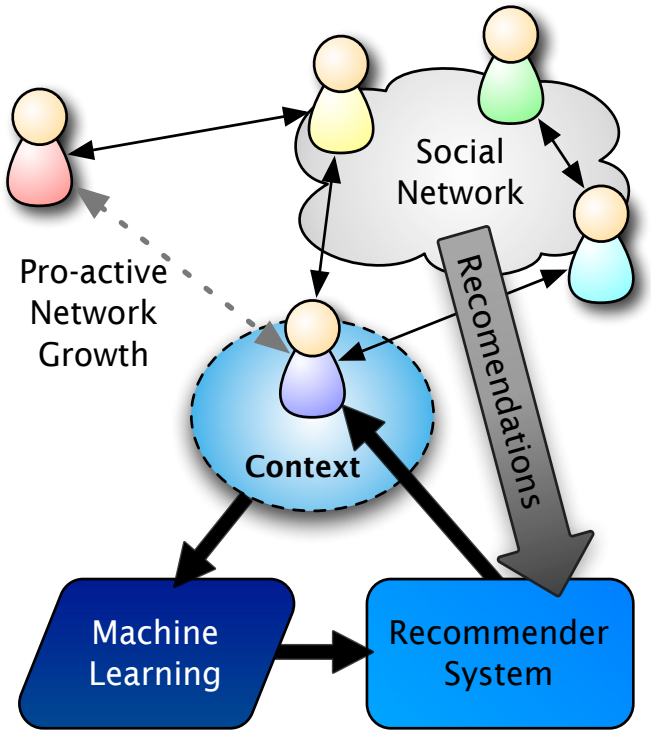
Virtual Centre of Excellence in
Mobile and Personal Communications

Mobile VCE

Instant Knowledge: Profiling Tool Brief

IK Concept

Instant Knowledge enhances the value of any organisation's most important asset—the information held by its employees. Rather than requiring staff to fill out skills profiles, which are very general, become outdated, and require significant effort, IK uses an application on employees' smart phones and laptops to gather information on what they are doing and who they are communicating with. This context is used to build dynamic skills profiles along with a social network map for the enterprise, which provides a resource to proactively offer recommendations to participants. Using IK, staff can always find the best person for the job.



IK Concept: It's not what you know, it's who you know, and who they know...

Profiling Tool Concept

An invaluable asset for any organisation is a database containing profiles of its staff detailing their current expertise. Yet acquiring such a database is notoriously hard. Trying to induce busy employees to fill in their own expertise has always proved difficult, not least because it is not always clear what skills should go into such a database, and how such a database can be kept up to date.

One apparently simple technological solution is to extract profiles automatically from the documents of staff held on their computers. These documents might include the documents and memos they write, their emails, and the documents they are sent by others. In combination with state-of-the-art recommendation tools this can quickly and accurately identify the expert in an organisation. However, this raises series concerns about privacy which no organisation can afford to ignore. Automatic profiling could potentially reveal passwords, bank details, private information, confidential memos, or just friendships which individuals prefer to keep private. Any automatic profiler thus faces a dilemma between revealing information useful for finding an expert and hiding information that could be damaging to user's privacy. There is no solution to this problem which is guaranteed to never leak sensitive information, but modern machine learning techniques can certainly go a long way in addressing the openness vs privacy dilemma outlined above.

Novelty & Contribution

In IK we have produced novel algorithms which allows accurate recommendations to be made at the same time as protecting as much information as possible.

Industrial Chair: Glyn Jones, Thales UK Research and Technology **glyn.jones@thalesgroup.com**
Academic Coördinator: James Irvine, University of Strathclyde **j.m.irvine@strath.ac.uk**

The algorithms use a corpus of published documents from an organisation to define the regions of legitimate interests of that enterprise. This is used to create a low dimensional subspace of words into which all profiles are projected. This can be done locally on a users machine so that no private information needs to be sent to another machine. This reduces profiles to a set of 100 numbers, from which private information such as passwords will be removed. Comprehensive tests show that despite removing significant amounts of information the profiles still provide high quality recommendations. In fact, the recommendations are improved, due to the removal of information which is not relevant to the expertise of the employees.

However, "problems worthy of attack prove their worth by hitting back", and protecting privacy certainly qualifies as a problem worthy of attack. Despite the success in keeping private individual words, tests show that just performing a projection as described above does not hide a users interests outside of the core business of an organisation. This is a consequence of the extraordinary power of modern recommender systems to piece together evidence to make a recommendation. For example, if users had documents covering their musical interests, then any query to a recommendation system on say Country and Western would pick out Country and Western devotees with high accuracy. This seems strange at first since a huge number of Country and Western terms will be filtered out. Nevertheless, because we use the same words in different contexts a particular constellation of words can identify the private passions of users with a high probability.

To provide a higher level of privacy we have developed the profiling algorithm further to identify documents that lie outside the core interests of an enterprise and to disregard them in producing a profile. This can again improve the recommendation quality for skills relevant to the organisation, while improving the privacy for users of the system.

Application Scenarios

This research was carried out as a core part of the IK project to deliver a recommendation system for finding the people with a particular expertise within an organisation. Critical to the success of such a scheme is to collect profiles for all employees and to keep it up to date. Security and privacy was a major concern of the project.

In the facebook world we live in, issues of exchanging information at the same time as protecting privacy are ubiquitous. One application beyond IK, might be the automatic creation of profiles for a professional social network similar to LinkedIn. Rather than having to fill in your own expertise a tool could be developed which would automatically produce a profile based on the documents held on a user's computer. Not only would this save users time and allow profiles to be kept up to date, it could substantially improve the identification of people with the right expertise.

Demonstration Results

The algorithms developed for this project have been thoroughly tested on a dataset consisting of profile information of over 800 users. To test privacy we added additional "private" information to the profiles to see whether we could identify that information. Our study shows that a word-based filter alone, although capable of hiding information such as passwords or bank account details, is nevertheless insufficient to hide private interests of individuals. To ensure a level of privacy which we believe most users would find acceptable requires a multi-level system where in addition to filtering at the level of words we also need to filters at the level of documents to exclude documents that are outside the interests of an organisation.

Conclusions

Automatically generating profile information is highly challenging if an appropriate level of privacy is to be achieved. This is an area that has appeared to be neglected in the research literature. We have developed algorithms that, for the first time, addresses the problem of protecting privacy at the same time as providing accurate recommendations of people with the right expertise.

Further Information

Videos and Technical Reports for all of the Instant Knowledge research outcomes are available to members on the Mobile VCE web site. For non-members the Instant Knowledge overview sheet is available at:

www.mobilevce.com/infosheets/InstantKnowledge.pdf

For further information and to register for information about future MVCE IK events please email Jerry Horton: **jerry.horton@mobilevce.com**

Industrial Chair: Glyn Jones, Thales UK Research and Technology **glyn.jones@thalesgroup.com**
Academic Coördinator: James Irvine, University of Strathclyde **j.m.irvine@strath.ac.uk**