Expertise Recommendation: A two-way knowledge communication channel

Debbie Richards, Meredith Taylor, Peter Busch

Computing Department, Macquarie University, North Ryde, NSW, 2109, Australia
{richards,mtaylor,busch}@ics.mq.edu.au

Abstract

In our service-based society, expertise is becoming increasingly recognized as a valuable and scarce resource. Thus, the demand for an integrated expert finder that understands whether it has found a current, relevant and credible expert will grow. We offer a three pronged approach which combines data mining, self reporting and referral by others together with a number of strategies which provide feedback and validation of the data gathered via data mining and self-reporting. The feedback mechanism offers a two-way communication channel between the service requester and the service provider, the expert.

1. Introduction

Given the increasing recognition that an organization's most valuable resources are its people and the knowledge they hold, expertise location is becoming an important strategy to accessing and sharing that knowledge. In contrast to expert-systems, also known as knowledge-based systems, which seek to capture what it is that the expert knows so that it can be captured and reused, an expert/ise recommender system suggests who might know about what. The goal of the system is to point someone with a question to the person who has the appropriate knowledge. In some cases the inquirer's main interest is in the answer, in other cases the main interest is to find an expert who will handle the problem [18]. Knowledge about the expert's areas of expertise is needed for such a system. To discover this knowledge it is common to use data mining techniques [12]. Another alternative is to collect this information via self-reporting techniques (e.g. [1]). Individually both approaches have shortcomings.

Data mining relies on the existence of data which is, or is able to be, sufficiently structured to be used as input into one or more algorithms. This raises a number of issues: expertise could be identified from many different sources (e.g. publications, webpages, press releases, projects, grants, etc); these sources will vary across individuals and organizations; the format of these sources will vary across individuals and organization; most of these sources are free-format, unstructured and unclassified (a major hurdle if one wishes to use a supervised learning technique). Not only is the input to data mining an issue, each algorithm has its own strengths and limitations often very closely tied to the structure, amount and type of data; often dependent on the (type of) domain. Furthermore, the "knowledge learnt" tends to be restricted to types of output such as association rules, clusters or classification rules. Even the meaning of interestingness, the identification of "interesting" concepts and validation of the output will vary for domain and/or algorithm.

Due to these various limitations, an alternative to data mining frequently used in recommender systems is the use of surveys/forms to be filled in by the domain expert which usually includes the selection of keywords relating to the individual's areas of expertise. The problem is then reduced to matching the searcher's query terms with the expert's keywords. This technique is often referred to as a yellow-pages approach to finding an expert as that is the way people usually find a plumber, lawyer or doctor. It is a simple and yet effective method for finding people who have certain skills. It works on the expectation that, for instance, only someone with legal training and qualifications will list themselves as a solicitor and that since they listed themselves, they are probably interested in receiving your call. Such assumptions are not always valid. The problems associated with the self-reporting approach include: experts failing to find the time to enter their data in the first place; data entered initially in a burst of enthusiasm by the individual or organization becoming obsolete or out-of-date; inaccurate and/or unvalidated self-reporting of expertise; and the levels and currency of expertise are typically not being captured or maintained.

The approach we offer includes a combination of both techniques, together with a number of verification and validation techniques to improve the consistency, completeness and reliability of the knowledge, with the caveat that we cannot ensure consistency and completeness.

2. Related work

Recommender systems share much in common with search engines which allow a user to enter keywords on a topic they are interested in and produce a list of links to resources based on those keywords and often also on the profile of the user. One of the most well known recommender systems is Amazon.com (http://www.amazon.com/) which allows the user to enter keywords and will search for books and other products based on those keywords. If the user does select one of the recommended products, the system will then suggest other products that it thinks the user might like based on the choices of other users who also selected the same product. That is, the system reasons that if a particular user likes the same product as 20 other users, they may also like some of the other products that those 20 users liked. This is known as collaborative filtering and is a common technique in modern recommender systems (e.g. [1]).

