Ranking Preferences Deduction Based on Semantic Similarity for the Stable Marriage Problem

Michaël Guedj UTEC INT, School of IT and New Technologies UTEC – CCI Seine-et-Marne Emerainville, 77436 Marne-la-Vallée Cédex 2 (France) michael.guedj@seineetmarne.cci.fr

Abstract—The stable marriage problem is a well-known problem with many practical applications. Most algorithms to find stable marriages assume that the participants explicitly express a preference ordering. This can be problematic when the number of options is large or has a combinatorial structure. We show, by simply asking the actors (men and women) to fulfill a personal profile with items positioning in a tree-structured semantic network, that it is possible to solve the problem of stable marriages without asking the actors to explicitly operate a ranking over the members of the opposite sex.

I. Introduction

The stable marriage problem is a well-known problem with many practical applications. The stable marriage problem and its many variants have been widely studied in the literature [9] [12] [22], in particular because of important practical applications, such as the National Resident Matching Program [21] and similar large-scale matching schemes. It is usually defined as the problem of matching men to women so that no man and woman, who are not married to each other, both prefer each other to their current partner [9]. Problems of this kind arise in many real-life situations, such as assigning residents to hospitals, students to schools, as well as in twosided market trading. A specific application is a web-based stable marriage system for matching sailors to ships in the US Navy. Surprisingly, a stable matching always exists whatever preferences are held by the men and women. The Gale-Shapley algorithm finds a stable matching in polynomial time [7]. The algorithm assumes that agents express their preferences (over the members of the other gender) explicitly as a totally ordered list of members of the other gender. In some applications, the number of men and women can be large. It may therefore be unreasonable to assume that each man and woman provides a strict ordering over the other gender. Also it is not always easy to choose: indeed some agents may not be able to choose among several partners. Our goal in this paper is to give a pretreatment methodology to derive the choices of the agents (men and women) without that these ones have actually given their choices.

The remainder of this paper proceeds as follows: first we present the related work; we then provide the foundation for understanding the problem of stable marriage and the semantic distance over a taxonomy; finally, we give our proposal.

II. RELATED WORK

[13] presents a comprehensive study of variants of the Stable Marriage Problem (SMP) in which the preference lists of the participants are not necessarily complete and not necessarily totally ordered (the preference lists involve ties); such variants may be critical for actual uses. [10] study variants of the classical SMP in which the preferences of the men or the women, or both, are derived from a master preference list. [17] consider using a compact preference formalism so called CP-nets [5] in stable marriage problems; in this way each man and each woman specifies their preferences over the other sex via such CP-net. They call the new underlying problematic: the compact stable marriage problems. Our research is strongly linked with this issue. [18] considered the use of fuzzy constraints in the context of SMP, to model compactly the preference orderings of men and women. We intend here to use general taxonomies; the profiles of the actors being concepts in these taxonomies. Contrary to [17] we do not ask the actors any statement of preferences (the CP-nets being a graphical model for compactly representing conditional and qualitative preference relations); and contrary to [18] we do not ask the actors any constraints statement. Indeed, we simply ask the actors to provide their profiles seen as a list of concepts, each concept taking part into a semantic network (in a tree structure) called taxonomy and encoding a hierarchization of the skills following the specialization/generalization. We deduce then, for each actor, a ranking based on a semantic similarity measure over the taxonomy. Such an approach is really interesting for real case practical uses; the filling of profiles being very natural unlike [17] and [18]. Moreover, we do not change the primitives of the solving algorithm; the proposed process being a pretreatment to obtain a suitable instance that can be processed by the resolution algorithm. Also, our proposal fits naturally stable marriage variants such as with ties. The pre-treatment is based on the job seeking problem or more exactly over the semantic approach that have been proposed to resolve this problem; problem of ranking adequate profiles. Let us remark that the output of the job seeking problem provide an input for the SMP (in the cases of course where the two kinds of actors are in equal number). Actually, several techniques and approaches have been proposed to construct automatic online

