Generating récit from sensor data: evaluation of a task model for story planning and preliminary experiments with GPS data

Belén A. Baez Miranda **Sybille Caffiau** **Catherine Garbay** **François Portet**
Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
`FirstName.LastName@imag.fr`

Abstract

Automatic story generation is the subject of a growing research effort which has mainly focused on fictional stories. In this paper, we present some preliminary work to generate récits (stories) from sensors data acquired during a ski sortie. In this approach, the story planning is performed using a task model that represents domain knowledge and sequential constraints between ski activities. To test the validity of the task model, a small-scale user evaluation was performed to compare the human perception of récit plans from hand written or automatically generated récits. This evaluation showed no difference in story plan identification adding credence to the eligibility of the task model for representing story plan in NLG. To go a step further, a basic NLG system to generate narrative from activities extracted from GPS data is also reported.

1 Introduction

Stories are a common construct used by humans to share their experience (with physicians, friends, relatives...) by which they tell what happened. In this paper, we focus on human activity stories that we call "activity récits" with the aim of generating these from real ambient data. According to (Adam, 2011), a *récit* is a set of events related to facts that have been effectively experienced, observed or captured. Our problem statement lays onto the narrative structure, the récit plan.

Computational Narratology (CN) is the study of narratives from the point of view of computation and information processing (Mani, 2013). Most of the current researches in CN are related to creativity, where the stories emerge from a set of predefined parameters, trying to imitate literary genres like fairy tales (Riedl and Young, 2010). However, we are interested in stories depicting human activity from real ambient data for which we have no control and little knowledge. In this paper, we focus on a ski touring application. Figure 1 shows an

10.00, awful weather, we went to Chamechaude, a usual destination in case of bad weather. In order to add some more climbing, we start 100 m below the Col de Porte, down the lift. The weather is not beautiful, objectively not very cold but we slip under a fine rain that freezes a bit. We climb quickly and we warm up quickly. Above the rain stopped and I even have the feeling it was too hot in the humid atmosphere! We took only a few breaks, and I do not remember having eaten or drunk anything [...] [Translated from French]

Figure 1: Example narrative and its corresponding raw data captured along a ski touring activity (two persons involved P1 and P2)

application to ski touring where skiers, alone or in groups, use special devices such as GPS (Global Positioning System), heart rate monitor, temperature etc. After their journey, they may share their experience, observations, and other evaluative elements (weather conditions, terrain and key places to visit) on websites such as www.skitour.fr.

The final goal of the research is to be able to generate a coherent and faithful story from the sensor raw data. In this paper, we present work about two research questions (among many others) linked to this final goal:

1. *What kind of model can ensure story representation and coherence? How can we evaluate it?*

2. *Is GPS data sufficient to generate initial stories?*

The first question has been partially studied in (Baez Miranda et al., 2014) where a task model approach was chosen to abstract and structure knowledge about a ski activity. However, this model was not evaluated. This paper thus reports an experiment in which the validity of the task model for récit plan, is evaluated by comparing the perception of the story plan using texts automatically generated from predefined task model instances (hand made) with human textual productions. This experiment is described in Section 2.

Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), pages 86–89,
Brighton, September 2015.

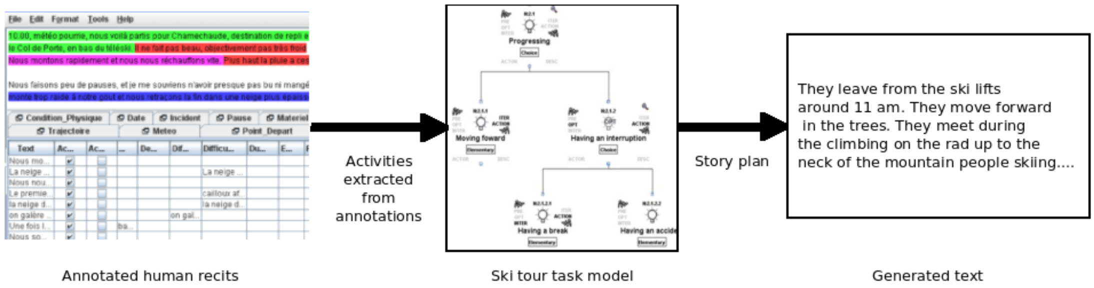

Figure 2: Steps for the récit generation for the task model evaluation.