Recommender systems for experts work in much the same way in that they take a user's search query and try to find someone who has the expertise to answer that query. The user will then be given the expert's contact details. These systems are mainly used internally within organizations. Validation that the information provided was wanted and useful is missing in many systems. Amazon attempts to obtain feedback using the message box shown in Fig. 1.

Fig.1. Search feedback from Amazon.com

Aïmeur et. al. [1] further explored the concept of validation of automatic identification of experts. They describe a recommender system called HELP which attempted to locate expertise and experts within an organization. The system included a database of questions asked previously by users and their respective answers. When a user searched for a solution to a particular problem, the question/answer database was first searched to see if their problem or a similar one had already been dealt with. If it had, the user was then presented with the solution, which they could choose to either accept or reject. If they rejected the solution, or if one wasn't found, the system would then search its user database for someone with the potential expertise to solve the problem. Users were required to register their own areas of expertise with the system. The rating system shown in Fig. 2 served to somewhat validate the recommendation the user was given by storing the responses in the profile of the expert. If someone claimed to be an expert on a certain topic but consistently received low ratings, then they would not be recommended by the system if it was possible to recommend someone with a higher rating.

As an alternative to self-reporting recommender system, are fully automatic approaches to locate experts such as Who Knows [14] and SAGE [2] using inputs such email (e.g. Agent Amplified Communications [6]), bulletin boards (e.g. Contact Finder [7]), Web pages (e.g. YENTA [5] and MEMOIR [10]), program code (e.g. Expertise Recommender [9] and Expert Finder [16]), and technical reports (e.g. KCSR Expert Finder [4]). These techniques could be applied to the artifacts of social software systems (i.e. email, WebLogs and Wikis) to provide automatic expert location. However, a review of these systems [12] found problems related to heterogeneous information sources, expertise analysis support, reusability and interoperability.

Expert Finding Systems can be described according to the seven categories [18]: how is one determined to be an expert; the degree of domain specificity or independence; how is the expertise model generated; how is the expertise model queried; what matching techniques are used by the query engine; how is the output presented; and how is the output personally customised. We will use these categories to evaluate our approach. To-date we have performed content analysis on key literature in the field to develop our own graphical model of expertise finding (systems) which reveals the key themes (as nodes) and their relationships (edges). There is not enough space to present that model here or describe how it was created, but in brief key themes from the literature included: organizations, employees, individuals, experts, expertise and knowledge. Technology, such as WWW, CSCW and social network analysis [11], played an important role in creating systems for recommending,
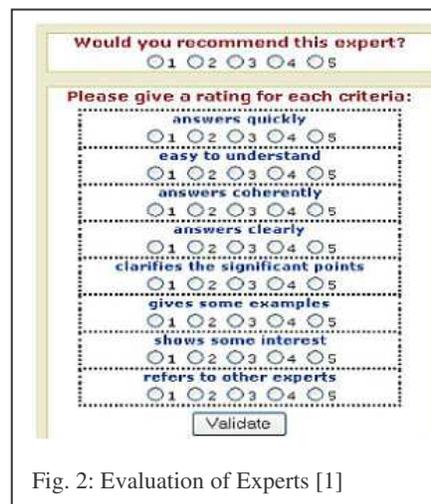

Fig. 2: Evaluation of Experts [1]

identifying and selecting experts within a certain area and developing models of expertise. Documents were used to connect employees and projects, but the latter were not central to locating expertise.

3. Exploratory case studies

As our goal is to create a recommender system that has the confidence and support of its intended users and that goes beyond the yellow-pages model, it was essential that we not just review the literature but also analyse the experiences of practitioners frequently concerned with the task of locating experts and expertise. Thus we conducted two case studies.

