the store marked in *large checker board* together, then the customer will not feel worried, but instead joyful. It is accepted that the joy, experienced by his/her friend when visiting a shop marked in *large checker board* will reduce the *worry* of the customer under consideration and the *joy* will be conveyed to him.

The goal is to offer a route from the shop number 0 to shop the number 15, which will bring the customer's friend the greatest *joy*. The two customers shop together. It is also required to follow only major paths. Three compromises are allowed. Visiting a shop in *large checker board* or *black* is considered to be more unacceptable than following secondary paths.

The route suggested by the VLA is given in "Figure 1a". It is seen that it passes through four shops marked in *white*. Two primary and three secondary paths are used. Three compromises are made: three secondary paths followed.

It means that more in number (three) more acceptable compromises have been made and less in number (no one) less acceptable ones. But visiting a shop, marked in *large checker board*, brings joy to the customer's friend. And the purpose of the shopping was to please the friend of the customer's as much as possible. It was supplementary accepted that when the two friends were shopping together, a visit to a *large checker board* -labeled store would not worry the customer in question. This mitigates the unacceptability of the decision to visit a shop marked in *large checker board*. Therefor passes through two shops marked in *white* and through one shop marked in *large checker board* is acceptable too. In this case two primary and two secondary paths will be used and all compromises will be allowed three "Figure 1b".

3. DISCUSSION AND CONCLUSION

The paper describes a modification of the reinforcement learning algorithm. The agent find the optimal path to the goal by following criteria models including emotional represented as new matrixes. Additional abstract goals are set to manage the way of reaching the goal. The models of criteria are learnt and used for goal finding. If the goal cannot be reached by meeting the set criteria, the agent could just ignore a given criterion and find a compromise way. Experiments have been conducted, illustrating the performance of the modified algorithm. The agent can make only appropriate compromises in order to reach the set goal. This modification would be useful when developing complex social scenarios, negotiations. The aim of the modification is to empower the learning agents to: control the way of reaching a goal; better understand the customers; be able to justify their decisions. The modified algorithm has been applied to training smart gift shopping-cart learning agents. This task allows for building and exploring connections between computer science, robotics and psychology. Empathy and Negotiations modeling are other appropriate tasks to be set for this modification of the reinforcement learning algorithm. Negotiation requires understanding an opponent's preferences [13][14][15]. The possibility to control the way of achieving a set goal allows for modeling both intelligent negotiation agents and distinct combination of negotiation tactics. Empathic agents can be seen as agents that can to place themselves into the position of another user's or agent's emotional situation and respond appropriately [15].

4. ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided within the Technical University of Sofia, Research and Development Sector, Project for PhD student helping N202PD0007-19 "Intelligent Cognitive Agent behavior modeling and researching".

5. REFERENCES

[1] Cindy Chan, Cassie Mogilner. 2015. Experiential Gifts Foster Stronger Relationships than Material Gifts. The symposia "The Psychology of Gift Giving and Receiving". The Society of Personality and Social Psychology (SPSP), Annual Convention in Long Beach, California, https://www.spsp.org/news-center/press-releases/psychology-gift-giving-and-receiving

[2] Andong Cheng, Meg Meloy, Evan Polman. 2015. "Picking Gifts for Picky People: Strategies and Outcomes. The symposia "The Psychology of Gift Giving and Receiving". The Society of Personality and Social Psychology (SPSP). Annual Convention in Long Beach, California.

[3] Kit Yarrow, Why Retail therapy works. Five therapeutic benefits of shopping – and how to spot a habit gone awry. Psychology today, Golden Gate University on San Francisco. https://www.psychologytoday.com/us/blog/the-why-behind-the-buy/201305/why-retail-therapy-works

[4] Leonard Lee. 2013. The emotional shopper: Assessing the effectiveness of retail therapy, Foundations and trends in Marketing, Vol. 8, No. 2 (2013), pp. 69-145, ©2015 L. LeeDOI: 10.1561/1700000035.

[5] Budakova D., Dakovski L. (2019) Smart shopping system. *8th International scientific conference (TechSys'19)*. Plovdiv, Bulgaria, 16-18 May 2019. doi:10.1088/issn.1757- 899X; ISSN: 1757-899X; ISSN: 1757-8981.

[6] Budakova D., Dakovski L., Petrova-Dimitrova V. (2019) Smart Shopping Cart Learning Agents Development. *19th IFAC-PapersOnLine, Conference on International Stability, Technology and Culture, (TECIS 2019).* Volume 52, Issue 25, 26-28 September, 64-69, Sozopol, Bulgaria, Elsevier ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.12.447

[7] Sutton R. S. and Barto A. G. (2014) *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, London, England, [Online]. Available: http://incompleteideas.net/book/ebook/the-book.html, [retrieved: 07, 2020].

[8] Gosavi A. (2009) Reinforcement Learning: A Tutorial Survey and Recent Advances. *INFORMS Journal on Computing.* Vol. 21 No.2, pp. 178-192, 2009.

[9] Abraham H. Maslow. 1998. Motivation and Personality. (Paperback), Addison-Wesley Education Publishers, 2nd Edition, Paperback, 400 pages, ISBN: 0060442417 (ISBN13: 9780060442415).

[10] Moffaert K. V. (2016) Multi-Criteria Reinforcement Learning for Sequential Decision Making Problems, *Dissertation for the degree of Doctor of Science: Computer Science, Brussels University Press*, ISBN 978 90 5718 094 1.

[11] Moffaert K. V. and Nowé A. (2014) Multi-objective reinforcement learning using sets of pareto dominating policies. *Journal of Machine Learning Research*, 15:3483–3512.

[12] Natarajan S., Tadepalli P. (2005) Dinamic Preferences in Multi-Criteria Reinforcement Learning. *22nd International Conference on Machine Learning.* Bonn, Germany.

[13] Gehghani M., Gratch J., Carnevale P. J. (2012) Interpersonal Effects of Emotions in Morally-charged Negotiations. *Proceedings of the Annual Meeting of the Cognitive Science Society,* Volume 34, 1476-1481.

[14] Roediger J. E., Lucas S.G., Gratch, J. 2019. Assessing Common Errors Students Make When Negotiating. 19th ACM International Conference on Intelligent Virtual Agents (IVA'19). ACM, Paris, France, 30-37, DOI: http://doi.org/10.1145/3308532.3329470.

[15] Paiva A., Leite I., Boukricha H., and Wachsmuth I., (2017). Empathy in Virtual Agents and Robots: A Survey. *ACM Trans. Interact. Intell. Syst.* 7, 3, Article 11 (September 2017), 40 pages. https://doi.org/10.1145/2912150.