The second question, though on a completely different aspect than the first one, is linked to an inherent problem in any system based on sensors. What kind of information can be inferred from them? Is this information sufficient? Since, ski touring corpus texts are mainly structured by the route, we report an initial basic data-to-text system that generates texts from GPS data in Section 3.

2 Evaluation of récit plan

To represent the story plan and cope with the precise case of human activity, we propose to use the notion of task model (Caffiau et al., 2010), which has been used previously by (Cavazza et al., 2002) for fictional but interactive stories. As presented in (Baez Miranda et al., 2014), this task model is core to our approach to story generation. In the approach, raw data is firstly captured and interpreted. The resulting interpretations are structured and linked together in a second step, according to the task model. One sequence of the task model is then identified as the story plan and used to drive the generation stage. This aim to result in an activity récit that emerges directly from the sensor data but is organised according to the task model expressing the human activity.

To evaluate the temporal perception of the récit, we followed the steps depicted in Figure 2. Several ski tour récits from www.skitour.fr were collected and annotated by the authors using a schema based on the task model. Then, the annotations were used as input to the task model. This story plan was then linearised into text using chronological order. Note, that to evaluate only the task model and to avoid side effect due to data processing, no raw data was used in this process. For more detail about the process, the reader is referred to (Baez Miranda et al., 2014).

18 French speakers (12 men and 6 women) aged between 19 and 38 were asked to rebuild chronological sequences of ski touring activities after reading separately three récits. The text selection was performed based on the text size, complexity of the ski touring sortie; clarity of the description of the sortie, linguistic quality, and finally the number of protagonists of the sortie and the level of expertise shown in the narration of the sortie. The duration of the experiment was 25 min in average. Each text was presented to each participant (within participants design) in either two versions, (i) the original human written one from the collected corpus of ski touring récits and (ii) the generated text based on the task model. The experiment consisted in sorting cards of basic activity into the sequence of the actual sortie using adhesive tape and a paper-made timeline. Once the reading was finished, the reader choose the cards corresponding to the events encountered in the text. Then, all the cards were arranged on the timeline according to the chronological order perceived during the lecture. The participants did not know whether the text presented was generated automatically or not.

The distance between the participant's answers and the reference story plan was computed using an edit distance similar to the Word Error Rate (WER). An ANOVA performed on the distance value showed a significant effect of text (human vs. computer) ($F(1,18) = 7.583$, $p=0.0131$). A participant effect was also found ($F(17,18)=2.281$, $p=0.0457$). Regarding the size of the participant's sequences, a difference between the human texts and the generated ones was found ($F(1,48) = 5.604$, $p = 0.022$) and a text effect ($F(1,18)=3.666$, $p=0.033$), that appears significant when the text is taken as factor. It seems thus that the generated texts induce significantly less errors during the activity identification than the original ones ($F(1,18)=8.993$, $p=0.00771$).

Regarding the distance, the generated texts present a chronological order more explicit and that may explain why participants were able to perceive easier the structure of the events sequence and to reconstruct the path. In human texts, the chronological order is more implicit because of the text configuration, which can include many satellite details or events omissions, like ellipses.

