Task-based Assessment of Automatic Metadata Extraction

Alberto Messina, Fulvio Negro
RAI, Radiotelevisione Italiana, Centre for Research and Technological Innovation, Italy
Turin, Italy
{a.messina,f.negro}@rai.it

Werner Bailer
JOANNEUM RESEARCH, DIGITAL – Institute for Information and Communication Technologies
Graz, Austria
werner.bailer@joanneum.at

1. INTRODUCTION

Adding interactivity to TV and video is often hindered by the lack of appropriate metadata about the content. This metadata is needed in order to provide appropriate links within the content and to other items, or to enrich the audiovisual content. However, producing such metadata comes at a high cost. It is an acquired concept that media production processes can benefit from the use of automatic information extraction tools, which analyze multimedia content and provide information for content description, indexing and search. This work is also motivated by the EBU MIM/SCAIE[1] working group, established 2007, which aims at bringing automatic information extraction tools into media production processes. However, these automatic tools do not produce perfect results, and manual correction might be required to achieve a certain target quality of the produced metadata. It is difficult to assess the impact of a specific information extraction tool (e.g., genre classification) on the overall process in terms of quality improvements or cost savings w.r.t. manual processes. In order to address this issue, we propose to look at assessment of automatic information extraction methods from a novel perspective. Rather than evaluating these tools in an isolated lab setting, the tools are assessed in the context of a specific task of a real media production workflow.

In user interface (UI) design, it was proposed already more than 25 years ago in [1] to follow a task-based approach. Thus, a large share of the literature on formal task models comes from the area of model-based user interfaces (for a historical overview see [2]). In this paper, we first discuss the notion of tasks, and the formalization we apply in order to achieve a machine processable representation. We then describe first results of the application of the formal task models for assessing the total cost of performing a task manually or with automatic tools with different performance levels, and describe the simulation demonstrating the proposed approach.

2. FORMAL TASK MODELS OF MEDIA PRODUCTION PROCESSES

In this paper, the term *task* denotes a sequence of actions performed by one or more users to achieve a defined goal in the production process, usually using a set of tools. The task has a defined set of input items and produces a set of output items. For example, a “Content Search” task could be defined as “The action performed by the employee of a broadcaster to find an audiovisual content item (the output item) with a specified topic (the input item)”. We denote the language to define and describe task models as *task metamodel*. A *task model* is an abstract representation of a task, i.e. an orchestrated set of actions performed by actors in order to reach a specific objective. The objective is expressed in terms of conditions that have to be satisfied by the reference domain. Pre- and post-conditions may pertain to single objects or sets of objects or to the entire domain. The task model is formalized using a task metamodel. The example mentioned above, which is classifiable as task model, is stated using the task metamodel “English language”.

We have reviewed a number of candidate representations for task metamodels, considering those that are adopted, provide a machine readable definition, define a serialisation (preferably XML) and for which there is also tool support. Metamodels that are specific to UI design have been excluded. We have selected ConcurTaskTrees (CTT) [3], which has been proposed as a graphical model for tasks. The basic structure of the model is a tree representing the breakdown of tasks into subtasks. On each level, temporal dependencies between subtasks (e.g., serial or parallel) can be modeled. The model has been extended over time and is probably the most commonly used of all task metamodels. We use an extension called collaborative CTT [4], which allows modeling the cooperative execution of a task by multiple actors (users and systems). A task model covering the role of each actor is defined, and cross-links (e.g., information exchange, interactions) between tasks from different roles are described.

3. BENCHMARKING CONTENT ANALYSIS TOOLS IN TASK CONTEXT

This section outlines how formal task models can be used to benchmark algorithms for automatic information extraction in a system context. The basic observation is that existing benchmarks of individual components do not always reflect the applicability of the methods for a certain task in a real process. For example, if we take the benchmark data for a person identification software based on recognition of the speaker voice, typically we get figures related to reference data sets used by the research community to compare results, and we do not have any indication on how errors (e.g., false detections) impact in real usages of the technologies. Here are two other examples, where there are mismatches between the commonly used benchmarking approaches of components and their contribution to solving an actual task. Typically, precision and recall of shot boundaries w.r.t. a ground truth is evaluated. However, missed shot boundaries coinciding with scene boundaries will strongly impact the result, while missed shot boundaries within scenes and several false positives might be tolerable. For person identification, there might be cues in multiple modalities (e.g. face, text insert, name mentioned) that contribute to successful identification. Depending on the structure of the data set the overall result varies with the performance of components working on the different modalities. For example, if persons are identified by text inserts, and video OCR performs very well, this might mask missed detections of a face detector.

[1] http://tech.ebu.ch/groups/pscaie

4. ASSESSING COST-EFFECTIVENESS

4.1 Approach

Real media production processes are a complex combination of human factors and system operations, and as such quite distant from the aseptic laboratory settings in which automatic tools are developed. Furthermore, workflows are the result of established practices that involve not only practical technical constraints but also personnel-related issues like shifts, contractual regulations and professional roles. As a result, costs connected with workflows cannot be estimated taking into account individual operations, but considering the whole process. As a consequence, expected workflow optimizations introduced by the introduction of automated tools cannot easily be assessed.

