

Profile Matching in Electronic Social Networks Using a Matching Measure for Fuzzy Numerical Attributes and Fields of Interests

Andreas de Vries

South Westphalia University of Applied Sciences, Hagen, Germany Email: <u>de-vries.andreas@fh-swf.de</u>

Received 2 July 2014; revised 5 August 2014; accepted 13 August 2014

Copyright © 2014 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

Abstract

The problem of profile matching in electronic social networks asks to find those offering profiles of actors in the network fitting best to a given search profile. In this article this problem is mathematically formulated as an optimization problem. For this purpose the underlying search space and the objective function are defined precisely. In particular, data structures of search and offering profiles are proposed, as well as a function measuring the matching of the attributes of a search profile with the corresponding attributes of an offering profile. This objective function, given in Equation (29), is composed of the partial matching degrees for numerical attributes, discrete non-numerical attributes, and fields of interests, respectively. For the matching degree of numerical profile attributes a fuzzy value approach is presented, see Equation (22), whereas for the matching degree of fields of interest a new measure function is introduced in Equation (26). The resulting algorithm is illustrated by a concrete example. It not only is applicable to electronic social networks but also could be adapted for resource discovery in grid computation or in matchmaking energy demand and supply in electrical power systems and smart grids, especially to efficiently integrate renewable energy resources.

Keywords

Profile Matching Algorithm, Matchmaking Algorithm, Matching Degree, Electronic Social Network, Matching Fields of Interest, Grid Computing, Renewable Energy, Energy Transition

How to cite this paper: de Vries, A. (2014) Profile Matching in Electronic Social Networks Using a Matching Measure for Fuzzy Numerical Attributes and Fields of Interests. *Applied Mathematics*, **5**, 2619-2629. http://dx.doi.org/10.4236/am.2014.516250

1. Introduction

Social activities in electronic networks play an increasingly important role in our every-day lives. We are exchanging important information via electronic mails, wikis, web-based forums, or blogs, and meet new friends or business contacts in Internet communities and via social network services. Parallel to this growing socialization of the World Wide Web, the requirements on the electronic services become more ambitious. Huge data quantities have to be processed, user-friendly interfaces are to be designed, and more and more sophisticated computations must be implemented to offer complex solutions.

One of the most important subjects in complex networks is the search for specific items or objects in it. For instance this comprises the search for web sites which are relevant with respect to a specific term, or the request for specific resources as consumers do for electricity in smart grids. Therefore there is, and has always been, a great interest in enabling and maintaining efficient special network services which perform precisely these tasks.

This paper studies a special aspect of general electronic networks, but in particular of social network services, the profile matching problem. In essence it asks, given a search profile, for offering profiles matching it best. This problem is in principle well-known in Grid computing, where computational tasks are seeking for appropriate resources such as CPU time and memory space on different computers. In electronic social networks, however, the problem is more general because not only specified attribute ranges such as resource sizes are to be compared but more or less vaguely describable interests. Questions of this kind have recently been under consideration, for instance to implement multiple interest matching in personel business networks [1], to study automated user interest matching in online classified offering systems [2], or to apply semantic fuzzy search techniques to web service matchmaking [3].

Here, however, a different approach is presented, leaving the important but difficult problem of semantic search aside and enabling an automated matching of general types of attributes. The aim of this paper is to formulate a mathematical model for the problem of matching attribute ranges and fields of interests in electronic social networks. It tackles the following fundamental questions. How can an appropriate system and its data structures be designed? How is the mathematical formulation of a matching problem as an optimization problem? In particular, what is its search space? What is its objective function? It is obvious to use a fuzzy function to calculate the matching degree of two numerical ranges or to use the characteristic function to check for the equality of discrete attribute sets, but how could a function calculating a matching degree of two fields of interest look like? One of the central results of this paper is the proposal of a precise definition of such a function computing the matching model is not confined to electronic social networks but can also operate to enable and control electricity resource allocation in electrical power systems or in smart grids. By this approach technological challenges of energy transition programs such as the *Energiewende* to integrate renewable energy into electrical power systems can be tackled.

The paper is organized as follows. After a definition of electronic social networks is given in the next section, a mathematical model of the matching problem as an optimization problem is proposed, especially the data structure of search and offering profiles, the search space, and the matching degree as the objective function. A short discussion concludes the paper.

