Digging out implicit semantics from user interaction

Stephane Marchand-Maillet

Viper group – University of Geneva
http://viper.unige.ch
marchand@cui.unige.ch

ABSTRACT

User interaction may take many forms in multimedia systems. Current systems mainly waste this implicit and natural source of semantic knowledge and rather create tedious and unnatural interaction protocols. We advocate for a complete integration of natural interaction protocols and semantic knowledge capture, mainly thru mining interaction sessions. We assert that users possess the ability to quickly examine and summarise these documents, even subconsciously. Examples include specifying relevance between a query and results, rating preferences in film databases, purchasing items from online retailers, and even simply browsing web sites. Data from these interactions, captured and stored in log files, can be interpreted to have semantic meaning, which proves indispensable when used in a collaborative setting where users share similar preferences and goals.

1. STATE-OF-THE-ART

The growth of the internet and the technology explosion have contributed to a high demand for new methods of information filtering and retrieval. The amount of multimedia content created daily is accelerating to a point where management, annotation and retrieval are becoming problematic. The one problem that these operations, based on the sole content, are facing is the semantic gap. This often cited term [20, 5, 7, 2, 9] describes the gap between the low-level information computers can describe (or summarize) and the high-level concepts which humans can perceive. What is it that tells us that an image of a man wearing a ski mask running from a bank most probably signifies a robbery? Certainly, algorithms can be trained to recognize this specific imagery, but this is only one of endless possibilities of high-level semantic information extracted from a document. There would need to exist a recognizer for every possible scenario. One can make the humorous reference to hiring one doctoral student to build a recognizer for every concept [15], but the reality is that without some mapping from syntax to semantics, this is (and has been) the most followed path.

A more recent way to alleviate the problems caused by the semantic gap is to use crowdsourcing as a way to supplement the missing semantic information. Crowdsourcing is a moniker which describes the outsourcing of a problem or task to a large number of users in an attempt at finding a solution [6]. Although the definition lends itself to an explicit arrangement of the distribution of problem workload, it also covers implicitly sourced user-power, where tasks are less defined and subtle. Such inferred semantic relationships may take many forms, but the more popular can be categorized as: browsing logs, where users casually peruse document collection with no formal information need; click-through data, where information is sought but evidence of interest (the "click") does not necessarily imply relevance [21, 3] and relevance feedback judgments, where the user has a definite query and explicitly rates search results with respect to relevance in order to refine that query [10, 8]. Manually providing semantic data is also common, such as manual shallow tagging of documents with keywords. However, here we focus on the less intrusive ways of collecting user interaction, as it is important for the users to feel free to operate naturally, without spending time consciously annotating and entering data.

When facing the semantic gap, systems have much to gain from user involvement when the data can be collected effectively. The relatively new social web and the growth of user-driven content and collaboration is fuelling new research on the efficient use of user interaction data. For some time now, it has been shown that user browsing habits and general web trends can be extracted from web server logs [22, 11, 1]. These methods are now being applied to collections of multimedia documents lacking semantic metadata, web search result click-through [3], and virtually any area where user interaction can be harvested. A number of novel methods of extracting explicit user interaction data have been documented recently in the literature. For example, Ahn *et al.* [16] procured incentive to label images on the internet using a game-based approach, which paired random users together to find agreement on semantic labels. A highest score list secured bragging rights as the motivation. Other similarly spawned crowdsourcing games from the same scientists include Peekaboom (object recognition) [18] and Phetch (image retrieval) [17]. Likewise, in the LabelMe project [13], incentive for researchers to use a region-level labelled image database encourages participation in the labelling process itself. While explicit user involvement such as these can potentially yield higher quality semantic data, the drawback is

Dagstuhl Seminar Proceedings 08251
Contextual and Social Media Understanding and Usage
http://drops.dagstuhl.de/opus/volltexte/2009/2020

that the incentive must pass a certain threshold of enough users to make the collection technique worthwhile. Often such systems garner a large following due to word-of-mouth in the social medium of the internet, only to experience a drop in popularity after the novelty has worn off.

2. WAY OUT

This problem motivates research in collecting implicit or functional-based user involvement data. This approach puts the generation of interaction data at a level where the users do not feel like they are performing a task because the user interaction is engineered to be subtle. Examples include the afore-mentioned long-term learning of relevance feedback or click-through data in information retrieval systems. Here the user is performing a functional task such as searching for a document; a side effect is that the user may select relevant documents or simply specify which examples are relevant to further refine the query. Associations can then be assumed on the keywords used in the query and the "relevant" documents examined and furthermore semantic similarities between the documents examined in the context of the search original query. Similarly, if a user is browsing a document database, a browsing strategy may afford a certain amount of semantic data, such as browsing a random sampling by concept (e.g. lizards of South America). The exact semantic labeling of such data may be more difficult, but it can also be useful to view this data simply as a graph of semantic similarity on the document database where the nodes denote documents and weighted edges semantic similarity.