recruitment systems. The traditional keyword based techniques mainly depend on exact matching between keywords extracted from job posts and candidate resumes; they suffer from low precision because they usually ignore the underlying semantic aspects of the terms that are extracted from both job posts and resumes [2]. To effectively locate and match individuals and positions, within or from outside an organization, it is important to use semantic technology [6] [15]. The authors of [4], [15], [23] [26] propose automatic recruitment systems that employ semantic resources that have been built based on integrated classifications and standards. [20] uses extra probabilistic edges to extend the applications with skills that the applicant possibly possesses. Particularly, [4] and [15] uses [27]'s similarity measure to evaluate the degree of match between job offers and applicants. [8] uses a similar approach but employs different similarity measures related to different scenarios. The semantic approaches have shown interesting results in accomplishing the matching process. However, development of complete and reliable ontologies that capture up-to-date knowledge about specific domains remains tricky [16] [14]. Our proposal will be based on [4] which gives an actual way to compare two profiles, whose items are concepts of a taxonomy.

III. STABLE MARRIAGE

The stable marriage problem is the following: given n men and n women, each man with a list ranking the women in order of preference and similarly each woman with a preference list of the men, the objective is to couple the men and the women in such a way that each one has no interest in having another husband/wife.

We expose here the mathematical basis we use. We note an ordered list between the symbols "<" and ">"; e.g. < 1,2,3 > and < 2,3,1 > are two different ordered lists. For a finite set X, we note |X| the cardinal of X (i.e. the number of its elements) and \hat{X} the set of all permutations of X; if $X = \{x_1, x_2, x_3\}, \ \hat{X} = \{< x_1, x_2, x_3 >; < x_1, x_3, x_2 >; < x_2, x_1, x_3 >; < x_2, x_3, x_1 >; < x_3, x_2, x_1 >; < x_3, x_1, x_2 > \}$. For a permutation x of k elements, we note x(i) the i-th element of x.

We give below the definitions necessary for the formulation of the well known algorithm of Gale and Shapley [7] that solves the problem of stable marriages. We mainly use the exposure from [11]; the algorithm itself is formulated as in [17].

Definition 1 (Instance of the SMP). An instance of the SMP is a tuple (M, W, n, π) where:

- 1) M is the set the of men;
- 2) W is the set the of women;
- 3) $|M| = |W| = n \in \mathbb{N}$ is the size of the sets M and N;
- 4) π is a (preference) function defined as: $\pi: M \cup W \to \widehat{M} \cup \widehat{W}$; $\forall m \in M, \pi(m) \in \widehat{W}$ ($\pi(m)$ is the preference list of m) and $\forall w \in W, \pi(w) \in \widehat{M}$ ($\pi(w)$ is the preference list of w).

Definition 2 (Matching). A matching for an instance of the SMP is a set $\mathfrak{M} \subset M \times W$ such that each $m \in M$ and each $w \in W$ appears in exactly one ordered pair in the set \mathfrak{M} .

Hence, in a matching, each man is paired with exactly one woman, and each woman with exactly one man. In fact, $\mathfrak{M}:M\to W$ defines a bijection; and noting $\mathfrak{M}(m)=w\iff (m,w)\in\mathfrak{M};\ \mathfrak{M}(m)$ is then the partner of $m\in M$ in the matching, $\mathfrak{M}^{-1}(w)$ being that of $w\in W$.

Definition 3 (Preference Relation). (Preference Relation of a Man) For $m \in M$ and $w, w' \in W$, we note $w' <_m w$, and one says that m prefers w to w', if:

- 1) $w = \pi(m)[i];$
- 2) $w' = \pi(m)[j];$
- 3) i < j.

We define the preference relation of a woman symmetrically.

Definition 4 (Blocking Pair). A pair $(m, w) \in M \times W$ is a blocking pair for the matching \mathfrak{M} if:

- 1) $(m, w) \notin \mathfrak{M}$ (m is not in a relationship with w in \mathfrak{M});
- 2) m and w both prefers each other to their own partners in the matching \mathfrak{M} :

$$\mathfrak{M}(m) <_m w \text{ and } \mathfrak{M}^{-1}(w) <_w m.$$

Definition 5 (Stable Matching). A matching \mathfrak{M} is stable or a solution, if there is no blocking pair for \mathfrak{M} .