2. Electronic Social Networks

A *social network* consists of a finite set of actors and the direct relations defined between them. An actor here may be an individual, a group, or an organization, and the direct relation between two actors may indicate that they directly interact with each other, have immediate contact, or are connected through social familiarities such as casual acquaintance, friendship, or familial bonds [4]-[6], see also ([7], ğ3.6). Thus a social network is naturally represented by a graph in which each node represents an actor and each edge a direct relation. Empirically, the mean number of direct relationships of an individual in a biological social network depends on the size of the neocortex of its individuals; the maximum size of such relationships in human social networks tends to be around 150 people ("Dunbar's number") and the average size around 125 people [8].

Since the popularization of the World Wide Web in the middle of the 1990's and in particular around 2005 after the introduction of the *Web* 2.0 paradigm [9], there have rapidly emerged several Internet based social networking services, maintaining "circle of friends" networks, business platforms, knowledge networks, or virtual world online games ([5], g13.5). Examples of each of these types of networks services are as follows:

Orkut (www.orkut.com), Facebook (facebook.com) or Google+ (plus.google.com); LinkedIn (linkedin.com) or XING (xing.com); Wikipedia (wikipedia.org); World of Warcraft, Star Wars: The Old Republic (swtor.com), or Second Life (secondlife.com).

In this paper, an *electronic social network* is defined as a network of at least three human individuals or organizations which use essentially, albeit not exclusively, electronic devices and media to get in contact and acquaintance to each other, to meet new partners, to communicate, and to exchange information. Examples of electronic social networks are Internet social networks, as well as videoconference sessions and conference calls, especially if they serve to meet new people as in party lines, or as long as they admit spontaneous communication between each network actor [10]-[12].

3. The Matching Problem

In computer science, the term *matching* or sometimes *matchmaking* in general refers to the process of evaluating the degree of similarity or agreement of two objects. Each object is characterized by a set of properties or attributes, which in many systems are given by name-value pairs [13]. Matching plays a vital role in many areas of computer science and communication systems. For instance, it is studied for resource discovery and resource allocation in grid computing where matching services are needed to intermediate between resource requesters and resource providers [14]. Other examples are given by the problem of matching demands and supply of business or personal profiles in online auctions, e-commerce, recruitment agencies, or dating services.

3.1. Basic Definitions

In most matching problems, the objects under consideration take asymmetric roles, viz., some *search* for information or request for a service, others *offer* information or provide a service. A single object may naturally do both activities at a time, in electronic social networks this even is the usual case. In the sequel we will therefore more accurately consider the matching of a *search profile*, containing information for a request, and an *offering profile* presenting information for a supply or a provided service.

Given a specific search profile, the *matching problem* then is to find those offering profiles which match it best, in a sense to be specified in the sequel. Generalizations of this problem ask for best global matchings, given a whole set of search profiles and a set of offering profiles. For instance, the *global pairwise matching problem* seeks pairs of search/offering-profiles such that the entity of the pairs matches the best under the constraint that any profile is member of at most one pair. The *global multiple matching problem* searches for possibly multiple combinations of search and offering profiles which as a whole match the best. The pairwise version of the problem typically occurs for dating services or classical marriage matching tasks, whereas the multiple version appears in grid computing or in brokering interest groups.

In this paper we will focus on the local version of the matching problem, *i.e.*, finding an optimum offering profile to a specified search profile. Thus the matching problem is an optimization problem, and to formulate it precisely we have to specify the search space and the objective function. The search space will turn out to be the set of pairs of the fixed search profile and the offering profiles, and the objective function will be a function measuring the "matching degree". We will work out these notions in the next sections.

3.2. Profile Data Structures

A *profile* consists of its owner corresponding to an actor of the electronic social network, a list of *attributes* of a given set *A* together with their values, a list of *attribute stencils* where each stencil represents a pair of an attribute name and its value range, and a list of *fields of interest* specifying their respective levels of interest. Attributes are properties of the profile owner such as age, height, weight, eye color, or hair color, and we therefore subsume them under the class "Owner" (Figure 1).

In principle, there are two different types of attributes, subsumed in the two disjoint sets N and D such that the set A of attributes separates as

$$A = N \bigcup D. \tag{1}$$

The set N consists of the *numerical attributes* of the owner which take integer or real numbers as values, the set D consists of discrete non-numerical values.

