Market-based Decentralized Profile Infrastructure: Giving Back to the User

Olfa Nasraoui,
Dept. of Computer Engineering & Computer Science,
Speed School of Engineering,
University of Louisville
Olfa.nasraoui@louisville.edu

Abstract

After years of loyalty with an e-commerce company (e.g. Amazon.com), the users or consumers still do not own their profiles (products viewed, rated or purchased); neither can they move it with them freely from one business or context to another. In fact many organizations profit directly from the sale and exchange of a limited view of a user's profile without even their knowing or consenting. However this profile can form extremely precious information that can benefit personalization and combat information overload in a variety of different domains both for the user and for the business. There are likely to be wide correlations between a user's tastes in books, movies, and many other products or content items that are not necessarily sold on the same website, and can form the basis of a more global and accurate collaborative filtering recommendation process. In this paper, we propose a market-based profile infrastructure to put the users more in control of their own profiles, and mechanisms to allow the user to profit from this profile, thus democratizing the recommendation process. We outline some of the challenges and research issues that need to be addressed and a few possible solutions

1. Motivations

To date, a loyal consumer/user of a Web business still cannot *own* their profile (e.g. products viewed or purchased), and cannot move it with them freely from one business or context to another. However this profile forms extremely precious information that can benefit personalization and combat information overload in a variety of different domains (not limited to books) for the following reasons:

1. There are likely correlations between a user's tastes in books, movies, and many other products or content items that are not sold on the same website, including: food, wine, clothing, sports, arts, "content" like "news and blogs", as well as music, videos, etc. Hence there is a need for ***"single profile integration" across multiple websites***.

2. The above correlation can only be enriched if further integrated with *many other user profiles* in a collaborative filtering framework, hence predicting a user's interests not just from the same user but also from other similar users' interests "by association". Hence there is a need for ***"multiple profile integration"***.

3. Currently each ***website*** has a limited view of user profiles (limited to the scope of what is sold/served by this website). Extending the scope to other websites would give a more global view of a user profile.

4. Currently each ***user's scope*** is limited: only their own profile is available, hence no sharing with other users. As a result, a user cannot possibly be the one who "invokes" a collaborative filtering recommendation. Instead it is always the server that initiates and benefits from such collaborative filtering.

All the above reasons make compelling the need for *an intermediate solution that fosters both single profile integration (across multiple websites) and multiple profile integration (across the same website). The solution would stand midway between the server (or the business) and the client (or the user).*

2. Market-based Profile Infrastructure

We propose a market-based profile infrastructure that is based on the creation of a dynamic and distributed market based system where each user's profile is maintained and updated and exchanged with other users and especially with online server/merchant agents in a bidding-like system/hence a market based economy. The principles of this system are:

- The technical platform consists of an *intermediate* solution that fosters both single profile integration (across multiple websites) and multiple profile integration (across the same website). The solution would stand midway between the server (or the business) and the client (or the user). This is similar to *peer to peer information sharing*: there is no single central control on the user profiles, though there could be a central repository of many user profiles in server communities or clustered repositories.

- A user, who ***owns his/her own profile***, ***earns some credit*** each time that their profile is invoked by a recommendation process (or transaction), hence accumulating credits that can be used toward any purchase within this marketplace, or credit that simply gets accumulated in the user's electronic account (e.g. an e-wallet or paypal account).
- The individual credits may be very small (e.g. a fraction of 1 cent per invocation of the user profile), but may accumulate to a profitable level with a large number of invocations, especially in an increasingly competitive and global e-commerce marketplace. The crediting should work in a similar manner to *Google Adsense* and would operate in a similar manner to e-bay auctions, except that interactions would be completely automated.
- The user, for the first time, not only ***"owns"*** their own profile, but also ***"can sell it like a commodity"*** and benefit from it. The entire solution is enabled by a ***technical infrastructure*** that allows such exchange of information to take place. ***Such a profit has previously been restricted only to the businesses (who frequently sell, buy and trade user profile data) without any benefit or gain to the user, and without their consent.***

Realizing this infrastructure requires: ***(i) Technical*** formulation of a logical and feasible architecture for the above infrastructure based on (1) current internet technologies such as Web service protocols and P2P networks [10], (2) additional data collection, routing, exchange, and privacy agents, and (3) market based agents that serve to implement the credit/payment operations that occur with the invocation of each user profile; ***(ii) Analytical study*** of the dynamics and equilibrium if any within this infrastructure under varying scenarios, and ***(iii) Simulation*** of the proposed infrastructure under different scenarios and variable parameters to study the effect of each type of data integration and exchange.