3.1 Case study 1

Firstly we conducted interviews with seventeen personnel during the months of September and November 2006 within a Defence R & D organization which is expertise intensive. A series of questions was presented to our interviewees in an attempt to get them to present their experiences on accessing expertise in each of their fields (e.g. "What features or criteria do you use to determine if someone is an expert?", "How do you find what projects/problems people are working on?"). The questions were also designed to elicit barriers they faced to gaining expertise/finding an expert as well as assessing the quality of the expertise they were provided (e.g. "What are the impediments or barriers to identifying an expert?", "What are the impediments or barriers to validating an expert?", "Is there information you are interested in having access to but currently you are unaware of where you could find it or whether what exists is accessible or reliable?"). It is this last parameter that is the most difficult to assess.

It was clear from our initial investigations in 2006, that a fully automated approach which advises who to contact, or a semi-automated approach using techniques such as SNA to answer "who knows who" will not deliver an optimal or widely accepted solution. Currently, there are basically two ways of accessing 'know-how' within the organization: drawing upon one's social networks (a people-centric approach); reading publications/project descriptions to determine who has the relevant expertise (a more algorithmic/automated type of approach). Issues that arose in accessing expert/ise include:
- Establishing trust and quantifying levels of trust
- Turnover of experts and thus loss of expertise
- Access to experts across organizational boundaries
- Measuring currency and relevancy of expertise
- Short timeframes (hours/days) for decision making
- Information is typically classified and restricted.

In this organization the key concern was "how do you find someone to work on a particular project".

Using our interview transcripts we conducted content analysis as we had done with the literature. This revealed that people and projects, and to a lesser extent, tools and groups play central roles, in identifying who is an expert. The structure of the organization also played a major role in being able to find and access an expert.

3.2 Case study 2

The second case study we conducted was within our own university, being another expertise intensive organization. The four key areas of expertise identified to us were the following: Research Office (RO); Development and External Relations Division (DERD); Marketing and Public Relations (MPR) and Teaching and Learning (TL). We interviewed representatives from these four areas, asking them the same questions as in the first case study (removing questions that were specific to the first organization) as well as asking them what their general requirements might be for an expert recommender system. The requirements included:
- Searching by age (suggested by DERD for offering age-specific awards).
- Searching for expertise in general as well as expertise in a particular field (suggested by DERD for offering non-domain-specific awards).
- Details of well-established and/or recognised experts, such as professors, and lesser established, eg Masters and PhD student, experts were needed, as well as non-academics performing relevant research. (Suggested by DERD as they have found that people studying towards a higher degree were more likely to be interested in applying for an award).
- Details about an expert's teaching awards, publications, and grants, as well as information about the type of teaching skills they possess (e.g. if they have experience in student centered learning, or team teaching). (Suggested by TL as a method of judging the quality of the expert's teaching skills and finding out, for instance, if they would be suitable to give guest lectures or mentor new staff members).
- A model showing the relationships between experts (Suggested by RO to show joint grants and papers).
- Certain flags to restrict searches on grant applications that are confidential (suggested by RO).
- Experts who are registered in the system should be allowed to nominate availability (suggested by MPR as the response time of an expert was an issue).

4. Approach

Fig.3. Our Triangulated approach
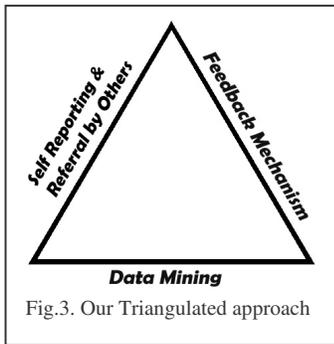
As depicted in Fig. 3, our approach uses data mining as a foundation from which our initial data is captured and against which our data is regularly crosschecked. The key inputs to data mining include individual web pages, project/grant repositories, citation indexes (e.g. CiteSeer (http://citeseer.ist.psu.edu/) and publications databases. Within our university we will use the Integrated Research Information System (IRIS) as one of our data sources. IRIS consists of a collection of all publications and impact factors of individuals within the university. The latter is currently our most useful resource as it is the most structured.