Correspondingly, the stencil of an attribute is determined by the attribute's name and its range, being of a



Figure 1. UML diagram of the data structure of a profile and its relationship to the owner's attributes, the attribute stencils and the fields of interest. An attribute stencil consists of an owner's attribute name and its (searched or offered) range.

certain set called *type* \mathcal{T} , with

$$\mathcal{T} = \mathcal{N} \bigcup \mathcal{D}, \tag{2}$$

where

$$\mathscr{N} \subset \left\{ \left[a, b \right] : -\infty \le a, b \le \infty \right\},\tag{3}$$

denotes the set of ranges for the numerical attributes, and

 $\mathscr{D} \subset \{E : E \text{ is a finite set or enum}\},\tag{4}$

denotes the set of possible value sets for discrete non-numerical attributes. That is, \mathscr{N} is a set of closed intervals $[a,b] \subset \mathbb{R} \cup \{\infty\}$, and \mathscr{D} is a finite set or enum, specified by the respective owner attributes determined by the system model. We allow the empty set \mathscr{D} as null element both in \mathscr{N} and \mathscr{D} . If a given range $R \in \mathscr{T}$ contains only one element, say $R = \{x\}$, then the stencil is often shortly written "p = x" instead of " $p \in \{x\}$ "; if on the other hand $R = [x, \infty)$ then we may write "p > x" instead of $p \in R$. For instance, "height = 180" means "height $\in [180, 180]$ ", or "height > 180" means "height $\in [180, \infty]$ ".

On the other hand, a field of interest is a name-value pair specifying the field itself as well as its level ranging on a scale from -1 to 1, coded by the interpolation of the following table,

Level	Meaning
-1	aversion
0	indifference
1	enthusiasm

The set of fields of interests is denoted by I and is a subset of words of a specified alphabet Σ ,

$$I \subseteq \Sigma^*. \tag{6}$$

Usually, Σ is the set of Unicode symbols. The set *I* determines the set of all fields of interests available to the system. Depending on the system design, it may be a fixed set of words, or an arbitrary word over the alphabet Σ .

Example 1. In grid computing, a main matching problem is resource discovery and resource allocation [15] [16]. Assume a toy grid consisting of two resource providers Haegar and Bond, and two resource requests by some computational process. In our terminology, Bond and Haegar each provide an offering profile, whereas the requests are represented by search profiles. Moreover, in the widely used matchmaking framework Condor-G [17]-[21], the profiles are called ClassAds (classified advertisements). Let us assume a system of four objects with search and offering profiles according to the following tables.

Search profiles				
Owner = 194.94.2.21	Owner = 194.1.1.3			
$CPU \ge 1.6 GHz$	Memory $\geq 2 \text{ GB}$			
Memory $\geq 1 \text{ GB}$				

Offering	profiles
Owner = bond.cs.ucf.edu	Owner = haegar.fh-swf.de
$CPU \le 3.6 \text{ GHz}$	CPU = 2.5 GHz
Memory $\leq 4.0 \text{ GB}$	Memory $= 1.0 \text{ GB}$

In each column of a profile there is listed its owner and some attributes and their values. Then a solution to the global matching problem is obviously given by the matchings (194.94.2.21, Haegar) and (194.1.1.3, Bond). \Box

3.3. The Search Space

Let be a set \mathfrak{S} of search profiles \mathfrak{s} and a set \mathfrak{O} of offering profiles \mathfrak{o} be given as input, where

$$\mathfrak{S}, \mathfrak{D} \subseteq \mathcal{P}(N \times \mathscr{N}) \cup \mathcal{P}(D \times \mathscr{D}) \cup \mathcal{P}(I \times [-1,1])$$

$$\tag{7}$$

where $\mathcal{P}(A)$ denotes the power set of an arbitrary set A. Then the search space S_G of a *global* matching problem is given by all pairs of search and offering profiles, *i.e.*, $S_G = \mathfrak{S} \times \mathfrak{D}$. In this paper, however, we are considering the local matching problem, given a single search profile \mathfrak{s} , *i.e.*, $\mathfrak{S} = \{\mathfrak{s}\}$, and the search space

$$S = \{\mathfrak{s}\} \times \mathfrak{O}(\mathfrak{s}) \tag{8}$$

where $\mathfrak{O}(\mathfrak{s}) = \{\mathfrak{o} \in \mathfrak{O} : \operatorname{owner}(\mathfrak{o}) \neq \operatorname{owner}(\mathfrak{s})\}$. Let

$$N_{\rm s} \subseteq N, \quad D_{\rm s} \subseteq D, \text{ and } I_{\rm s} \subseteq I$$
 (9)

denote the set of searched numerical attributes, the set of discrete non-numerical attributes, and the set of fields of interests, respectively, specified by the search profile. Then a search profile \mathfrak{s} itself is a set of the given attribute stencils \mathfrak{n}_s , \mathfrak{d}_s , and of fields of interest \mathfrak{i}_s ,