However, it could be possible to find that some activities were identified in the human texts but not

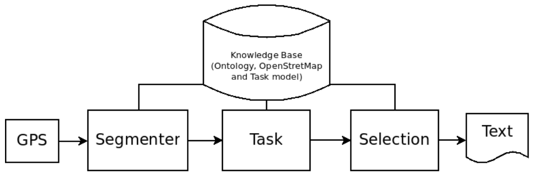

Figure 3: The basic GPS-to-récit system.

in the generated ones. This could be explained by many reasons, such as a possible lack of coverage during the task model construction; activities not identified during the corpus analysis or the fact that, due to the presence of ambiguity in human texts, the participants did not distinguish the activities correctly. Improvements in the task model and in the collection and analysis of the corpus would thus be needed in order to make the approach more robust.

These results show that task model is an eligible support to abstract activity events and structure then. In our approach, instances of task model (récit plan) emerge from ambient sensor data, in the next section, we present a preliminary experiment to extract human activities events (concrete tasks in task model) from GPS data.

3 Generating Récits from real data: The case of the GPS traces

In ski touring, the most important component of the story is the progress of the tour in the followed track. As a matter of fact, the goal of the sortie (e.g., peak, lake, col, etc.) is very often also the goal of the story, although other goals can be found in the human authored corpus (e.g., doing the sortie in the shortest time). The first step is thus to extract the movement and break activities.

To do so, a basic system sketched in Figure 3, has been designed to process the successive geographic localisations provided by a GPS device. First, GPS data from one sortie is temporally segmented based on the altitude. Then, these segments are abstracted into activities. The selection of activities is then performed using the task model so as to obtain a sequence of activities which is valid with respect to the model[1]. Then each activity of the sequence is lexicalised and a simple GRE is performed. Sentence planning is performed using rigid syntactic patterns which are unified with the lexicalised tasks and then realised as text.

3.1 Corpus collection

A small "parallel corpus" was formed through voluntary skiers, involving (i) acquired numerical data and (ii) narratives written by the skiers after their sortie. Physiological and actimetric data were specifically collected for this sortie using a smartphone running the RecordMe application (Blachon et al., 2014) and physiological sensors. These data involve time, location, altitude, heart and breath rate, etc. Extracts of numerical and textual data are shown in Figure 1. This corpus is composed of 5 records (three of which are of couple of skiers) but will grow in the near future.

3.2 Processing

The GPS segmentation consists in aggregating altitude points into segments of points that can be approximated by a straight line with a low amount of error. The Douglas-Peucker algorithm (Douglas and Peucker, 1973) was used for its simplicity. At the end of the process, a list of segments is obtained each of them being labelled as having either a `positive, null or negative` dénivelée[2]. All successive segments with the same dénivelée label are then merged.

Then, the segments are classified based on the average speed of the segment into 'ascending', 'moving forward', 'descending' or 'break'[3]. These segments populate an ontology (Baez Miranda et al., 2014) and are then enriched with links to the next and previous activities, the start and end time, the dénivelée, the average speed as well as the set of participants performing them.

Other important information is Point of Interest. These are encountered along the way (e.g., the Achard lake, the chairlift). These provide: first, an alternative description since ski tour sortie are rarely described by latitude and longitude but by using natural geographical description (See (Turner et al., 2010) for reference) ; second, subgoals to the récit structure since some POIs are main steps to reach the final goal. POI can be extracted using services such as OpenStreetMap which collects information about POI all over the world. For instance a query about the area of the 'Croix de Chamrousse' [4] gives the results presented in Figure 4. From this, every natural elements can be retrieved and associated to the tasks through co-occurrence links.

The abstraction of segments into tasks is for the moment very crude as it consists only of classification based on speed and slope (e.g., a speed of 15km/h in a descending segment is a 'descent' ski activity). Activity selection is then performed following the chronological order and the task model.

[1] Note that this selection is very crude at the moment since not all types of activity can be retrieved from the GPS data.

