Integration of TV News Stories using Markov Logic

Stanley Kok
Pillar of Information Systems Technology and Design
Singapore University of Technology and Design
Singapore 138682, Singapore
stanleykok@sutd.edu.sg

Introduction

The wealth of audio, visual and textual data offered by TV news stories is a source of tremendous potential benefits. However, news stories are often proliferated across many channels, hampering many applications (such as multimedia information retrieval) from presenting a single concise and coherent view of all the news from every channel. We focus on news stories that are delivered on Singapore's free-to-air channels. Because of Singapore's multi-lingual population, TV programs are delivered in its four official languages of English, Malay, Mandarin and Tamil. News programs in each language typically cover the same national news, but differ in stories that pertain to a specific language community (e.g., coverage of Chinese festivities appear in Mandarin news but not Malay ones). We propose an approach based on Markov logic (Richardson and Domingos 2006) to deduplicate the news stories on all channels, thereby integrating them into a concise, coherent collection. In the next two sections, we briefly review Markov logic and describe our approach.

Markov Logic

In first-order logic (Genesereth and Nilsson 1987), formulas are constructed using four types of symbols: constants, variables, functions and predicates. Constants represent objects in a domain of discourse (e.g., people: `Anna`). Variables (e.g., `x`) range over the objects in the domain. Predicates represent relations among objects (e.g., `Friends`) or attributes of objects (e.g., `Tall`). Variables and constants may be typed. An *atom* is a predicate symbol applied to a list of arguments, which may be variables or constants (e.g., `Friends(Anna, x)`). A *positive literal* is an atom, and *a negative literal* is a negated atom. A *ground atom* is an atom all of whose arguments are constants (e.g., `Friends(Anna, Bob)`). A clause is a disjunction of positive/negative literals. A *world* is an assignment of truth values to all possible ground atoms. A database is a partial specification of a world; each atom in it is true, false or (implicitly) unknown.

Markov logic is a probabilistic extension of first-order logic. A *Markov logic network (MLN)* is a set of weighted first-order formulas. Together with a set of constants, it defines a Markov network (Pearl 1988) with one node per ground atom and one feature per ground formula. The weight of a feature is the weight of the first-order formula that originated it. The probability distribution over possible worlds x specified by the ground Markov network is

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{i \in F} \sum_{j \in G_i} w_i g_j(x)\right) \qquad (1)$$

where Z is a normalization constant, F is the set of first-order formulas in the MLN, G_i and w_i are respectively the set of groundings and weight of the ith first-order formula, and $g_j(x) = 1$ if the jth ground formula is true and $g_j(x) = 0$ otherwise. Markov logic enables us to compactly represent complex models in non-i.i.d. domains. General algorithms for inference and learning in Markov logic are discussed in Richardson and Domingos (2006). A publicly-available implementation is available at http://alchemy.cs.washington.edu (Kok et al. 2006).

Proposed Approach

A straightforward way of integrating news stories from different language channels is to match the visual content of the stories. We could define a metric that captures the similarity of images in different news stories, and specify a predicate `VisuallySimilar` to represent that metric. We could then use a Markov logic rule such as `VisuallySimilar(news1, news2)` $\Rightarrow$ `SameNews(news1, news2)` to capture the extent to which visual content determines the uniqueness of news stories. However, such an approach neglects the audio and textual modes of information that are available in TV content. The audio speeches of newscasters frequently contain information that can be used to match stories. Because the stories are in different languages, we have to first transcribe the speech into text via automatic speech recognition (ASR), and then translate the text into a *lingua franca* via machine translation (MT). Only then can we determine the similarity of the speech content, and use it to match news stories. As a start, we shall use English as the *lingua franca*, and only consider English and Mandarin news channels[1]. An obstacle to this approach is that of obtaining training data in the form of paired speech utterance and text transcription for ASR, and in the form of aligned sentences in two languages for MT. Fortunately, some (but not all) English and Mandarin news channels are professionally subtitled in their respective languages[2]. From these channels, we can obtain the paired speech utterances and subtitles for training an ASR system to understand the Singaporean intonation. How about the training data for MT? This problem can be resolved with Singapore's drama series. Mandarin drama programs in Singapore are all professionally subtitled in English. Using the aforementioned ASR system, we can transcribe the Mandarin speech to textual Mandarin sentences and align them to the English subtitles, thereby obtaining the paired bilingual sentences that are required for training an MT system. The resulting ASR and MT systems can then be used to transcribe and then translate (unsubtitled) Mandarin news programs to obtain English textual transcriptions. Using the ASR system, we could also transcribe speech in (unsubtitled) English news to English text. At this point, we can define a similarity metric between the English sentences from both English and Mandarin news sources, and use it to match news stories. Additional research directions include:

- Using other domain-specific rules to improve matching of news stories (e.g., time and day of news broadcast),
- Learning rules that help to match news stories automatically from videos as in my previous work (Kok and Domingos 2005; 2009; 2010), and
- Scaling up the matching algorithm as in my previous work (Namata, Kok, and Getoor 2011).

References

Genesereth, M. R., and Nilsson, N. J. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.

Kok, S., and Domingos, P. 2005. Learning the structure of Markov logic networks. In *Proceedings of the Twenty-Second International Conference on Machine Learning*, 441–448. Bonn, Germany: ACM Press.

Kok, S., and Domingos, P. 2009. Learning Markov logic network structure via hypergraph lifting. In *Proceedings of the Twenty-Sixth International Conference on Machine Learning*, 505–512. Montreal, Canada: Omnipress.

Kok, S., and Domingos, P. 2010. Learning Markov logic network using structural motifs. In *Proceedings of the Twenty-Seventh International Conference on Machine Learning*. Haifa, Israel: Omnipress.

Kok, S.; Sumner, M.; Richardson, M.; Singla, P.; Poon, H.; and Domingos, P. 2006. The Alchemy system for statistical relational AI. Technical report, Department of Computer Science and Engineering, University of Washington, Seattle, WA. http://alchemy.cs.washington.edu.

Namata, G. M.; Kok, S.; and Getoor, L. 2011. Collective graph identification. In *Proceedings of the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 87–95. San Diego, CA: ACM Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann.

Richardson, M., and Domingos, P. 2006. Markov logic networks. *Machine Learning* 62:107–136.

[1]Extending our proposed approach to Malay and Tamil would only require a straightforward usage of input news stories in the appropriate language.

[2]Unlike USA, close captioning is not widely available in Singapore.