Cross System Personalization by Learning Manifold Alignments

Bhaskar Mehta
Fraunhofer IPSI,Dolivostrasse 15, Darmstadt 64293, Germany

Introduction

The World Wide Web provides access to a wealth of information and services to a huge and heterogeneous user population on a global scale. One important and successful design mechanism in dealing with this diversity of users is to *personalize* Web sites and services, i.e. to customize system contents, characteristics, or appearance with respect to a specific user. Each system independently builds up user profiles and may then use this information to personalize the system's content and service offering. Such isolated approaches have two major drawbacks: firstly, investments of users in personalizing a system either through explicit provision of information or through long and regular use are not transferable to other systems. Secondly, users have little or no control over the information that defines their profile, since user data are deeply buried in personalization engines running on the server side.

Cross system personalization(CSP) (Mehta, Niederee, & Stewart 2005) allows for sharing information across different information systems in a user-centric way and can overcome the aforementioned problems. Information about users, which is originally scattered across multiple systems, is combined to obtain maximum leverage and reuse of information. My previous approach to cross system personalization relied on each user having a *unified profile* which different systems can understand. The unified profile contains facets modeling aspects of a multidimensional user which is stored inside a "*Context Passport*" that the user carries along in his/her journey across information space. When a user goes to a new system which, the user's *Context Passport* is presented to a system which can then understand the context in which the user wants to use the system. The basis of 'understanding' in this approach is of a semantic nature, i.e. the semantics of the facets and dimensions of the unified profile are known, so that the latter can be aligned with the profiles maintained internally at a specific site. The results of the personalization process are then transfered back to the user's Context Passport via a protocol understood by both parties. The main challenge in this approach is to establish some common and globally accepted vocabulary and to create a standard every system will comply with. Without such a convention, the exact mapping between the unified user profile and the system's internal user profile would not be known.

Machine learning techniques provide a promising alternative to enable cross system personalization without the need to rely on accepted semantic standards or ontologies. The key idea is that one can try to learn dependencies between profiles maintained within one system and profiles maintained within a second system based on data provided by users who use both systems and who are willing to share their profiles across systems – which we assume is in the interest of the user. Here, instead of requiring a common semantic framework, it is only required that a sufficient number of users cross between systems and that there is enough regularity among users that one can learn within a user population, a fact that is commonly exploited in social or *collaborative filtering*.

Automatic Cross System Personalization

For simplicity, we consider a two system scenario in which there are only two sites or systems denoted by A and B that perform some sort of personalization and maintain separate profiles of their users; generalization to an arbitrary number of systems is relatively straightforward and is discussed later. We assume that there is a certain number of N_c common users that are known to both systems. For simplification, we assume that the user profiles for a user u_i are stored as vectors $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^n$ and $\mathbf{y}_i \in \mathcal{Y} \subseteq \mathbb{R}^m$ for systems A and B, respectively. Given the profile $\mathbf{x}_i$ of a user in system A, the objective is to find the profile $\mathbf{y}_i$ of the same user in system B, so formally we are looking to find a mapping

$$F_{AB} : \mathbb{R}^n \to \mathbb{R}^m, \qquad \text{s.t.} \quad F_{AB}(\mathbf{x}_i) \approx \mathbf{y}_i \qquad (1)$$

for users u_i. Notice that if users exist for which profiles in both system are known, i.e. a training set $\{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, \ldots, l\}$, then this amounts to a standard supervised learning problem. However, regression problems typically only involve a single (real-valued) response variable, whereas here the function F_{AB} that needs to be learned is *vector-valued*. In fact, if profiles store say rating information about products or items at a site, then the dimensionality of the output can be significant (e.g. in the tens of thousands). Moreover, notice that we expect the outputs to be highly correlated in such a case, a crucial fact that is exploited by recommender systems. For computational reasons it is inefficient and often impractical to learn independent regression functions for each profile component. Moreover, ignoring inter-dependencies can seriously deteriorate the prediction accuracy that is possible when taking such correlations into account. Lastly, one also has to expect that a large fraction of users are only known to one system (either A or B). This brings up the question of how to exploit data without known correspondence in a principled manner, a problem generally referred to as *semi-supervised learning*. Notice that the situation is symmetric and that unlabeled data may be available for both systems, i.e. sets of vectors $\mathbf{x}_i$ without corresponding $\mathbf{y}_i$ and vice versa. In summary, we have three conceptual requirements from a machine learning method: a) Perform vector-valued regression *en bloc* and not independently, b) Exploit correlations between different output dimensions (or response variables); and c) Utilize data without known correspondences. In addition, the nature of the envisioned application requires: a)Scalability of the method to large user populations and many systems/sites, and b)Capability to deal with missing and incomplete data.