The motivation for such a definition is that given a blocking pair (m, w) in some a matching, m and w would leave their partners for each other.

Definition 6 (Gender Optimal and Pessimal Stable Matching). A stable matching is man-optimal (resp. man-pessimal) if each $m \in M$ is paired (in the matching) with the $w \in W$ he prefers most (resp. least) of all possible partners. We define woman-optimal and woman-pessimal stable matching similarly.

The algorithm of Gale and Shapley (Algorithm 1) consists of a number of rounds in which each unengaged man proposes to the most preferred woman to whom he has not yet proposed. Each woman receiving a proposal becomes engaged, provisionally accepting the proposal from her most preferred man. In subsequent rounds, an already engaged woman can trade up, becoming engaged to a more preferred man and rejecting a previous proposal, or if she prefers him, she can stick with her current partner. The algorithm takes $O(n^2)$ steps and construct a matching that is male-optimal, since every man is paired with his highest ranked feasible partner, and female-pessimal, since each woman is paired with her lowest ranked feasible partner; these results being justified by the Theorem 1 and the corollaries 1 and 2; we refer the reader to [11] for the proofs.

Theorem 1. The Gale-Shapley algorithm computes a manoptimal and woman-pessimal stable matching in $O(n^2)$ -time.

Corollary 1. For any instance of the stable marriage problem, there exists a stable matching.

Algorithm 1 Gale-Shapley Algorithm

```
1: procedure GALE-SHAPLEY(The sets M and W of resp. men and women)
2: while there is a bachelor man m do
3: w \leftarrow the first woman in \pi(m) to which m has not yet proposed 4: if w is single then
5: match m with w
6: else if m' <_w m where m' is w's current partner then
7: match m with w and set m' bachelor
8: else
9: w rejects m and m remains bachelor
10: end if
11: end while
12: end procedure
```

Corollary 2. A stable matching is man-optimal if and only if it is woman-pessimal.

As example, consider n=3 and let $W=\{w_1,w_2,w_3\}$ and $M=\{m_1,m_2,m_3\}$ be respectively the set of women and men. The following sequences of strict total orders complete the input (M,W,n) as an instance of SMP:

- $m_1: w_1 > w_2 > w_3$ (m_1 prefers w_1 to w_2 and w_2 to w_3); $m_2: w_2 > w_1 > w_3$; $m_3: w_3 > w_1 > w_2$;
- $w_1: m_1 > m_2 > m_3$; $w_2: m_3 > m_1 > m_2$; $w_3: m_2 > m_1 > m_3$.

For this instance, the Gale-Shapley algorithm returns the stable marriage $\{(m_1, w_1), (m_2, w_2), (m_3, w_3)\}$.

IV. SEMANTIC MATCHING

The semantic matching allows a comparison of profiles through the use of taxonomies. [27] proposes an approach for semantic search by matching conceptual graphs; their work is then adapted to the context of the job seeking problem by [4] then [8]. Different from traditional recommendation systems which recommend items to users, job recommender systems recommend one type of users (e.g., job applicants) to another type of users (e.g., recruiters). In particular, job recommender system is designed to retrieve a list of job positions to a job applicant based on his/her preferences or to generate a list of job candidates to a recruiter based on its requirements. The job seeking problem is very interesting for the SMP because its objective is to provide a list of preferences to the actors (men and women); this preferences list being considered (in addition to the actors) as the input hypothesis of the SMP. Semantic matching is a technique which combines annotations using controlled vocabularies with background knowledge about a certain application domain. In [4], the domain specific knowledge is represented in the form of various concept hierarchies and can be used to determine the semantic similarity between concepts [19]. Thus a comparison between job descriptions and applicants profiles is possible based on their semantic similarity and not merely relying on the containment of keywords.

This approach sustained especially by [4] and [8] is based on ideas from [27] and [3].

A. Hierarchy of Skills

The background knowledge of the recruitment domain is represented in a machine understandable format that allows

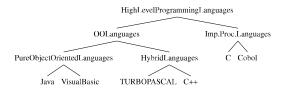


Fig. 1. Skills Hierarchy (taken from [15])

to compare job descriptions and applicant profiles based on their semantic similarity [19] instead of merely relying on the containment of keywords. Such a format is usually [4] [15] [20] a hierarchy of skills so-called taxonomy, which captures the inheritance relationship. An example of such a taxonomy is given in Figure 1 and is taken from [15].