$$\mathfrak{s} = \mathfrak{n}_{s} \bigcup \mathfrak{d}_{s} \bigcup \mathfrak{i}_{s}, \tag{10}$$

where

$$\mathfrak{n}_{s} = \left\{ \left(p, R_{s} \left(p \right) \right) \colon p \in N_{s} \right\} \subset N_{s} \times \mathcal{N}$$

is the set of attribute-range pairs, with the given mapping $R_s: N_s \to \mathcal{N}$ from the set N_s of the searched numerical attributes to their associated desired ranges (R_s associates to each numerical attribute p in N_s an interval $R_s(p) = [a,b]$),

$$\mathfrak{d}_{s} = \left\{ \left(p, E_{s} \left(p \right) \right) \colon p \in D_{s} \right\} \subset D_{s} \times \mathscr{D}$$

is the set of attribute-set pairs, with the mapping $E_s: D_s \to \mathcal{D}$ from the given set D_s of searched discrete attributes to their desired sets or enums, and

$$\mathfrak{i}_{s} = \left\{ \left(p, l_{s} \left(p \right) \right) \colon p \in I_{s} \right\} \subset I_{s} \times \left[-1, 1 \right]$$

is the set of searched fields of interest with their desired levels, with the given mapping $l_s : I_s \rightarrow [-1,1]$. Note that for a usual software system each of the pairs (p, R_s) , (p, E_s) , (p, l_s) can be easily implemented as a table or a hash map. Analogously, an offering profile is given by

$$\mathfrak{o} = \mathfrak{n}_{0} \bigcup \mathfrak{d}_{0} \bigcup \mathfrak{i}_{0}, \tag{11}$$

where the three sets are defined the same way as in the search case, but with the index "s" (for "search") replaced by "o" (for "offering").

Example 2. Assume a small social network for pooling interest groups, consisting of three persons, Alice, Bob, and Carl, who provide search and offering profiles according to the following tables.

Search profiles					
Owner = Alice		Owner = Carl			
Age ∈ [20, 40]		Age ∈ [20, 30]			
		Height > 180			
Tennis = 1.0		Basketball = 1.0			
Chess = 0.5					
Offering profiles					
Owner = Alice	Owner = Bob	Owner = Carl			
Age = 25	Age = 26	Age = 31			
Height = 165	Height = 182	Height = 195			
Tennis = 1.0	Tennis = 0.5	Basketball = 1.0			
Chess = 0.5	Basketball = -1.0				
Basketball = 0.5					

In each column of a profile there is listed its owner, some attributes and their values, and the fields of interests with their levels. For instance, Alice looks for someone between 20 and 40 years of age being enthusiastic in tennis and having some penchant to chess, whereas Carl seeks a tall person in the 20's with highest preference for basketball. Looking at the offering profiles in this social network, one sees that Alice may contact Bob, but Carl cannot find an ideal partner in this community. On the other hand, Alice would be a "better" partner for Carl than Bob, since she is partly interested in basketball. Formally Alice's search profile, for instance, is given as follows. The sets for the searched attributes and fields of interest are

$$N_s = \{age\}, \quad D_s = \emptyset, \quad I_s = \{\text{tennis, chess}\},$$
 (12)

the mapping R_s is given by

$$R_{s}(p) = \begin{cases} [20, 40] & \text{if } p = \text{``age'',} \\ \emptyset & \text{otherwise.} \end{cases}$$
(13)

and the mapping l_s is given by the table

$$\frac{p \quad \text{tennis} \quad \text{chess}}{l_{s}(p) \quad 1.0 \quad 0.5} \tag{14}$$

The mapping E_s does not exist since $D_s = \emptyset$. To sum up, Alice's search profile is given by

$$\mathfrak{s}_{Alice} = \{ (age, [20, 40]) \} \cup \{ (tennis, 1.0), (chess, 0.5), (basketball, 0.5) \},$$
(15)

Note in particular that $n_{s,Alice} = \emptyset$. On the other hand, the offering profiles read

$$\mathfrak{O}(\mathfrak{s}_{\text{Alice}}) = \{\mathfrak{o}_{\text{Bob}}, \mathfrak{o}_{\text{Carl}}\}$$
(16)

where

$$\mathfrak{o}_{Bob} = \{ (age, [26, 26]), (height, [182, 182]) \} \cup \{ (tennis, 0.5), (basketball, -1.0) \}.$$
(17)

$$\mathfrak{o}_{Carl} = \left\{ (age, [31, 31]), (height, [195, 195]) \right\} \cup \left\{ (basketball, 1.0) \right\}.$$
(18)

