**Generating Pictorial Storylines via Minimum-Weight Connected Dominating Set Approximation in Multi-View Graphs**

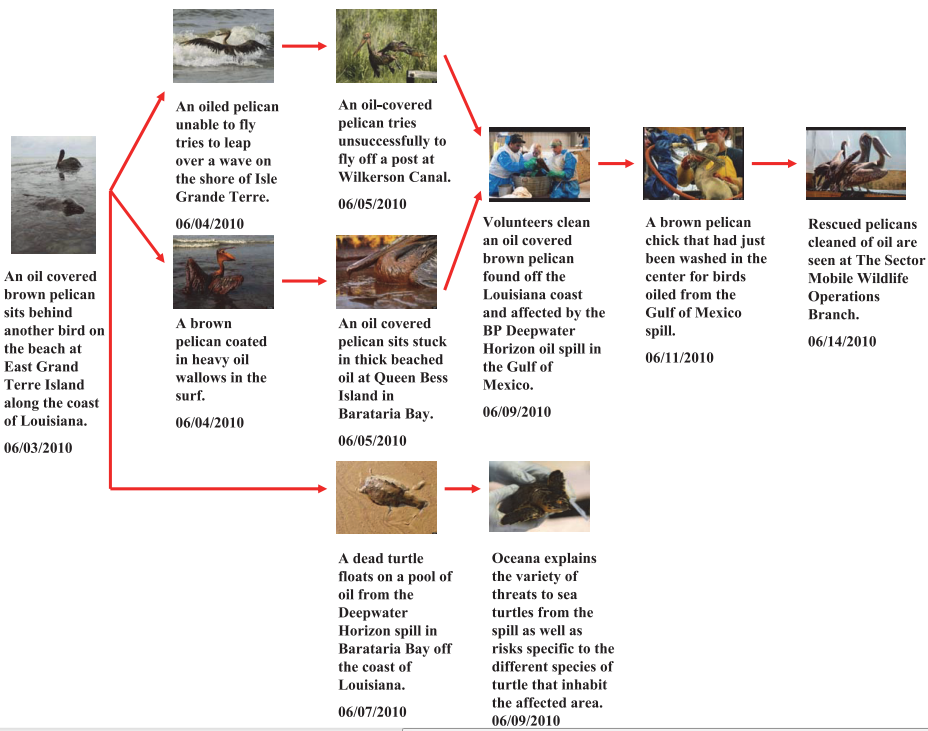
Dingding Wang, Tao Li, Mitsunori Ogihara

Association for the Advancement of Artificial Intelligence, 2012

Summary submitted by *Anunaya Srivastava*

The goal is to generate a pictorial storyline, which is chronologically ordered chain of related events, from a set of given objects. An object is defined as a combination of an image and its related text. The user query is a set of user defined keywords or phrases describing an on-going event in real life. The generated storyline is a graph structure where each node is labelled by a summary of an individual phase of the event and each edge represents causal relationship between two phases. Thus, this paper combines text processing and image processing to form a cross-modal storyline which is not only informative but also visually appealing to the user.

An example of text-image storyline –



The approach used in this paper is very similar to the paper ‘Generating Event Storylines from Microblogs’(GESM) which I have already summarized in the following document.



In fact, GESM cites this paper for developing its framework. GESM has borrowed the idea of formulating the storytelling problem as a graph-based optimization problem and solving it using Minimum Weight Dominant Set(MDWS) summarization technique followed by Directed Steiner Tree algorithm to generate a storyline.

Since, I have already explained the above two algorithms in the attached documents I will not elaborate them here. Instead, I will highlight the difference between the two. GESM builds storyline from tweets which are extremely dynamic, short, noisy and have re-tweets; while this paper uses images and text associated with the text. This paper thus generated a cross-modal storyline. Secondly, while GESM requires Dynamic Pseudo Relevance Feedback(DPRF) to extract relevant tweets from the user query, there is no such requirement in this paper. After extracting relevant tweets using DRPF, both the papers use the same approach for summarizing and generating storyline except the fact that this work used both image and text data. Hence, there is a slight modification in making the multi-view graph.

*A* ***multi-view graph*** *is a quadruple G=(V,W,E,A), where V is a set of vertices (nodes), W is the weights of V, E is a set of undirected edges, which represents the similarities between vertices, and A is a set of directed edges (arcs), which represents the time continuity of the vertices.*

Hence, while making a multi-view graph each node is represents an object i.e. a combination of text and image. Set of edges E is calculated using text as well as image similarities. For text similarity, the standard ‘bag-of-words’ model is used(discussed in GESM). For images similarity, the author calculates their features from colour and texture by adopting colour and edge Directivity Descriptor (CEDD) [Chatzichristofis, S. A., and Boutalis, Y. S. 2008. Cedd: colour and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In Proceedings of ICVS’08, 312–322]. The author uses cosine measure to calculate similarity.