Streaming Virtual Patient Records

Pedro Pereira Rodrigues[1–3] and Ricardo Cruz Correia[2–4]

[1] LIAAD - INESC TEC, Portugal
[2] Faculty of Medicine of the University of Porto, Portugal
[3] CINTESIS - FMUP, University of Porto, Portugal
[4] CRACS - INESC TEC, Portugal

{pprodrigues,rcorreia}@med.up.pt

1 Introduction

Clinical record integration and visualization is one of the most important abilities of modern health information systems (HIS). Its use on clinical encounters plays a relevant role in the efficacy and efficiency of healthcare. However, the amount of data currently being produced, stored and used in hospital settings is stressing information technology infrastructure; integrated HIS of central hospitals may gather millions of clinical documents. One solution is to consider a virtual patient record (VPR), created by integrating all clinical records, which must collect documents from distributed departmental HIS. Our vision is that the problem is better modeled in the distributed streaming data paradigm, since distributed streams of documents are produced in the entire hospital, which are to be collected by the VPR, which has then to process and manage the corresponding centralized stream of documents.

2 Patient records and evidence-based medicine

Evidence-based medicine relies on three information sources: patient records, published evidence and the patient itself [1]. Even though great improvements and developments have been made over the years, on-demand access to clinical information is still inadequate in many settings, leading to less efficiency as a result of a duplication of effort, excess costs and adverse events [2]. Furthermore, a lot of distinct technological solutions coexist to integrate patient data, using different standards and data architectures which may lead to difficulties in further interoperability [3]. Implementing a Virtual Patient Record (VPR) system provides an adequate and cost-effective solution for most clinical information needs, as data is gathered from different HIS distributed across the entire hospital. As more and more patient information is stored, it is very important to efficiently select which one is more likely to be useful and promote it in a scenario where scarcity of resources (screen space, storage space, bandwidth and doctors time) is very real [4]. To assess the likelihood of document access, the system's log file could be used. The log file is where all actions performed by users of information systems are recorded. Intentionally and originally created

and kept for audit purposes, these logs can provide very interesting insights into the information needs of health-care professionals in some particular situations, although most of the times the quality of these logs is not delivering [5].

3 A proposal for a streaming virtual patient record

The current setting is a central hospital that has several departmental information systems (DIS) that produce clinical documents (e.g. radiology documents and lab results) that might be relevant for the practice of healthcare. The access to this documents is better achieved by a centralized information system that integrates all the different departmental systems, aggregating the documents that are most relevant for the current encounter [3]. Between May 2003 and May 2004, a virtual patient record (VPR) was designed and implemented at Hospital S. João, a university hospital with over 1350 beds. An agent-based platform, Multi-Agent System for Integration of Data (MAID), ensures the communication among various hospital information systems [6]. Clinical documents are retrieved from clinical department information systems (DIS) and stored into a central repository in a browser friendly format. MAID is now running for the last 9 years, regularly scanning 14 DIS and collecting about 7000 new documents each day. Currently, over 340 doctors are using the system on a daily basis [7].

The vision we present here is that the problem is better modeled in the distributed streaming data paradigm, defining a Streaming Virtual Patient Record (SVPR), since distributed streams of documents are produced in the entire hospital, which are to be collected by the VPR, which has then to process and manage the corresponding centralized stream of documents.

3.1 Current virtual patient record

Each departmental information system produces a stream of documents, the production rate of which is highly dependent on date and time of production, but also on other factors specific to the department itself (e.g. type of document). This creates a heterogeneous network of distributed data streams.

In the current setting, the central VPR receives an increasing rate of 200+ documents per hour, corresponding to the daily 5300+ patients in the hospital (including inpatients, outpatients and emergency rooms) which need to be processed. The VPR tracks information on visualizations performed by 4850+ users. Older documents tend to be less frequently visualized in encounters but, for example, nearly half of the visualizations in 2010 and 2011 targeted documents older than January 2009, so old documents cannot be discarded. During an encounter with a patient, and considering that only documents from that patient are relevant (which is most of the times true) the pool of documents that one of the currently active 2375 users might access is 537.9 (average number of active documents per active patient). Clearly, the user cannot be presented with a list of 530+ documents, so a ranking is needed to prioritize the most relevant.