B. The Concept Similarity

The similarity between two concepts (nodes in the taxonomy) c_1 and c_2 is determined by the distance $d_c(c_1,c_2)$ between them, which reflects their respective positions in the concept hierarchy. Concept similarity is defined in [27] (then used by [4] and [8]) as:

$$sim_c(c_1, c_2) = 1 - d_c(c_1, c_2)$$

Every node in a concept hierarchy is assigned a milestone value [27], which is calculated with the formula:

$$milestone(x) = \frac{1}{2k^{level(x)}}.$$

where:

- k is a factor larger than 1 indicating the rate at which the value decreases along the hierarchy;
- and level(n) is the depth of the node n in hierarchy.

Since the distance between two given concepts in a hierarchy represents the path from one concept to the other one over the closest common parent *ccp*, the distance is calculated as shown below [27]:

- $d_c(c_1, c_2) = d_c(c_1, ccp) + d_c(c_2, ccp)$
- $d_c(c, cpp) = milestone(ccp) milestone(c)$.

This model implies two assumptions:

- The semantic differences between upper level concepts are bigger than those between lower level concepts (in other words: two general concepts are less similar than two specialized ones);
- 2) The distance between brothers is greater than the distance between parent and child.

C. The Competence Level Similarity

[4] provides means for specifying required competence levels (cl) in job postings. Hence, the method provided by [4] not only considers taxonomic similarity of concepts but also compares competence levels in order to find the best match. The competence level similarity is determined by the following formula [41:

$$sim_p(cl_1, cl_2) = \begin{cases} 1 - \alpha(cl_1 - cl_2) & \text{if } cl_1 > cl_2 \\ 1 & \text{otherwise} \end{cases}$$

where $0,25 \leq \alpha$ is a factor indicating the rate at which the value of sim_p decreases with increasing deviation between competence levels.

D. Global Similarity

The approach sustained by [4] gives employers also the opportunity to specify the importance of different job requirements. The concept similarity is then justified by the indicated weight, i.e. the similarity between more important skills will have greater influence on the similarity between a job position posting and an applicants profile. Putting all together, the formula for calculating the similarity of a job position posting (c) and a job position seeker (s) is [4]:

$$Sim(c, s) = \sum_{i \in I} w(c_i) \cdot \max_{j} \left[sim_c(c_i, s_j) \cdot sim_p(p(c_i), p(s_j)) \right]$$

where $\sum_{i \in I} w(c_i) = 1$ and $p(c_i)$ (resp. $p(s_j)$) is a competence level in c_i (resp. s_j).

Each required skill from the job position posting (c_i) is compared with each skill in the profile of some a job seeker (s_j) . This includes the calculation of both concept and competence level similarities. The similarity values of the best matching pairs are multiplied by the corresponding weight and summed up yielding the final similarity.

V. STABLE MARRIAGE USING THE UNDERLYING SEMANTIC STRUCTURE

Our goal in this paper is to give a preprocessing methodology to derive the choices of the agents (men and women) without that these ones have actually given explicitly their choices. Indeed, a major barrier to a practical use of the stable marriage resolution is the request for the actors to rank members of opposite gender; whence variants such in which the preference lists of the participants are not necessarily complete and not necessarily totally ordered (with ties) [13]. To do so, we consider another problem: the job seeking problem, whose the (dual) objective is to give to the job seekers a ranked list of companies (sorted relevantly to their profile) and similarly to give to the applicant companies a ranking of the job seekers that may interest them. We remark that the output of the job seeking problem (where e.g. the job seekers are the men and the companies applicant are the women) provides an input to the stable marriage problem (in the case of course where the two kinds of actors are in equal number). To resolve the job seeking problem, approaches based on distance over semantic networks has shown interesting results [4]; such an approache is explained in Section IV.