As can be shown, both types of user involvement (collaborative filtering and long-term learning over relevance feedback) share fundamental problems. For large data sets, the interaction data is highly sparse, with usually more than 95% missing values [14, 4, 19, 12]. Large data sets also introduce a problem of efficient computation. In collaborative filtering, predictions must be made in real time, so complex models and similarity indices must be calculated ahead of time, which normally means some degree of off-line processing. The same is true for long-term learning in information retrieval. To keep the system current, the underlying models must be built or updated after query sessions, which also requires off-line processing to keep retrieval latency low. Both domains also share the principle of data propagation. From this perspective, we have a sparse dataset and we want to make predictions for some missing values. A type of propagation must occur, where missing values are approximated based on existing values. In information retrieval, the propagation is the relevance of images with respect to query concepts, and in collaborative filtering, the propagation that occurs is the prediction of user ratings for specific items. From the propagation perspective, our problem is well-tailored to be resolved on architecture such as that put forward by community-based systems. Jointly considering networks of different types for sharing and propagating resources and knowledge is key to successfully obtained semantic information management at large-scale.

A "Data Network" is created by federating the data around notion of similarity. Bootstrapping that network may simply be done by content-based similarity analysis. Here, the "Physical Network" (eg the Internet or a P2P network), offering distributed computing helps in resolving scalability issues. From that base data network, the challenge is to enrich it by implicitly or explicitly collected data, as discussed above. Again, the presence of a "Physical Network", support of a user community will help in providing resources (eg for crowdsourcing). This "User Network" (ie users connected by some affinity) will act on the Data Network and implicitly enrich it provided sufficient incentives are provided, as discussed above.

3. REFERENCES

[1] Pierre Baldi, Paolo Frasconi, and Padhraic Smyth, *Modeling the internet and the web: Probabilistic methods and algorithms*, John Wiley & Sons, West Sussex, England, 2003.

[2] M. Cord and P. H. Gosselin, *Image retrieval using long-term semantic learning*, IEEE International Conference on Image Processing, 2006.

[3] Nick Craswell and Martin Szummers, *Random walks on the click graph*, In Proceedings of SIGIR 2007, 2007.

[4] Mukund Deshpande and George Karypis, *Item-based top-N recommendation algorithms*, ACM Trans. Inf. Syst. **22** (2004), no. 1, 143–177.

[5] J. Fournier and M. Cord, *Long-term similarity learning in content-based image retrieval*, 2002.

[6] Jeff Howe, *The rise of crowdsourcing*, Wired Magazine **14** (2006), no. 06.

[7] Takeo Kanade and Shingo Uchihashi, *User-powered "content-free" approach to image retrieval*, Proceedings of International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society 2004 (DLKC04), 2004, pp. 24–32.

[8] S. Marchand-Maillet and E. Bruno, *Exploiting user interaction for semantic content-based image retrieval*, Tech. report, Computer Vision and Multimedia Laboratory, Computing Centre, University of Geneva, 2003.

[9] Donn Morrison, Stéphane Marchand-Maillet, and Eric Bruno, *Semantic clustering of images using patterns of relevance feedback*, Proceedings of the 6th International Workshop on Content-based Multimedia Indexing (London, UK), June 18-20th 2008, (to appear).

[10] Henning Müller, Wolfgang Müller, David McG. Squire, Stéphane Marchand-Maillet, and Thierry Pun, *Long-term learning from user behavior in content-based image retrieval*, Tech. report, Université de Genève, 2000.

[11] O. Nasraoui, C. Cardona, C. Rojas, and F. Gonzalez, *Mining evolving user profiles in noisy web clickstream data with a scalable immune system clustering algorithm*, 2003.

[12] Netflix, *The Netflix Prize*, Web site: `http://www.netflixprize.com/`, 2006.

[13] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman, *LabelMe: A database and web-based tool for image annotation*, Int. J. Comput. Vision **77** (2008), no. 1-3, 157–173.

[14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, *Application of dimensionality reduction in recommender systems–a case study*, 2000.

[15] Cees Snoek, *Content-based video indexing*, Presentation given at Summer School on Multimedia Semantics (SSMS'07), Glasgow, UK, 2007, Slides URL:

[16] Luis von Ahn and Laura Dabbish, *Labeling images with a computer game*, CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems (New York, NY, USA), ACM Press, 2004, pp. 319–326.

[17] Luis von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum, *Improving accessibility of the web with a computer game*, CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems (New York, NY, USA), ACM, 2006, pp. 79–82.

[18] Luis von Ahn, Ruoran Liu, and Manuel Blum, *Peekaboom: a game for locating objects in images*, CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems (New York, NY, USA), ACM, 2006, pp. 55–64.

[19] Jun Wang, Arjen P. de Vries, and Marcel J.T. Reinders, *A user-item relevance model for log-based collaborative filtering*, Proc. of European Conference on Information Retrieval (ECIR 2006), London, UK, 2006.

[20] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski, and B. Field, *Semi-automatic image annotation*, 2001.

[21] Gui-Rong Xue, Hua-Jun Zeng, Zheng Chen, Yong Yu, Wei-Ying Ma, WenSi Xi, and WeiGuo Fan, *Optimizing web search using web click-through data*, CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management (New York, NY, USA), ACM, 2004, pp. 118–126.

[22] Osmar R. Zaïane, Man Xin, and Jiawei Han, *Discovering web access patterns and trends by applying OLAP and data mining technology on web logs*, Advances in Digital Libraries, 1998, pp. 19–29.