The output of the data mining step is provided to each person included in the expertise database. In the university we may choose to include only academics; academics plus administration staff; or only "research-active" academics (this definition will vary across institutions) in the data mining process. Using data mining we can attach dates related to the expertise found to assist with currency. We can also keep statistics on the level of expertise using simple measures such as the number of publications in that area and term frequency inverse document frequency (TFIDF).

Referring back to Fig. 3, a supporting dimension is the use of the self-reporting technique to expertise identification. Within this dimension we allow for others, such as a PhD supervisor, to nominate/refer another person, such as their student, hence the label "Self Referral". However, this technique is not used in isolation but builds upon the data mining foundation. To address the issues of external validation, expertise currency and motivation to enter and maintain the data, individuals are sent the output from data mining at regular intervals, say twice yearly. Given the importance of reputation and track record in the university system we believe academics would be motivated to check what the system, as it reflects the data in the world about them, says and to correct any errors or omissions.

If the system only used self-reporting, some will choose to have no profile through this system, but few would choose to have an inaccurate picture of themselves portrayed to others. It is also much easier to review a portfolio which has been automatically generated for you than have to format one for yourself. In addition to sharing the data validation and maintenance with the human, the system provides a standard way of describing its experts which facilitates searching.

Providing an automated means of determining who is an expert in what and to what extent together with a means for involving the human expert in validation and augmentation of the profile acquired through data mining is still not enough. From discussions, interviews and personal experience the key question to be asked of the advice of any recommendation system is "was it useful?" As shown in Fig. 3, the third support to our approach is a feedback mechanism.
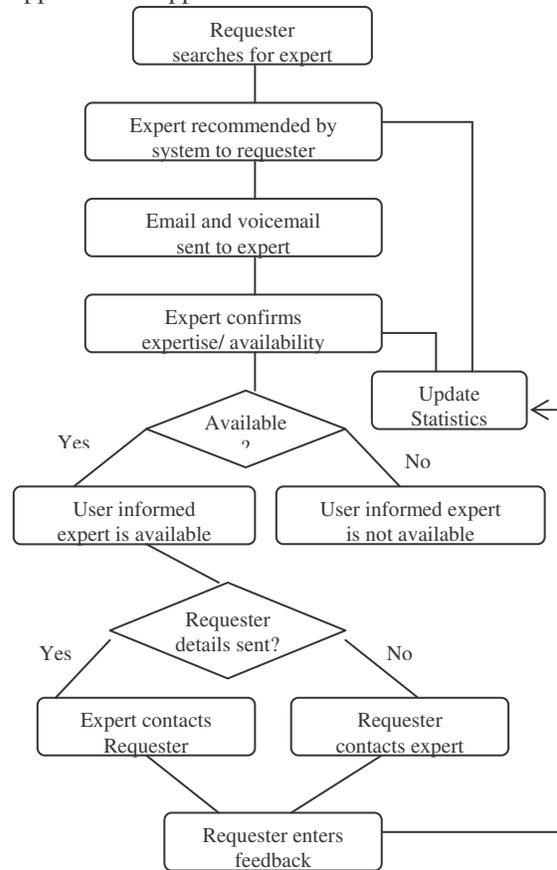

Fig 4. Search flow between Requester, Expert and System

Feedback will be gathered from both sides, hence the notion of two-way communication. Firstly, as depicted in the activity diagram in Fig. 4, when a recommendation is made an email and voicemail message will be sent to the individual/s suggested by the system as experts. The identified expert/s will have

the opportunity to respond to the person requesting their services regarding their availability and willingness to be contacted. The system could be configured in different ways regarding the exchange of contact details, (see the checkbox at the bottom of Fig 6). When selected, the search for an expert results in the query being sent to the experts identified by the system together with the details of the service requester. The expert can select a link in the email to confirm that they are indeed able to offer expertise in the domain/problem area specified and their availability. If the service requester has sent their details, the expert can contact the requester or wait to hear from them as they would do if no requester details were given.