3. Description of Proposed Infrastructure

One way to implement the proposed infrastructure is by emulating some ideas both from peer to peer (P2P) networks where information is exchanged in a decentralized manner, and from currently operating Web service platforms, where services are advertised/registered via a registration mechanism in a registry, and different actors interact and find each other by checking these registries. Everything else is typically automated by well controlled ontologies and languages. One main difference between a user profile marketplace and the Web service infrastructure is that the user does not necessarily have a "server" or a unique URL where they can be reached as in traditional Web servers and Web services. One way to circumvent this limitation is by using the user's own computer or personal device like a cell phone to store such information. Also, an intermediate architecture where user information is logged on their device in real time and then transmitted to one of several repositories that could be on actual servers/registries, but *not necessarily a central* server. They could be for example "super peers" as in P2P networks, which are designated peers that act in a role halfway between a true server and a true peer in the exchange and routing of information.

Recent proliferation of social networking websites offers yet another vital "carrier" and platform for such an infrastructure if a user profile can be stored and invoked securely as part of a market strategy.

In the future, with the widespread use of RFID readers and assuming that RFID tags will be required on most merchandise, the collection and exchange of user profiles will be greatly facilitated and automated. For instance RFID readers can be implanted on the user's personal device, such as a cell phone to integrate *even offline* transactions with online transactions. *Online* transactions can be logged entirely by *software* on the client computer. However, *RFID can bridge the gap between online and offline worlds*, especially if the cell phone, equipped with an RFID reader also communicates this information with the user profile's designated repository, be it the user's own computer or one of several remote intermediate repositories. Finally, cell phones (as well as IP addresses of various internet connected devices) can enrich the user profile by also adding *geographical location information*: Hence patterns can be discerned at the *local versus global level for better prediction*.

Another possible platform would be based on integrating consumer profiles together with their e-wallets. E-wallets, while still not being widely accepted, have recently been implemented and evaluated based on the existing "*mondex*" card that shares the flexibility and privacy features of cash, while allowing participation in electronic commerce even at micro-economical level (very small value transactions) [1].

Privacy must finally be integrated within the above infrastructure. First, there must be a taxonomy of what can be logged or not logged among the user's interactions and purchases. For example some users do not like to log anything that is medical or health related, nor anything that is financial/bank related, etc. These restrictions will be directly implemented on the user's "data collection" side, and hence they will be enabled and respected.

3.1 Challenges

There are several challenging problems attached to the proposed infrastructure, some of which are listed in Table 1, together with related issues and sub-problems and ideas for possible solutions.

Table 1: Challenges facing a market-based profile infrastructure

Issue	Sub-problems and possible solutions
Distributed computing	- Network topology, information exchange mechanisms, - Scalability and adaptation/incremental learning, indexing for fast access/search/ retrieval/ computing/ update/ storing - Distributed real-time DBs: concurrency and updates - Service-based computing and access (Web services)
Monetary system to realize the distributed transactions	Existing finance and market models Current e-commerce infrastructure e-wallets such as mondex; or paypal accounts
Physical infrastructure	overlaid on existing network structure: e.g. current P2P networks [10], existing social networking websites (e.g. MySpace.com), dedicated distributed repositories affiliated with the user's bank, and earning a commission for their indexing
Ontological engineering: to be able to recognize related items, related companies, etc (i.e. a bar code is not sufficient! For example, we need to know that items 1123 and 2344 are both books or both types of wheat bread…etc)	Need category information (taxonomy) or textual description of an item (better than SKU codes)
Theoretical analysis and modeling	-Study and simulate the market dynamics -Study equilibrium vs. parameters: what are the cost and benefit for each player? -Optimization: of cost/profit -Graph models: tri-partite models the relations {Users X Companies X Items}: This in turn can form the basis for community discovery to help localize Collaborative Filtering (CF) -Query incentive networks, Social networks
Privacy Preservation: What are the risks/Threats? Hacking, vandalism, fraud, attacks	- Authentication and secure access -Privacy-preserving CF (on distributed platform?): separate the basket from the identity/demographics (never link to a credit card or soc. Security) -the main entity is a basket versus time, possibly with an expiration date (after this date, the user must receive "new" credit for each new invocation of their profile) -Existing and developed research/techniques/frameworks already exist to handle many of the above threats [2,3]
Ethics/legal	-Who wins, what is fair? -Governance: who stores and controls the data? -How does this infrastructure interact and play with existing laws? i.e. the legality of accessing and processing the user's data (with their consent)