With the definitions

$$x_{B} = (\mathfrak{s}_{\text{Alice}}, \mathfrak{o}_{\text{Bob}}), \quad x_{C} = (\mathfrak{s}_{\text{Alice}}, \mathfrak{o}_{\text{Carl}}), \tag{19}$$

the search space $S = \{\mathfrak{s}_{Alice}\} \times \{\mathfrak{o}_{Bob}, \mathfrak{o}_{Carl}\} = \{x_B, x_C\}$ consists of two feasible solutions. \Box

3.4. Matching Degree as Objective Function

The matching degree of a search profile and an offering profile is a real number $f \in [0,1]$, with f = 0

meaning "total mismatch" and f = 1 meaning "perfect match". In general, the matching degree will be the weighted sum of several partial matching degrees, one for each property separately. Moreover, the matching degree of a numerical attribute is calculated in a different way than the matching degree of a non-numerical discrete attribute or of a field of interest. A function measuring the matching degree of ranges of an attribute has to quantify how a given offered attribute stencil $[a_o, b_o]$ fits into the stencil pattern given by the corresponding range in the search profile. In case of a numerical attribute, the stencil is given by a closed interval $[a,b] \in \mathcal{D}$ in case of a discrete-value attribute it is a set or enum E. For a field of interest the matching degree is a continuous function of the level of interest of the search profile and the level of interest of the offering profile which is constructed below.

3.4.1. Matching Degree for a Numerical Attribute

To determine the matching degree of a searched value range $[a_s, b_s]$ with a given advwertised value range $x \in \mathbb{R}$, we define the fuzzy step function $h_e(x) = h_e(x; a, b)$ with $a \le b$ and 0 < e < 1, as

$$h_{e}(x;a,b) = \begin{cases} \frac{x}{ae} - \frac{1-e}{e} & \text{if } (1-e)a < x \le a, \\ 1 & \text{if } a < x \le b, \\ \frac{1+e}{e} - \frac{x}{be} & \text{if } b < x \le (1+e)b, \\ 0 & \text{otherwise.} \end{cases}$$

$$1 \quad ae \quad be \\ 1 & \text{otherwise} \end{cases}$$

$$(20)$$

The parameter *e* is called the *fuzzy level*. It denotes the relative length of the fuzzy transition region. The smaller *e*, the narrower this region, and the more accurate an offered attribute value must fit into the searched interval. In the limit $e \rightarrow 0$, the function h_e is a sharp step function, for a = b and e > 0 it represents a triangle function, and for $a \rightarrow -\infty$ or $b \rightarrow \infty$ it tends to the Heaviside step functions $H_b(-x)$ or $H_a(x)$, respectively.

For instance, if the searched attribute is "height > 180" and an offered attribute is "height = 165" then for a fuzzy level of e = 10%, we have

$$h_{0.1}(165;180,\infty) = \frac{165}{18} - \frac{0.9}{0.1} = 0.1\overline{6},\tag{21}$$

i.e., the matching degree is 16.7%. Then the matching degree of two numerical ranges $[a_s, b_s]$ as search range and $[a_o, b_o]$ as offered range are given by

$$m_{n}([a_{s},b_{s}],[a_{o},b_{o}];e) = \max[h_{e}(b_{o};a_{s},b_{s}),h_{e}(b_{s};a_{o},b_{o})].$$
(22)

3.4.2. Matching Degree for a Non-Numerical Attribute

If the values of specific attribute are constrained to be of a finite set, or an enum, say E then the matching degree is determined by the Boolean characteristic function χ_E defined by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{otherwise.} \end{cases}$$
(23)

If the searched attribute, for instance, is "name \in {'Smith', 'Taylor'}" and the offered attribute is "name = 'Tailor'" then "E = {'Smith', 'Taylor'}" and " χ_E ('Tailor') = 0", *i.e.*, the matching degree is zero. By contrast the matching degree for "name = 'Tailor'" is " χ_E ('Tailor') = 1". Since the owner of an offering profile can offer at most one value for an attribute, we have

$$m_d(E, \{x\}) = \chi_E(x).$$
 (24)

3.4.3. Matching Degree of a Field of Interest

First we notice that the matching degree as a function of the levels of interest l_s for the search profile and l_o for the offering profile must be asymmetric. For instance, if $l_s = 0$ and $l_o = 1$, *i.e.*, the search is indifferent

with respect to the field of interest, and the offering profile has $l_0 = 1$, then the matching degree should be greater than 0, but if the search requires $l_s = 1$ and $l_0 = 0$ then the matching degree should be zero. In the first case, the searcher is indifferent about the field of interest, in the second case he demands high interest.

