Semantic-Aware Illustration Insertion for Custom Publishing

Lei Liu, Jerry Liu, Shanchan Wu
HP Labs
{lei.liu2, jerry.liu, shanchan.wu} @hp.com

Abstract

Illustrations make reading material come alive. Aside from providing additional information to the text, reading material containing illustration engage our spatial memory, provides anchor points in the reading path, and increases memory retention of the material. However, despite the plethora of available multimedia, adding illustrations to text continues to be a difficult task for the amateur content publisher. To address this problem, we present a semantic-aware illustration insertion system for custom publishing. Compared to using a search engine such as Google Images, our system has the advantage of being able to discern among different topics within a long text passage and recommend the most relevant images for each detected topic with semantic "visual words" based relevance.

Problem statement

Illustrations make reading materials come alive. Aside from providing additional information to the text, reading material containing illustration engages the spatial memory, provides anchor points in reading, and increases memory retention. However, despite all the freely available media accessible through search engines, adding illustrations to text continues to be a difficult task for the amateur content creator. These content creators may be a blogger sharing her subject matter expertise, a father creating a PTA newsletter, or even a teacher authoring her own class material. These subject matter expert can author text quite fluently but may often find locating meaningful illustrations to be a painful task requiring significant time and effort. Thus, custom publications from non-professionals often lack the richness of illustration found in their professional counterparts.

Our solution

We address this problem by developing an semantic-aware illustration insertion system for custom publishing. In a nutshell, given query text of any length, our system first detects the underlying topics from the text. For each topic, we generate a set of keywords to represent the meaning of this topic. We then recommends a list of images that are relevant for each detected topic. Our system makes the following contributions: (1) **recommends images for text queries of arbitrary length**, (2) **detects underlying topics from multi-topic content,** and (3) **introduces a novel semantic "visual words" based image ranking system**. Using text content from the web page where the image originated from, we determine "visual words" as the semantic topic features with probabilistic topic modeling technique. Our processing steps in our system is shown in Figure 1.

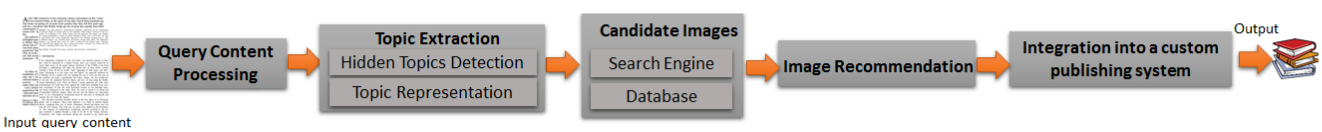

Figure 1. System Architecture

Authors use our system by first submitting text content of any length. These text may be paragraphs, pages or an entire book. The submitted content is fed to ***Query Content Processing*** module responsible for preprocessing the query passages, including removing noisy and stop words, stemming, etc. ***Hidden Topics Detection*** then detects the underlying topics from the submitted query content. We represent each topic with set of terms to indicate the concept for each single topic in ***Topic Representation*** module. ***Image resources*** module fetches relevant images for each extracted topic. We then rank the discovered images based on the user defined criterion in ***Image Recommendation***. Our system can recommend a set of images for any length query content in a ranked list based on "visual words" semantic relevance. The details of each component are discussed in the remaining sections.

Hidden Topics Detection & Representation

Given a text passage as a query, we utilize topic models to discover the abstract topics that underlying the query. Intuitively, provided that a selected passage is about one or more topics, one expects particular words to appear in each topic more or less frequently. After generating the topics, each topic is represented by a set of words that frequently occur together. Examples of topic models include probabilistic latent semantic indexing (PLSI)[5], Latent Dirichlet Allocation (LDA) [6], etc. We use LDA in experiment section of this paper for demonstration purpose. LDA is formally defined a joint distribution of the hidden and observed variables, where $\beta_{1:k}$ is the topics and each β_k is a distribution over the vocabulary. The topic proportions for the *d-th* document are θ_d, where $\theta_{d,k}$ is the topic proportion for topic *k* in document *d*. The topic assignments for the *d-th* document are z_d, where z_{dn} is the topic assignment for the *n-th* word in document *d*. Finally, the observed words for document *d* are w_d, where w_{dn} is the nth word in document *d*, which is an element from the fixed vocabulary.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)$$

After detecting the topics from query content, we extract a small number of words, usually five, with highest probability score in β_k to represent the concept of each detected topic.

Candidate Images: Each topic found from the previous step is provided to the Candidate Images Module, which searches for the top *K* (*K=40* default value)most relevant images with five topic representation words as query from either search engine or image database. The output of this module is a set of *K* most relevant images, with one for each topic. For each candidate image, we extract image URLs, original URL where the image orginated, content from the source webpage, and image size and resolution (height and width).

Image Recommendation: Candidate images can be recommended based on the semantic "visual words" based relevance between query content and content of original page containing the candidate images. This factor is critical to provide a precise image to satisfy user needs. For example, if a passage from the article discusses "*Explain why fireworks different colors with the knowledge of fireworks chemistry?*", then the image in Figure 2 is more appropriate than Figure 3, a general chemistry book cover.