There are some recent learning methods that can be utilized for vector-valued regression problem, but some of them do not fulfill the above requirements. Kernel dependency estimation (KDE) is a technique that performs kernel PCA on the output side and then learns independent regression functions from inputs to the PCA-space. However, KDE can only deal with unlabeled data on the output side and requires the solving of computationally demanding preimage problems for prediction (Bakir, Weston, & Schlkopf 2004). Another option is Gaussian process regression with coupled outputs (Keerthi & Chu 2006). Here it is again difficult to take unlabeled data into account while preserving the computational efficiency of the procedure. The same is true for more traditional approaches like Multi-Layer-Perceptrons with multiple outputs.Instead of using regression methods, we thus propose the use of *manifold learning* in this context. Manifold learning methods generalize linear dimension reduction techniques that have already been used successfully in various way for collaborative filtering. Moreover, they are usually motivated in an unsupervised setting that can typically be extended to semi-supervised learning in a rather straightforward manner. More specifically, we propose to use the *Locally Linear Embedding* (LLE) approach (Saul & Roweis 2003) as our core method. LLE constructs a low-dimensional data representation for a given set of data points by embedding the points in a way that preserves the local (affine) geometry. Compared to other manifold learning and non-linear dimension reduction algorithms, such as Sammon's MDS or Isomap, the LLE approach is computationally attractive and highly scalable, since it only relies on distances within local neighborhoods. Moreover, as presented in (Ham, Lee, & Saul 2003), constrained LLE (CLLE) can be utilized to learn mappings between two vector spaces by semi-supervised alignment of manifolds. The former work also provides empirical evidence that CLLE can outperfom standard regression methods. The key idea is to embed user profiles from different systems in low-dimensional manifolds such that profiles known to be in correspondence (i.e. profiles of the same user) are mapped to the same point. This means the manifolds will be aligned at correspondence points. A more general version of CLLE has been derived in (Ham, Lee, & Saul 2005), which takes the Laplacian eigenmap approach as the starting point.

Issues & Challenges

Given two collaborative filtering systems A and B with user databases $\mathcal{X}$ and $\mathcal{Y}$, with a subset of users being common to each system, we would like to apply the constrained LLE algorithm discussed in the previous section. However, straight forward application of the algorithm is not possible, because of the sparsity of the user databases. Sparsity makes the matrices $\mathcal{X}$ and $\mathcal{Y}$ nearly singular and LLE cannot compute the manifold embeddings. To deal with this, we need explore a few approaches for filling missing data. Another significant challenge is to deal with privacy of the user's data. In an $n-$system scenario, much more information about a user can be accumulated that with any of the single systems a user interacts with. Developing encrypted protocol, or distributed algorithms to deal with the privacy issue is a significant challenge. A final challenge is to create a practical framework for cross-system personalization to be realized and examine how current systems need to evolve/adapt according to the requirements imposed by this approach

Current Status

After investing some time with semantic approaches, I am now investigating machine learning techniques for CSP. Based on the MovieLens dataset, I have begun to explore the efficacy of various approaches to deal with missing values and the effect of the number of users crossing over. The evaluation is based on measuring the MAE and $Top - N$ precision of the generated recommendation of a new user from System A crossing over to a system B.

References

Bakir, G.; Weston, J.; and Schlkopf, B. 2004. Learning to find pre-images. volume 16, 449–456. Cambridge, MA, USA: MIT Press.

Ham, J.; Lee, D.; and Saul, L. 2003. Learning high dimensional correspondence from low dimensional manifolds. In *ICML Workshop*.

Ham, J.; Lee, D.; and Saul, L. 2005. Semisupervised alignment of manifolds. In Cowell, R. G., and Ghahramani, Z., eds., *AISTATS 2005*, 120–127. Society for Artificial Intelligence and Statistics.

Keerthi, S., and Chu, W. 2006. A matching pursuit approach to sparse gaussian process regression. In Weiss, Y.; Schlkopf, B.; and Platt, J., eds., *Advances in Neural Information Processing Systems 18*. MIT Press.

Mehta, B.; Niederee, C.; and Stewart, A. 2005. Towards cross-system personalization. In *Universal Access in Human Computer Interaction*.

Saul, L. K., and Roweis, S. T. 2003. Think globally, fit locally: Unsupervised learning of low dimensional manifold. *Journal of Machine Learning Research* 4:119–155.