Definition 3. An interest matching degree function is a function $m: [-1,1]^2 \rightarrow [0,1]$ such that the following conditions are satisfied.

The first condition expresses the *perfect matching of the diagonal*, the second the *search indifference*, and the last the *search necessity*. \Box

A possible matching degree function is given by

$$m_i(l_s, l_o) = \max\left[\varphi(l_s, l_o), 0\right]$$
(26)

where

$$\varphi(x,y) = 1 - \frac{(c^2 - 1)(x - y)^2}{c^2 - 2 + (x - cy)^2}$$
(27)

with

$$c = \frac{1 + \sqrt{7}}{2} \approx 1.823.$$
 (28)

By construction, $m(l_s, l_o)$ satisfies the conditions in (25) and therefore is an interest matching degree function.

It is asymmetric with respect to its arguments, since we have $m(l_s, l_o) \neq m(l_o, l_s)$ if and only if $l_s^2 \neq l_o^2$. On the other hand, it is an even function, *i.e.*, $m(l_s, l_o) = m(-l_s, -l_o)$.

3.4.4. The Objective Function

Putting together all partial matching degrees considered above, *i.e.*, the matching degree (22) for numerical attributes, the matching degree (24) for numerical attributes, the matching degree (26) of fields of interest, we construct a function $f: S \rightarrow [0,1]$ as a weighted sum of them. We notice that any $x_s \in S$ represents a feasible solution of the matching problem and has the form $x_s = (\mathfrak{s}, \mathfrak{o})$ where \mathfrak{s} is the given search profile (10) and \mathfrak{o} is one of the given offering profiles (11) in the network. Then f is defined for each $x_s \in S$ by

$$f(x_{s}) = \sum_{p \in N_{s}} \frac{m_{n}(R_{s}(p), R_{o}(p))}{|N_{s}| + |D_{s}| + |I_{s}|} + \sum_{p \in D_{s}} \frac{m_{d}(E_{s}(p), E_{o}(p))}{|N_{s}| + |D_{s}| + |I_{s}|} + \sum_{p \in I_{s}} \frac{m_{i}(l_{s}(p), l_{o}(p))}{|N_{s}| + |D_{s}| + |I_{s}|}$$
(29)

where $R_{...}(p)$ and $E_{...}(p)$ denote the attribute ranges of the attribute p, $l_{...}(p)$ the levels of the field of interest p, the ellipsis "…" referring to "s" for the search profile and "o" for the offering profiles, respectively; the vertical bars $|\cdot|$ embracing a set denote the number of its elements.

Thus for the computation of the matching degree, the attributes and fields of interest of the search profile \mathfrak{s} are leading, *i.e.*, it is \mathfrak{s} which determines what is tried to be matched. If therefore a property p of the search profile does not occur in the offering profile, then the matching degree functions $m_n(p)$ and $m_d(p)$ vanish by definition. If, however, a searched field of interest $p \in \mathfrak{i}_s$ does not occur in the offered profile, then it is the level $l_o(p)$ which vanishes by definition. The crucial difference between null values of attributes and null values of fields of interest in the offering profile is therefore the following: searched attributes are mandatory, in the sense that a lack of this attribute in an offering profile means a mismatch with respect to it; however, if a field of interest does not occur in the offered profile, it is indifferent to its owner, but depending on the level of interest in the search profile the matching degree may be positive nonetheless.

Example 2 (revisited). For Alice's search space we have the two solutions (19), i.e.,

$$f(x_B) = \frac{m_n([20,40],[26,26])}{3} + \frac{m_i(1,0.5) + m_i(0.5,0)}{3} = \frac{1}{3} + \frac{0.5636 + 0.6308}{3} = 0.7315$$
(30)

and

$$f(x_c) = \frac{m_n([20,40],[31,31])}{3} + \frac{m_i(1.0,0) + m_i(0.5,0)}{3} = \frac{1}{3} + \frac{0 + 0.6308}{3} = 0.5436$$
(31)

Hence Bob's offering profile has a matching degree of 73.15% with Alice's search profile, whereas Carl's matches it only by 54.36%. \Box