The screens which make up the user interface play a critical part of the approach by allowing the development of a profile, or user model, of the expert as well as the service requester. The user models will provide knowledge to the system that will be used as a filter for searching and matching. The need for customizing both sides became apparent from interviews revealing that the (type of) person being sought varied for different departments. As shown in Fig. 5, it was necessary to capture the needs and preferences of the person looking for someone (e.g., they are a news journalist and need someone to provide an expert opinion for television in layman's terms in the next hour) and also the preferences communication

Fig 5. A simplified validation screen sent to the expert as a result of data mining from webpages, publication/ citation databases, etc.

skills and availability of the expert (e.g. they have given radio interviews in the past but are away for two weeks). The user models can potentially be populated from data found in other places, such as from webpages maintained by the individual or corporate pages, but realise that the feasibility of this depends on the level of webpage standardization across the organization and the degree to which content management is enforced. Thus in our solution we leave this as potential future work. As part of the user model are statistics relating to the responses from experts regarding their expertise and availability and the requesters satisfaction with the service provided. This service has many aspects to it, some of which, such as ability to communicate and availability, can be used to rate the expert and order future results to searches. Recommended experts who frequently receive negative feedback (or poor ratings if a rating system similar to that used by [1] is used) will be recommended less (or perhaps ranked lower in the final output) by the system than experts who have consistently received positive feedback, for example. This feedback mechanism will act as a referral system, suggested by some of the people we spoke with as the type of system they would be interested in, rather than a yellow-pages model. They didn't want contact details of anyone professing to know a particular domain, they wanted to find someone based on the experience of someone else, who they preferably knew and trusted.

While the system will use the statistics for generating and ordering recommendations in response to a query, for privacy and ethical reasons, only the expert will be able to see his/her own statistics regarding their overall usefulness and availability as perceived by the service requesters. We see this personal feedback as potentially useful for professional development and self evaluation. Similar to the way in which data from student evaluations of a teacher may be used for professional development and dis-closed at the experts discretion for promotion or other reasons. As represented, this add-itional feed-back/ valid-ation pillar provides a triangulated approach bringing together auto-mated machine-based knowledge discovery and manual validation.

Fig. 6. Simplified Search screen to find an expert

5. Conclusions

Rating systems, such as we propose and that used by [1], raise several ethical issues. For instance many people may object to the concept of rating another person and may refuse to participate. On the other end of the scale, some users may give recommended experts an unnecessarily bad rating simply because they don't like them on a personal level. In addition, the possibility that a person's personal or teaching skills could be criticised would be a sensitive issue for many and may result in a large number of experts refusing to be registered in the system. In our approach we aim to try different methods of user feedback as well as limiting the visibility of an expert's feedback results and preserving the anonymity of the users who provide the feedback in an attempt to avoid the ethical issues. In addition, we will also consider the use of more personalised feedback (e.g. a reporter wishing to interview an expert will be rating them on different criteria than someone wishing to work with the expert on a project).

As a key part of our approach is the combination of self-reporting/referral and data mining. Some data can only be obtained via self-reporting (e.g. indicating if you are available to do media interviews or guest lectures). However, information about which units one teaches, expertise areas, grants, etc. can be gained from personal websites and internal databases. An issue we are yet to fully resolve would be how to reconcile differences between these sources and between the outputs of data mining and self-reporting.

Recommender systems greatly speed up and simplify the searching process, whether the item being searched for is a book, film, or another person. This project is interested in a recommender system which maintains user profiles to match experts with service requesters. Validation of the user profiles and the system recommendations using a combination of automated and human-based techniques seeks to combine and reinforce these two main approaches which individually have numerous weaknesses. Closing the loop between the seeker and the sought is aimed at providing both parties with confidence and motivation to use the system. We anticipate that our findings will be of use to other recommender systems and search engines, such as Google, in general. Finally, to provide a generalized framework for expertise location, we will consider what modifications are necessary to allow other knowledge-intensive organizations to use the framework and toolkit.