3.2. Models of Interaction for Market-based Profile infrastructure

Below, we describe a few potential models that can support market-based profiles.

3.2.1 Constrained Markets

We start by specifying the *players* that participate in a market based profile system, their roles, and interactions. Like many markets, this market consists at least of sellers and buyers. The item*/product sellers* are the companies that traditionally sell products, while the *item/product buyers* are customers who have made prior transactions with these or other sellers. A seller is interested in computing good recommendations for items that they sell based on the profiles of available buyers with matching profiles. A recommendation has a value to the seller if it improves the seller's recommendation model (we will come to this later) or if it results in a sale. The buyers are interested to offer their profile information in return for a reward from the seller. In essence all the players are sellers: they either sell products (or services or commodities), or they sell their profiles (in this case they are the consumers). Hence, from now on, to avoid any confusion in their roles, we will refer to the players as the *companies* (trying to *buy profiles* to compute recommendations) and the *consumers* (who want to earn rewards by *selling their profiles*). The value of a profile for a company is proportional to its potential for improving the recommendation model, and this can be hard to quantify. A raw recommendation model, whether it is a user-item rating matrix or an item-item association matrix will generally benefit by increasing its density, or inversely decreasing its sparsity. This is very similar to saying that each additional rating (or item present in a profile for sale) has an incremental value to be added to improve the model. However not all items are equal, since some "valuable" items may have very low ratings and are therefore needed more than items that are already densely rated in the recommendation model. For example, items may be "valuable" in as far as they are *essential* to contribute to forming *more accurate* distance-based user neighborhoods in user-based CF, or simply because they are of high value (for instance they have a high profit value for the company). While the latter can be easily quantified by a company, the former seem to be related to the densification of the user-item rating matrix as discussed above. Below we formulate the optimization from the points of views of the company and the consumer as primal and dual complementary linear programs.

The Company's Optimization Problem (Primal): Essentially, companies are in pursuit of buying x_j consumer profiles from each class (j) of profiles that meet certain expectations in terms of which items or attributes (i) they contain, and they want to do so at a minimum combined cost ($c^t x = \Sigma_j c_j x_j$), where c is a cost vector, with c_j, the cost assigned to profile j. The constraints can be formulated as a set of inequalities of the form $A^t x \geq r$, where matrix A's component a_{ij} = the number of rating of item (i) in profile (j) and r_i is the minimum number of desired ratings of item i accumulated from the x profiles.

The Consumer's Optimization Problem (Dual): Meanwhile, consumers are in pursuit of selling their own profile which contains r_i ratings for each item (i) sought by the company, at the optimal price π_i per item i, while being competitive with the individual requirements (in terms of ratings) and costs for all classes of consumer profiles, given by $\Sigma_i \pi_i a_{ij} \leq c_j$, $j = 1,\ldots,n$. Moreover, the consumers would like to sell their profile at maximum revenue ($\pi^t r$), since the revenue from selling the items rated in their profile is given by $\Sigma_i \pi_i^t r_i = \pi^t r$.

The primal and dual problems above are instances of convex linear programming problems which can be solved efficiently for a unique global solution. For a pair (x, π) to be respective optima for the two above dual problems, it is sufficient that they obey the complementary slackness condition [4], which states that π_i ($a_i^t x$- r_i) = 0 for all i, and that ($c_j - \pi^t A_j$) x_j = 0 for all j. Note that A_j is the j^{th} column of A, while a_i^t is the i^{th} row of A. All that this tells us is that there is a potential for equilibrium in the market of companies and consumers, however it is clear that many assumptions (like the dynamic nature of the bidding process) may not hold.