Notice that the objective function (29) is constructed in such a way that each searched item p of a search profile has equal weight. If, however, each item shall have its own weight w(p), then the objective function (32) is easily modified to

$$f(x_{\rm s}) = \sum_{p \in N_{\rm s}} \frac{w_n(p)}{w_{\rm tot}} m_n(R_{\rm s}(p), R_{\rm o}(p)) + \sum_{p \in D_{\rm s}} \frac{w_d(p)}{w_{\rm tot}} m_d(E(p), T_{\rm o}(p)) + \sum_{p \in I_{\rm s}} \frac{w_n(p)}{w_{\rm tot}} m_i(l_{\rm s}(p), l_{\rm o}(p))$$
(32)

where w_{tot} is the total sum

$$w_{\text{tot}} = \sum_{p \in N_{\text{s}}} w_n(p) + \sum_{p \in D_{\text{s}}} w_d(p) + \sum_{p \in I_{\text{s}}} w_i(p)$$
(33)

of all weights $w_n: N_s \to \mathbb{R}^+$, $w_d: D_s \to \mathbb{R}^+$, $w_i: I_s \to \mathbb{R}^+$.

3.5. Mathematical Formulation

Let $S = \{\mathfrak{s}\} \times \mathfrak{O}(\mathfrak{s})$ be the set of pairs $x_s = (\mathfrak{s}, \mathfrak{o})$ of a given search profile \mathfrak{s} as in (10) and the offering profiles $\mathfrak{O}(\mathfrak{s})$ available to it in the social network as in (8), each offering profile being of the form as given in (11). Given the objective function $f: S \to [0,1]$ either as defined in (29) or as in (32), the profile matching problem is the maximum problem

$$\underset{x_{s} \in S}{\operatorname{argmax}} f(x_{s}). \tag{34}$$

A simple algorithm to solve this maximum problem is to exhaustively compute the matching degrees of all possible profile pairs $x \in S$ and to store those with the maximum value. Let n_p be the maximum number of properties (*i.e.*, numerical and discrete attributes as well as fields of interest) a profile in the network can have, and let n_{net} be the number of offered profiles. Then the number of necessary profile comparisons is estimated by $O(n_p^2 \cdot n_{net})$. For instance in Example 2 above the search space S of Alice we have $n_p = 4$, $n_{net} = 2$, *i.e.*, the solution of problem (34) by brute force requires ≤ 32 property comparisons; it can be directly verified that in fact we need 12+9=21 comparisons in this case.

4. Discussion

In this paper, a mathematical model of the matching of search and offering profiles in electronic social networks is proposed. Basing on the data structure described by **Figure 1** and distinguishing between matching of attribute ranges via stencils and matching of fields of interests via comparison, the matching problem is formulated as an optimization problem, with the search space consisting of a fixed search profile and several offering profiles as in (8), and the matching degree as its objective function in (29). The main difficulty is to define a function measuring adequately the matching degree of two fields of interest and obeying the necessary conditions listed in Definition 3. A proposed solution is the function given in (26) and depicted in **Figure 2**. The implementation of a matching service in an electronic social network basing on this matching optimization is straightforward.

The applicability of a profile matching algorithm is not restricted to electronic social networks but could be adapted for resource discovery in grid computation or in matchmaking energy resources in grids. In particular energy transition projects aiming to integrate renewable energy into electrical power systems have to solve the problem of matching energy supply and demand, caused by the high variability of renewable energy supply such as wind or solar power ([22], ğğ7.5, 8.2.1). This way profile matching might become one of the relevant technologies to support eager energy transition projects like the *Energiewende* towards a sustainable economy.



Acknowledgements

I am indebted to Thomas Kowalski for valuable discussions.