3.2 Requirements for a streaming approach

The virtual patient record already uses mobile agents to asynchronously collect documents from multiple departmental information systems, creating a single stream of collected documents. Parallel to that stream, we need to track the visualization of the documents by the user, in order to adapt the system to the probability of visualization of each document. Hence, there are two main data streams being managed: a stream of documents being produced; each element is a new document (or a new version of an existing one); and a stream of visualization events; each element is an event of visualization of a previously produced document. The proposed streaming VPR should be able to perform the following tasks:

1. collecting documents produced in distributed departmental HIS;
2. updating a previously collected document with a new version produced by the original HIS;
3. managing the stream of documents being collected by the streaming VPR;
4. keeping track of documents visualization events;
5. ranking available documents to present to the user of the streaming VPR;
6. deciding which documents should be kept in primary storage and which should be sent to secondary storage.

Such a system should deal with the high-rate of data production in terms of efficacy (the system is able to gather clinical documents and displayed them to requesting users) and efficiency (requested documents are readily available at time of request and preferably not stored in secondary storage devices).

4 Current state and proposed stream model

Tasks 1 and 2 are already performed by the current VPR [6] and being adjusted [8]. Task 4 is implemented using audit track logs and has been studied in batch [4, 7]. Task 5 is now being tested, using simple statistical methods to infer visualization probability [9]. Task 6 is not yet implemented. For task 3 (and corresponding adaptation of all other tasks), we need to define the most appropriate stream model. The first stream is gathered by integrating documents from heterogeneous information systems, which should be modeled according to the *insert-delete* or *turnstile* model [10], allowing that observations might be updated or deleted by future events (some clicnical documents are subject of validation and revision, deactivating the previous versions of that document). The second one is controlled in the VPR but it only tracks information on visualizations, so it should be modeled according to the *insert-only* or *time series* model [10], since the event of visualization is not subject of deletion. However, given the predictive task of our system, we could consider the *accumulative* or *cash-register* model [10], where each observation is an increment to a given sum, i.e. the counters of the current number of visualizations of each document.

Addressing the problem from the data stream approach will increase the ability to process and store clinical documents, as well as improve the usability

of the virtual patient record. Furthermore, the track logs can give insights into the information needs of healthcare professionals in a particular situation. The study of these logs, using streaming machine learning algorithms, should allow us not only to describe how the systems were used, but may also be useful to predict future use of the system and of the data items it contains.

Acknowledgments. This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project "FCOMP - 01-0124-FEDER-022701".

References

1. Wyatt, J.C., Wright, P.: Design should help use of patients' data. Lancet **352**(9137) (October 1998) 1375–8
2. Feied, C.F., Handler, J.a., Smith, M.S., Gillam, M., Kanhouwa, M., Rothenhaus, T., Conover, K., Shannon, T.: Clinical information systems: instant ubiquitous clinical data for error reduction and improved clinical outcomes. Academic emergency medicine **11**(11) (November 2004) 1162–9
3. Cruz-Correia, R.J., Vieira-Marques, P.M., Ferreira, A.M., Almeida, F.C., Wyatt, J.C., Costa-Pereira, A.M.: Reviewing the integration of patient data: how systems are evolving in practice to meet patient needs. BMC medical informatics and decision making **7** (January 2007) 14
4. Cruz-Correia, R.J., Wyatt, J.C., Dinis-Ribeiro, M., Costa-Pereira, A.: Determinants of frequency and longevity of hospital encounters' data use. BMC medical informatics and decision making **10** (January 2010) 15
5. Cruz-Correia, R., Boldt, I., Lapão, L., Santos-Pereira, C., Rodrigues, P.P., Ferreira, A.M., Freitas, A.: Analysis of the quality of hospital information systems audit trails. BMC Medical Informatics and Decision Making **13** (2013) 84
6. Vieira-Marques, P.M., Cruz-Correia, R.J.a., Robles, S., Cucurull, J., Navarro, G., Marti, R.: Secure integration of distributed medical data using mobile agents. Intelligent Systems **21**(6) (2006) 47–54
7. Rodrigues, P.P., Dias, C.C., Rocha, D., Boldt, I., Teixeira-Pinto, A., Cruz-Correia, R.: Predicting visualization of hospital clinical reports using survival analysis of access logs from a virtual patient record. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal (2013) 461–464
8. Patriarca-Almeida, J.H., Santos, B., Cruz-Correia, R.: Using a Clinical Document Importance Estimator to Optimize an Agent-Based Clinical Report Retrieval System. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. (2013) 469–472
9. Santos, B., Rodrigues, P., Cruz-Correia, R.: An automatic clinical document importance estimator for an existing electronic patient record - architecture and implementation. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal (2013) 537–539
10. Muthukrishnan, S.: Data Streams: Algorithms and Applications. Now Publishers Inc, New York, NY (2005)