3.2.2 Social networks, Query Incentive Networks, and Game Theoretic Models

Similar to information retrieval on P2P networks, there could be a finite lifetime and a finite reward on each transaction. According to [5], using the cosine similarity between users in a user-based Collaborative Filtering can be viewed as computing the probability that the two users will ever meet at any location while doing a random walk on the user-item graph (where the nodes are users and items, an edge connect a user and an item if this item belongs to the user's profile, and the user-item links are probabilities of following a certain user-item link). This is is easy to verify in the case of binary profiles (user-item vectors consist of 0s and 1s), since *Cosine* (u_i, u_j) = $\Sigma_k u_{ik} u_{jk}$ / $(\Sigma_k u_{ik} \Sigma_k u_{jk})^{1/2} \approx \Sigma_k$ P(k|i) P(k|j). We may extend this idea to a graph with only users as the nodes, and with the edges or links weighted by the cosine similarity between their user profiles (i.e. their item vectors), or are simply the strength of connection in a social network. The item-based Collaborative Filtering version of this problem is when the nodes in the graph correspond to the items only, and we consider two items to meet if they occur in the same transaction at any time. This is particularly the case when the item-item association matrix M is obtained by correlation analysis M_{ik} = correlation (i,k); or by association rule mining of rules of the form $i \rightarrow k$ (a measure of confidence of having an item k given an other item i is the conditional probability P(k|i)).

Another graph model can be based on the graph that represents a social network (for example *MySpace*). In this case nodes are linked by social relations, and these relations form the basis for routing a query and searching for an answer. In this case a search engine can be dedicated to scour and index this social network and be used to query for useful profiles at a later stage.

Regardless of which model to use, it is clear that a random walk on a graph model can form the basis for many CF recommendation strategies. In this case the recommendation process takes place by submitting a query to one or more nodes in the graph, allowing these nodes to pass this query on to their local neighbors, and then waiting for an answer which is returned when a satisfactory answer is found. This is essentially one way to implement information retrieval on a peer to peer network (such as Gnutella), except that without an incentive for users to participate in such operations, the effective active (responding) network at any time could be very limited. To answer this challenge, Kleinberg and Raghavan formulated a model for query incentive networks in [6]. Rather than posing queries to a *centralized index* of the system, users pose queries to the *network* itself. Requests for information are propagated along paths through the network, connecting those with information needs to those with relevant answers. In addition queries are submitted together with incentives for answering them, and these incentives also get propagated along paths in a network, with each participating node earning a portion of the reward, until either an answer is found or the propagating rewards get depleted. In [6], this type of information-seeking process was formulated as a game among the nodes in the network, and this game was shown to possess a natural Nash equilibrium. Furthermore, the authors tried to understand how much incentive would be needed in order for a node to achieve a reasonable probability of obtaining an answer to a query from the network, by studying the size of query incentives as a function of the rarity of the answer and the structure of the underlying network. There are several issues that need to be addressed when considering the query incentive network model. In addition to the basic reward setting strategy for each node in the network, the seller must decide the value of the reward to be offered for an answer (i.e. a profile) from the network. This reward can be based on an estimate of *the value added by incorporating one more customer's profile into CF*. The problem with the above approach is that it may not scale to millions of transactions per second. An index based retrieval may be the only option on several indexed Databases, but the DBs need to be refreshed with every user's new transactions. Another game theoretic approach can be achieved by analyzing the proposed infrastructure when implemented by means of dynamic and real-time automated auctions between companies and consumers. The *equilibria* of such systems, if they can be derived, can shed light on its promises, as well as whether it provides *fair* play for all the players, and *under what conditions*.

Why is a graph-based view of the recommendation process interesting within our framework?

There are two reasons for this interest:

1. The graph model supports a distributed profile base, so that no single authority owns all profiles
2. The graph-based search supports local search where a query is passed from one node to its neighbors, thus limiting threats to privacy.

What are the main challenges facing the graph-based view of the recommendation process?