References

- [1] Augusto, L.R., de Castro, R.C.F., Franco, L.G., Machado, L.G.P. and Seo, C.E. (2007) Multiple Interest Matchmaking in Personal Business Networks. US Patent US7953673 B2. <u>http://google.com/patents/US7953673</u>
- [2] Gribova, V. and Kachanov, P. (2009) An Approach to Automated User Interest Matching in Online Classified Advertising Systems. In: Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J. and Bevilacqua, V., Eds., *Emerging Intelligent Computing Technology and Applications*, Lecture Notes in Computer Science, Vol. 5754, Springer, Berlin, 665-673. <u>http://dx.doi.org/10.1007/978-3-642-04070-2_72</u>
- [3] Al Rabea, A.I. and Al Fraihat, M.M.A. (2012) A New Matchmaking Algorithm Based on Multi-Level Matching Mechanism Combined with Fuzzy Set. *Journal of Software Engineering and Applications*, 5, 110-118. <u>http://dx.doi.org/10.4236/jsea.2012.53018</u>
- [4] Barnes, J.A. (1954) Class and Committees in a Norwegian Island Parish. Human Relations, 7, 39-58. <u>http://dx.doi.org/10.1177/001872675400700102</u>
- [5] Easley, D. and Kleinberg, J. (2010) Networks, Crowds, and Markets. Reasoning about a Highly Connected World. Cambridge University Press, Cambridge. <u>http://dx.doi.org/10.1017/CBO9780511761942</u>
- [6] Wasserman, S. and Faust, K. (1994) Social Network Analysis. Cambridge University Press, Cambridge. <u>http://dx.doi.org/10.1017/CBO9780511815478</u>
- [7] Newman, M.E.J. (2010) Networks. An Introduction. Oxford University Press, Oxford.
- [8] Hill, R.A. and Dunbar, R.I.M. (2003) Social Network Size in Humans. *Human Nature*, 14, 53-72. http://dx.doi.org/10.1007/s12110-003-1016-y
- [9] O'Reilly, T. (2007) What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, **1**, 17.
- [10] Agarwal, S. and Lamparter, S. (2005) sMart—A Semantic Matchmaking Portal for Electronic Markets. Proceedings of the 7th International IEEE Conference on E-Commerce Technology, Munich, IEEE Computer Society. <u>http://citeseer.ist.psu.edu/agarwal05smart.html</u>
- [11] Berghella, M., Calí, A., Capata, A., Catarci, T., Cerrocchi, P., Masi, P., Oppedisano, M., Trevisani, E. and Vitaletti, A. (2005) SmartDate: User Adaptation in Location-Based Mobile Matchmaking. PSMD 05: Proceedings of the 2005 International Workshop on Plastic Services for Mobile Devices, Rome, 12 September 2005. <u>http://www.dis.uniroma1.it/psmd05/papers/021.pdf</u>
- [12] Cal, A., Calvanese, D., Colucci, S., Di Noia, T. and Donini, F.M. (2004) A Description Logic Based Approach for Matching User Profiles. *Proceedings of the 17th International Workshop on Description Logics (DL'04)*, **104**. CEUR Workshop Proceedings. <u>http://sisinflab.poliba.it/publications/2004/CCCDD04</u>
- [13] de Vries, A. (2007) XML Framework for Concept Description and Knowledge Representation. Aachen.
- [14] Bai, X., Yu, H., Ji, Y. and Marinescu, D.C. (2004) Resource Matching and a Matchmaking Service for an Intelligent Grid. International Journal of Computational Intelligence, 1, 197-205. <u>http://www.cs.ucf.edu/~xbai/ICCI04.pdf</u>
- [15] Foster, I. and Kesselman, C. (2004) The Grid 2. Blueprint for a New Computing Infrastructure. Elsevier, Amsterdam.
- [16] Prodan, R. and Fahringer, T. (2006) Grid Computing: Experiment Management, Tool Integration, and Scientific Workflows. Springer-Verlag, Heidelberg & Berlin.

- [17] Raman, R., Livny, M. and Solomon, M.H. (1998) Matchmaking: Distributed Resource Management for High Throughput Computing. *The 7th International Symposium on High Performance Distributed Computing*, Chicago, 28-31 July 1998, 140-146. <u>http://dx.doi.org/10.1109/HPDC.1998.709966</u>
- [18] González-Castaño, F.J., Vales-Alonso, J., Livny, M., Costa-Montenegro, E. and Anido-Rifón, L.E. (2003) Condor Grid Computing from Mobile Handheld Devices. *Mobile Computing and Communications Review*, 7, 117-126. http://dx.doi.org/10.1145/881978.882005
- [19] Lodygensky, O., Fedak, G., Cappello, F., Néri, V., Livny, M. and Thain, D. (2003) XtremWeb & Condor Sharing Resources between Internet Connected Condor Pools. *Proceedings of the 3rd International Symposium on Cluster Computing and the Grid*, 382-389.
- [20] Thain, D. and Livny, M. (2003) Building Reliable Clients and Services. Foster and Kesselman, 15, 285-318.
- [21] Thain, D., Tannenbaum, T. and Livny, M. (2005) Distributed Computing in Practice: The Condor Experience. *Concurrency, Practice and Experience*, **17**, 323-356.
- [22] Edenhofer, O., Madruga, R.P., Sokona, Y., Seyboth, K., Matschoss, P., Kadner, S., Zwickel, T., Eickemeier, P., Hansen, G., Schlömer, S. and von Stechow, C. (2012) Renewable Energy Sources and Climate Change Mitigation. Special Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. <u>http://www.ipcc.ch/pdf/special-reports/srren/SRREN_Full_Report.pdf</u>



IIIIII II

 \checkmark

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.