Our proposal consist to operate a preprocessing over the actors before running the Gale-Shapley Algorithm. To do this we have to consider a ranking measure, based on the semantic similarly measure defined in [4] and presented in Section IV. This measure is dependent on the considered agent, w.l.o.g. $m \in M$, that we noted Sim_m and defined as follow:

$$Sim_m: W \to [0,1]$$

$$Sim_m(w) = Sim(m, w)$$

Such a measure is defined for each men m (resp. each women w) and allows to provide for each m (resp. w) a ranking over the women (resp. over the men). (The ranking measure for the women is a function which takes as input, the set M of the men: $M \to [0,1]$). We summarize this approach in Algorithm 2. Let us note that we can be in a case of equal

Algorithm 2 Gale-Shapley Algorithm with Preprocessing to Deduce the Rankings Based on the Semantic Similarity

```
1: procedure GALE-SHAPLEY_WITH_PREPROCESSED_RANKING(The sets M and W of resp. men and women)
2: // Finding the preferences
3: for m \in M do
4: \pi(m) \leftarrow rank the women in W following Sim_m
5: end for
6: for w \in W do
7: \pi(w) \leftarrow rank the men in M following Sim_w
8: end for
9: Gale-Shapley(M, W)
10: end procedure
```

preference after the achieving of the finding of the preferences stage. However the Gale-Shapley algorithm requests for a strict order of preferences. Several choices can then be considered. If these cases are seen as exceptional, we can take as heuristic to arbitrarily order equal elements in the ranking. Otherwise, we have to consider resolution approach to tackle stable matching with ties [13] [24].

VI. CONCLUSION AND FUTURE WORK

The request done to the actors to rank the members of the opposite sex can be a hindrance for a practical real-world use of the problem of the stable marriage (or some variants/related problems). We have shown that an easy way to solve this issue is to deduce the rankings on the basis of the description of the profiles viewed as a list of concepts, positioning in a tree-structured semantic network. Future work include evaluation of the proposed approach with experimental results; studies cases including also variants/related problems to the stable marriage problem; as well as the extension of the semantics measure to take into account more complex semantic networks.

ACKNOWLEDGMENT

Our work would not have been possible without the support of the CCI Seine-et-Marne, Mr Fehd Bensaid Director of the UTEC and Mr Frédéric Bourcier Director of the IT and new technology department of the UTEC.

REFERENCES

- S. T. Al-Otaibi and M. Ykhlef. A survey of job recommender systems. In International Journal of the Physical Sciences, vol. 7(29), pages 5127– 5142, July 2012.
- [2] Nicholas J. Belkin and W. Bruce Croft. Information filtering and information retrieval: Two sides of the same coin? *Commun. ACM*, 35(12):29–38, December 1992.
- [3] A. Billig and K. Sandkuhl. Match-making based on semantic nets: The xml-based approach of baseweb. In Proceedings of the workshop on XML-Technologien fr das Semantic Web, pages 39–51, 2002.