Clearly, the biggest challenge when using the graph-based model would be how to obtain the desired information (the answer) ***(i)*** quickly and ***(ii)*** at a reasonable cost. Different search and incentive strategies, and different properties of the graph in question will largely dictate the cost in terms of time to reach an answer. However, to answer real-time scalability demands, it would be impractical to perform graph search for each transaction, especially for Web based commerce. Thus it is conceivable that these searches could be performed in the following ways to reduce computational burden:

- Perform searches in bundles of transactions or users, i.e. find the optimal profiles for a batch of transactions.
- Perform searches on sub-graphs, for example after an offline discovery of communities within the original graph [7,8]. A bi-partite graph (*Users* X *Items*) consisting of the consumers and their items (purchased or rated) can form the basis for community discovery using a variety of existing techniques. So would the graph (*Users* X *Companies*). Another possibility is a tri-partite graph model (*Users* X *Items* X *Companies*) that combines both players via the items bought or sold by users and companies respectively.

3.2.3 Bio-inspired Models

Because the proposed infrastructure consists of a dynamic network of actors, each trying to optimize their own criterion, one can take advantage of comparing it with certain biological organisms and societies that exhibit such behavior in the quest of survival. These include:

- **Co-evolutionary models**: where two or more species co-evolve simultaneously to optimize their survival.

- **Ant Colonies**: consist of a community of simple agents (ants) that succeed well in achieving collaborative tasks, using a set of simple rules governing each individual ant, further enriched by communication, thus giving rise to a special collaborative intelligence known as *stigmergy* [9].

- **Swarms:** Swarms of agents (e.g. bees) move in such a way that each agent is aware of the movement and fitness of its neighbors, thus resulting in complex behavior.

6. Conclusions

After years of loyalty with an e-commerce company (e.g. Amazon.com), the users or consumers still do not own their profiles (products viewed, rated or purchased); neither can they move it with them freely from one business or context to another. In fact many organizations profit directly from the sale and exchange of a limited view of a user's profile without even their knowing or consenting. In this paper, we proposed a market-based profile infrastructure to put the users more in control of their own profiles, and mechanisms to allow the user to profit from this profile, thus democratizing the recommendation process. From our discussion, we considered several challenges and research issues that need to be addressed and a few possible solutions, which are far from complete, and need further investigation.

7. References

[1] K.Y. Tam and S.Y. Ho, A Smart Card Based Internet Micropayment Infrastructure: Technical Development and User Adoption, Journal of Organizational Computing and Electronic Commerce, 2007, Vol. 17, No. 2, Pages 145-173

[2] Polat, H. and Du, W. 2003. Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques. In *Proceedings of the Third IEEE international Conference on Data Mining*. ICDM'03, Nov. 2003, Washington, DC, 625.

[3] Canny, J. 2002. Collaborative filtering with privacy via factor analysis. In *Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Tampere, Finland, Aug. 2002). SIGIR '02. ACM Press, New York, NY, 238-245.

[4] C. Papadimitriou and K. Steiglitz, Combinatorial Optimization, Algorithms and Complexity, Dover, NY, 1982.

[5] E. Cohen and D.D. Lewis, Approximating Matrix Multiplication for Pattern Recognition Tasks, Journal of Algorithms, Volume 30, Issue 2, February 1999, Pages 211-252.

[6] J. Kleinberg and P. Raghavan, Query Incentive Networks. In FOCS '05: 46th Annual IEEE Symposium on Foundations of Computer Science. Pittsburgh, PA, 132-141.

[7] G.W. Flake, S. Lawrence, C.L. Giles, F.M. Coetzee, Self-organization and identification of Web communities, IEEE Computer, 35 (3), 2002, 66-70.

[8] M.E.J Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E - Statistical, nonlinear, and soft matter, 2004.

[9 M Dorigo, E Bonabeau, G Theraulaz, Ant algorithms and stigmergy, Future Generation Computer Systems, 2000.

[10] T Suel, C Mathur, JW Wu, J Zhang, A Delis, M. Kharrazi, X. Long, K. Shanmugasundaram, ODISSEA: A Peer-to-Peer Architecture for Scalable Web Search and Information Retrieval, International Workshop on the Web and Databases, 2003.