6. References

[1] Aïmeur, E., F. Onana, F. S. Mani and A. Saleman, Anita, "HELP: A Recommender System to Locate Expertise in Organizational Memories" *AICCSA '07*. pp. 866-874.

[2] Becerra-Fernandez, I. "The Role of Artificial Intelligent Technologies in the Implementation of People-Finder Knowledge Management Systems" *Knowledge-Based Systems,* 13: 2000, pp. 315-320.

[3] Chung, R., P. Bateman, K. Cho, "Expertise gaps and profiles: An integrated view of expertise on knowledge transfer" *Proc. of IS-CORE 2005 W/shop of Int.l Conf. of IS* Dec 11-14 Las Vegas Nevada, 2005.

[4] Crowder, R., G. Hughes and W. Hall, "An agent based approach to finding expertise" in *Proc. 4th Int.l Conf. on Practical Aspects of Knowledge Mgt* Heidelberg Germany , 2002, pp. 179-188.

[5] Foner, L. "Yenta: A Multi-Agent Referral-Based Matchmaking System" in *Proc. 1st Int.l Conf. on Autonomous Agents* Marina del Rey California, 2002, pp. 301-307.

[6] Kautz, H., B. Selman, and M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering" *Communications of ACM,* 40(3), 1997, pp. 63-65.

[7] Krulwich, B. and C. Burkey, "The ContactFinder Agent: Answering Bulletin Board Questions with Referrals" *Proc. 13th Nat. Conf. on AI & 8th Innovative Appl. of AI Conf.* Vol 1, Portland, Oregon, 1996, pp. 10-15.

[8] Lave, J. and E. Wenger, *Situated Learning: Legitimate Peripheral Participation,* Cambridge University Press, Cambridge, U.K, 1991.

[9] McDonald, D., and M. Ackerman, "Expertise Recommender: A Flexible Recommendation System and Architecture" *Proc. ACM2000 CSCW Conf.,* Philadelphia PA, 2000, pp. 231-240.

[10] Pikrakis, A., T. Bitsikas, S. Sfakianakis, M. Hatzopoulos, D. DeRoure, S. Reich, G. Hill, and M. Stairmand, "MEMOIR – Software Agents for Finding Similar Users by Trails" in *Proc. 3rd Intl.Conf. on Practical Application of Intelligent Agents and Multi-agents*, London, UK, 1998, pp. 453-466.

[11] Scott, J., *Social Network Analysis: A handbook*, Sage Publications, London, U.K, 1991

[12] Sim, Y., R. Crowder and G. Wills, "Expert Finding by Capturing Organizational Knowledge from Legacy Documents" *Proc. IEEE ICCCE '06*, Kuala Lumpur, Malaysia, 2006.

[13] Snowden, D., "Narrative patterns: Uses of story in the third age of knowledge management" *Journal of Info. & Knowledge Mgt,* 00, 2000, pp. 1-5.

[14] Streeter, L., and K. Lochbaum, "An Expert/Expert Locating System Based on Automatic Representation of Semantic Structure" *Proc. 4th IEEE Conf. on AI Appl. Comp. Soc. of IEEE*, San Diego CA, 1998, pp. 345-349.

[15] Sveiby, K., *The New Organizational Wealth: Managing & Measuring Knowledge-Based Assets*, Berrett–Koehler, San Fran. USA, 1997.

[16] Vivacqua, A., "Agents for Expertise Location" in *Proc. AAAI Spring Symp. on Intelligent Agents in Cyberspace,* Stanford, CA, 1998, pp. 9-13.

[17] Wernerfelt, B., "The Resource-Based View of the Firm", *Strategic Management Journal*; 5, (2), 1984, pp. 171-180.

[18] Yimam-Seid, D. and A. Kobsa, "Expert Finding Systems for Organizations: Problems and Domain Analysis and the DEMOIR approach" *Jrnl Org.l Comp. & Electronic Commerc*e 13, 2003, pp. 1-24.