[2] a dénivelée is a difference in altitude between the starting point and the ending point

[3] 'ascending', 'moving forward', 'descending' are specific cases of the task 'moving forward'

[4] Chamrousse is a famous ski resort in the French Alps

```
<osm version="0.6" >
<node lat="45.1258501" lon="5.9025905">
 <tag k="ele" v="2253"/>
 <tag k="name" v="Croix de Chamrousse"/>
 <tag k="natural" v="peak"/></node>
<node lat="45.1255687" lon="5.9001744">
 <tag k="aerialway" v="pylon"/></node>
 ...
```

```
cat s
actor {P1,P2}
activity descent
locomotion_mode ski
goal {station}
source {Chamrousse_Peak}
time {10:21}
duration {22:32}
```

Figure 4: OpenStreetMap description and semantic representation

For each activity, if the addition of this activity to the set of selected activities makes a valid scenario wrt the model, the activity is added. In any case, the segments containing the main goal of the sortie and the start and end ones should be included into the set of selected activities.

Each activity is translated into a semantic frame. For instance, a descending activity for participant P1 can be represented by the structure in Figure 4. This structure is then matched to predefined set of syntactic structures which constrain lexical choices. The sentence could then be realised as “Departing from Chamrousse. At 08:16 P1 mounts to Col des 3 Fontaines during 1:52. At 10:08 he has a break to Croix de Chamrousse during 0:13 [...]”. The realisation is performed using simpleNLG (Gatt and Reiter, 2009).

4 Future work

The project is at its initial phase and there are many improvements to perform. One of the most important task for the text generation part is to adopt a more structured approach to microplannification. We are working on re-implementing the micro-planner used in the BabyTalk project (Portet et al., 2009). On the macro-planner side, the reasoning must be more integrated so that a dynamic planning is performed and missing data is taken into account. An important challenge is to handle several narrative threads since several skiers can participate to the sortie. Regarding the data processing, the next step will be to include more signals such as physiological ones that can inform about the physiological state of the skier along the track (tired, resting, etc.). This will permit more adaptation of the output toward either sport-like récit (focusing on performance) or leisure one (focusing on where skiers have been).

On the coherence side, to improve and to produce a more natural text, we need to explore other aspects such as temporality. Currently, the story plan from the task model can produce a sequence of events linked in causal way by establishing preconditions and effects during the task model construction. However, this is not reflected in the generated texts. So, we need to add discourse connectors that indicate this causal links. Rendering simultaneous tasks is also an important feature to add to the model. The task model can express this, but it is not yet reflected in the generated text.

Finally, generating an activity récit from sensor data raises specific issues, in particular regarding the paucity of data. Inferencing and reasoning processes are then needed to cope with this lack of information and keep the récit consistent.

References

Jean-Michel Adam. 2011. *Genre de récits. Narrativité et généricité des textes*. Academia.

Belén A. Baez Miranda, Sybille Caffiau, Catherine Garbay, and François Portet. 2014. Task based model for récit generation from sensor data: an early experiment. In *5th International Workshop on Computational Models of Narrative*, pages 1–10.

David Blachon, François Portet, Laurent Besacier, and Stéphan Tassart. 2014. RecordMe: A Smartphone Application for Experimental Collections of Large Amount of Data Respecting Volunteer's Privacy. In *UCAmI 2014*, pages 345–348, Belfast, UK.

S Caffiau, D L Scapin, P Girard, M Baron, and F Jambon. 2010. Increasing the expressive power of task analysis: Systematic comparison and empirical assessment of tool-supported task models. *Interacting with Computers*, 22(6):569–593.

Marc Cavazza, Fred Charles, and Steven J. Mead. 2002. Character-based interactive storytelling. *IEEE Intelligent Systems*, 17(4):17–24.

David H Douglas and Thomas K Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, 10(2):112–122.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of ENLG-2009*.

Inderjeet Mani. 2013. *Computational Modeling of Narrative*, volume 18. Morgan & Claypool.

François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816.

M. O. Riedl and R. M. Young. 2010. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39:217–268.

Ross Turner, Somayajulu Sripada, and Ehud Reiter. 2010. Generating approximate geographic descriptions. In *Empirical Methods in Natural Language Generation: Data-oriented Methods and Empirical Evaluation*, pages 121–140.