- [4] Christian Bizer, Ralf Heese, Malgorzata Mochol, Radoslaw Oldakowski, Robert Tolksdorf, and Rainer Eckstein. The impact of semantic web technologies on job recruitment processes. In Wirtschaftsinformatik 2005: eEconomy, eGovernment, eSociety, 7. Internationale Tagung Wirtschaftsinformatik 2005, Bamberg, 23.2.2005 - 25.2.2005, pages 1367–1382, 2005.
- [5] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. J. Artif. Int. Res., 21(1):135–191, February 2004.
- [6] Simona Colucci, Tommaso Di Noia, Eugenio Di Sciascio, Francesco M. Donini, Marina Mongiello, and Marco Mottola. A formal approach to ontology-based semantic match of skills descriptions. *J. UCS*, 9(12):1437–1454, 2003.
- [7] L. S. Shapley D. Gale. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [8] Maryam Fazel-Zarandi and Mark S Fox. Semantic matchmaking for job recruitment: an ontology-based hybrid approach. Proceedings of the 8th International Semantic Web Conference, 2009.
- [9] Dan Gusfield and Robert W. Irving. *The Stable Marriage Problem:* Structure and Algorithms. MIT Press, Cambridge, MA, USA, 1989.
- [10] Robert W. Irving, David F. Manlove, and Sandy Scott. The stable marriage problem with master preference lists. *Discrete Applied Mathematics*, 156(15):2959 – 2977, 2008.
- [11] Michael St. Jules. Stable matchings, 2013.
- [12] D.E. Knuth. Stable Marriage and Its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms. CRM proceedings & lecture notes. American Mathematical Soc.
- [13] David F Manlove, Robert W Irving, Kazuo Iwama, Shuichi Miyazaki, and Yasufumi Morita. Hard variants of stable marriage. *Theoretical Computer Science*, 276:261 279, 2002.
- [14] M. Maree and M. Belkhatir. A coupled statistical/semantic framework for merging heterogeneous domain-specific ontologies. In 2010 22nd IEEE International Conference on Tools with Artificial Intelligence, volume 2, pages 159–166, Oct 2010.
- [15] Malgorzata Mochol, Holger Wache, and Lyndon J. B. Nixon. Improving the accuracy of job search with semantic techniques. In *Business Information Systems*, 10th International Conference, BIS 2007, Poznan, Poland, April 25-27, 2007, Proceedings, pages 301–313, 2007.
- [16] Paslaru E. Mochol, M. and B. Simperl. Practical guidelines for building semantic erecruitment applications. In *International Conference on Knowledge Management*, IKNOW06, Special Track: Advanced Semantic Technologies (AST'06), pages 1–8, Graz, Austria, 2006.
- [17] Enrico Pilotto, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Algorithmic Decision Theory: First International Conference, ADT 2009, Venice, Italy, October 20-23, 2009. Proceedings, chapter Compact Preference Representation in Stable Marriage Problems, pages 390–401. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [18] Maria Silvia Pini, Francesca Rossi, and Kristen Brent Venable. Stable matching problems with soft constraints. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 1511–1512, Richland, SC, 2014. International Foundation for Autonomous Agents and Multiagent Systems.
- [19] Jonathan Poole and J. A. Campbell. A novel algorithm for matching conceptual and related graphs. In *Proceedings of the Third International Conference on Conceptual Structures: Applications, Implementation and Theory*, ICCS 95, pages 293–307, London, UK, UK, 1995. Springer-Verlag.
- [20] Gábor Rácz, Attila Sali, and Klaus-Dieter Schewe. Foundations of Information and Knowledge Systems: 9th International Symposium, FoIKS 2016, Linz, Austria, March 7-11, 2016. Proceedings, chapter Semantic Matching Strategies for Job Recruitment: A Comparison of New and Known Approaches, pages 149–168. Springer International Publishing, Cham. 2016.
- [21] Alvin E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92:991–1016, 1984.
- [22] Uriel G. Rothblum. Two-sided matching: A study in game-theoretic modeling and analysis: By alvin e. roth and marilda a. oliveira sotomayor, econometric society monographs, cambridge univ. press, cambridge, ma, 1990. 265 + xiii pp., \$54.50 (hardback). *Games and Economic Behavior*, 4(1):161–165, 1992.
- [23] Francisco García Sánchez, Rodrigo Martínez-Béjar, Leonardo Contreras, Jesualdo Tomás Fernández-Breis, and Dagoberto Castellanos Nieves. An

- ontology-based intelligent system for recruitment. *Expert Syst. Appl.*, 31(2):248–263, 2006.
- [24] S. Scott and University of Glasgow. A Study of Stable Marriage Problems with Ties. University of Glasgow, 2005.
- [25] Z. Siting, H. Wenxing, Z. Ning, and Y. Fan. Job recommender systems: A survey. In Computer Science Education (ICCSE), 2012 7th International Conference on, pages 920–924, July 2012.
- [26] F. Trichet, M. Bourse, M. Leclere, and E. Morin. Human resource management and semantic web technologies. In *Information and Communication Technologies: From Theory to Applications*, 2004. Proceedings. 2004 International Conference on, pages 641–642, April 2004.
- [27] Jiwei Zhong, Haiping Zhu, Jianming Li, and Yong Yu. Conceptual graph matching for semantic search. In *Proceedings of the 10th International Conference on Conceptual Structures: Integration and Interfaces*, ICCS '02, pages 92–196, London, UK, UK, 2002. Springer-Verlag.