The proposed approach is then to simulate the entire process under consideration introducing two distinct modalities for the specific function being optimized, fully manual operations and computer-assisted operations, and then perform an analysis on the minimum performance needed by the automatic tools to improve the overall cost figures of the process. In general, the condition to be met can be expressed as:

1. $\sum_{i=0}^{N}\left(C_i^A + C_i^{chk}\right) < \sum_{i=0}^{N} C_i^M$

where N is the number of functions in the workflow in which some automatic tool is introduced, C_i^A is the cost connected with the individual execution of the automatic tool implementing function *i*, C_i^{chk} is the cost connected with the manual check of the automatic tool's results, and $C_i^M$is the cost connected with the fully manual implementation of function *i*. Costs can have a manifold nature and depend both on personnel costs and on systemic costs, and these may vary from function to function in the workflow. This means that analytical estimation of Eq. 1 in real cases can be very complex and expensive. Thus, to practically evaluate such trade-off condition we consider the whole workflow as the function W to be evaluated and estimate total costs connected with each of the two workflow versions (automatic + check and fully manual):

2. $C_W^A + C_W^{chk} < C_W^M$

Developing Eq. 2 distinguishing systemic and personnel costs we obtain:

3. $C_{Ws}^A + C_{Wp}^{chk} + C_{Ws}^{chk} < C_{Wp}^M + C_{Ws}^M$

where obviously we assume that $C_{Wp}^A = 0$, i.e. that costs of personnel for the execution of the automatic tool are negligible.

4.2 Simulation

We have selected a task model for "identification of persons in news material" we have and performed a simulation analysis under the following conditions. (1) We assume that the function related to manual person identification could be substituted by a combination of automatic analyses: identification by face, by speaker's voice, or by open caption/graphics. (2) We implement the whole workflow in a simulation tool in the two versions (manual and automatic + check). (3) We run the simulation of the manual process. (4) We run a series of simulations of the automatic process, under a range of values for the F-measure of the automatic tools. (5) We compare results and find a tradeoff condition. As shown in Figure 1, The proposed approach can be used to relate the performance of an automatic content analysis tools to the cost saving in a specific process. Based on measured data from actual process, this methodology can help to assess whether the performance of a specific automatic content analysis tool is good enough for a specific practical application.

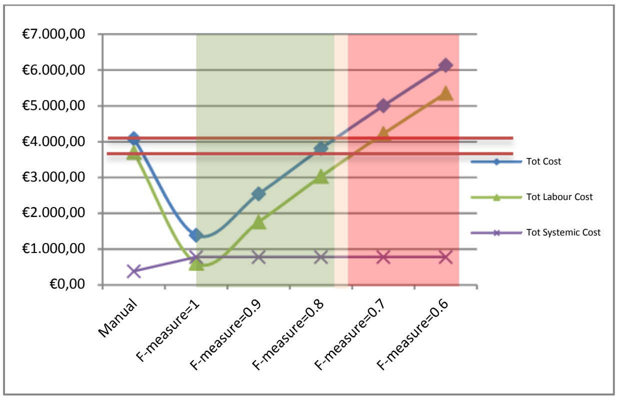

Figure 1. Simulation results.

5. DEMONSTRATION

The simulation described is based on an executable business process model, and can be run with different parameterizations for the automatic tools as well as manual steps included in the process. The results are cost estimation for the different metadata generation processes, which can be use for supporting the decision whether specific metadata generation processes can be cost-effectively automated. Although in this first version the numbers used in the simulation so far were experts' estimates rather than measured data, the results already provided useful insights about how to evaluate the tradeoff between performance and overall costs in specific system contexts. By considering different types of errors separately (e.g., because a false detection can be efficiently deleted, but a missed detection is costly to correct) the simulation can be further refined to a specific task context.

6. CONCLUSION

Generating the metadata necessary for interactive TV services is costly, and can potentially benefit from automation. We propose the use of task models for the assessment of automatic information extraction tools. This approach provides a novel perspective on evaluation of automatic information extraction, and is capable of providing information on practical applicability of tools, beyond lab benchmarks. The approach has already been used for running a cost simulation for comparing manual annotation with automatic annotation at different performance.

7. ACKNOWLEDGMENTS

This work has been partially supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under the grant agreement n° 287532, "TOSCA-MP".

8. REFERENCES

[1] Green, M. "The University of Alberta user interface management system," *Proc. ACM SIGGRAPH*, 1985.

[2] Meixner, G., Paternò, F., and Vanderdonckt, J. "Past, Present, and Future of Model-Based User Interface Development," *i-com*, 2011/10.

[3] Mori, G. and Paternò, F., and Santoro, C. "CTTE: Support for Developing and Analyzing Task Models for Interactive System Design," *IEEE Trans. Software Eng.*, 28(2), 2002.

[4] Paternò, F., Mancini, C., and Meniconi, S. "ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models," *Proc. IFIP TC13 International Conference on Human-Computer Interaction*, 1997.