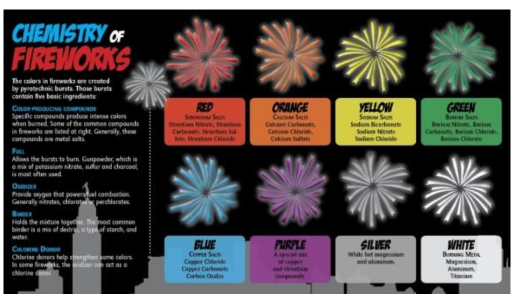

Figure 2. Chemistry of Fireworks

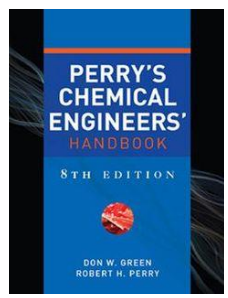

Figure 3. Chemical Book Cover

To perform content based image recommendation, we generate semantic topic features and rerank the candidate iamges based on the content relevance score. More details of topic based content relevance measurement can be accessed in two filed invention disclosures[2][3]. As a post processing step, our system allows user to specify preferences and filtering out the images that fail to fit her requirements. For example, if a user prefers higher resolution images, we can first reranking the top K candidate images based on "visual words" semantic topic features and then remove images that resolution fail to meet the requirement.

Integration into a Custom Publishing System: We provide a user-friendly and intuitive interaction experience for integrating this recommendation solution into a custom publishing service. An author selects a block of text from the edited content and initiates the image insertion process. The service then displays a gallery of multiple images recommended through the techniques described here. The author can selects one or more images to be inserted into the publication. This insertion position can be the beginning of the text passage or the end of the selected passage. The editor can resize and move the images to other positions in the publication if necessary. Our system also supports completely automated image insertion from recommendation where the system scans the entire content of the book and inserts images at pre-determined page or content interbals.

Evidence the solution works

We have integrated our system into the METIS learning platform currently being piloted with San Jose State University and Gunn High School in Palo Alto. While the pilot is ongoing, early feedbacks show that our system is seen favorably by the authors. Figure 4 illustrates an interaction experience with our system. A user select any passage (an example passage is shown in a red rectangle), and click image recommendation icon (outlined in green circle). Recommended image resources are then available for insertion on the popup screen.

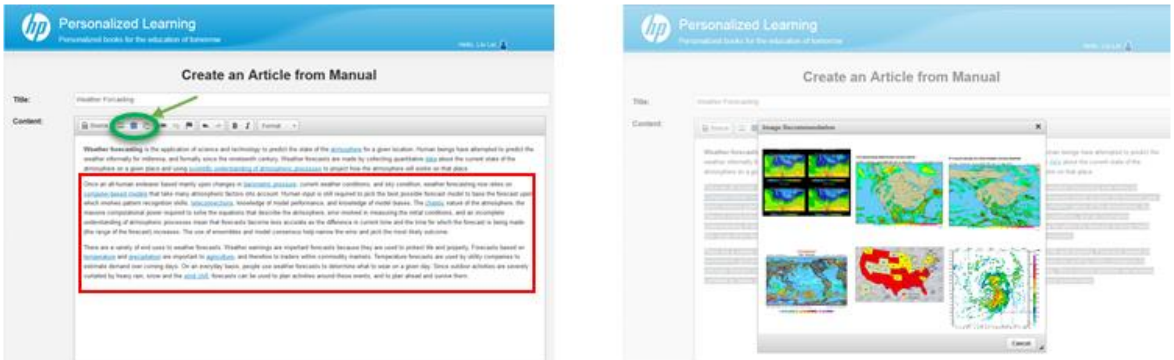

Figure 4. Image Discovery and Insertion for Customized Publishing

Competitive approaches

Given a long text passage with the objective to find images relevant to said passage, a usual practice is to submit the entire text string as a query to an image search engine, like Google Image [7]. However, as existing search engines are designed to only accept a few words as the query, the output from the search engine from a query string that is too long will often be an error message. Another strategy is to manually summarize the long query passage to create a query consisting of a few words to find the relevant images. However, this approach is inefficient and may not accurately represent the query passage. Another key disadvantage with today's systems is that although there may be more than one topic underlying the long query content, existing search engines fail to consider this factor and treat all of these concepts as a single topic with which to find the relevant images.

Current status and Next Steps

We have filed three patents to cover every aspect of image discovery and insertion for custom publishing system[2][3][4]. After feedback from our external pilot, we can work to integrate these capabilities into the open publishing platform efforts in IPS.

References

[1] http://bookbuilder.cast.org/
[2] Lei Liu, Georgia Koutrika, Jerry Liu, Steve Simsk, Identifying Relevant Topics for Recommending a Resource, United States Patent Application No. PCT/US2014/040566
[3] Lei Liu, Georgia Koutrika, Jerry Liu, External Resource Identification, United States Patent Application No. PCT/US2014/053683
[4] Lei Liu, Jerry Liu, Shanchan Wu, Hector Lopez, *Image Discovery and Insertion for Custom Publishing*, HP ID: 700219465
[5] Thomas Hofmann. Probabilistic Latent Semantic Indexing. Preceedings of the 22 Annual International SIGIR Conference
[6] David M. Blei, Andrew Y. Ng and Michael I. Jordan, Latent Dirichlet Allocation. Advances in Neural Information Processing Systems (NIPS), volume14, pages 601-608, Cambridge, MA, 2002
[7] http://www.google.com/imghp
[8] http://goo.gl/larFFr