

Editorial Context Solution Using Real Time Interactive Visualisation–Data Driven Story Telling

A. Akshai R, B. Meenakshi S, C. Rohit Krishnan, D. S Shweta , and E. Venkatesh B P

Abstract—Learning creates the ability to expound your mind and it is the minimum requirement of success in life. The erudite people read an average of 2-3 hours a day. Continuous learning can create new insights and perspectives in life. Efficient method of learning makes it possible to attract even non-readers. As mentioned, efficient method of learning be made up out of representations like snippets, maps and graphs. It involves extracting any type of content such as newspaper, e-book, etc and compress it to bring forth pictorial representation . It inculcates reformative education, saves time during the analysis of surveys and feedback analysis.

Index Terms—Sentiment analysis, visual representation, pictographic memory, feedback, query analysis, survey.

I. INTRODUCTION

One of the major problems associated with learning is the protracted essays and stories that makes the reader lose focus on the nitty-gritty of the magnum opus. In August 2013, Book trust commissioned DJS Research [1], an independent research company, to carry out a quantitative research project investigating reading habits and attitudes of adults in England, and to examine the relationship between reading habits, attitudes to reading and demographic factors such as socio-economic group, age and gender.

It is always easy for a person to remember what he sees than what he listens to. The pictographic memory makes it convenient for the user to observe, understand and learn concepts. This is the reason why people tend to remember the story of the movie they watched months ago in a better way than the essay they read today.

Recently, deep convolutional and recurrent networks for text have yielded highly discriminative and generalizable (in the zero-shot learning sense) text representations learned automatically from words and characters (Reed *et al.*, 2016) [2]-[4]. These approaches exceed the previous state-of-the-art using attributes for zero-shot visual recognition on the Caltech-UCSD birds database (Wah *et al.*, 2011), and also are capable of zero-shot caption-based retrieval. With these upgrades, it is possible to map a text sentence to its corresponding image in the hash table. The conversion of entire text document or voice note to its corresponding visual representation in the form of images, graphs and charts will enable learning in a better way [5].

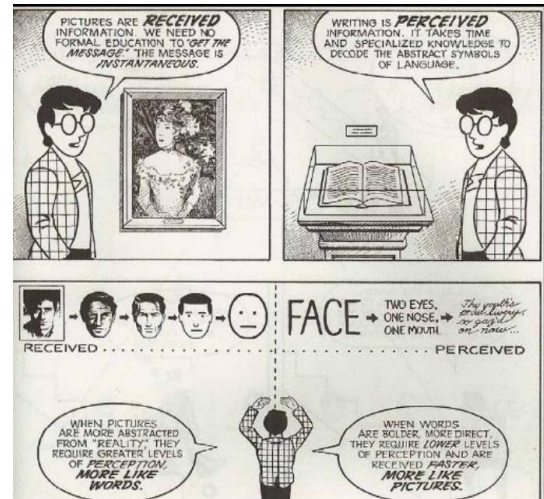


Fig. 1. Represents that understanding the things visually is easy than words.

II. LITERATURE SURVEY

Detecting text based image with optical character recognition for English translation and speech using Android, application that allows smart phones to capture an image and extract the text from it to translate into English and speech it out is no longer a dream. In this study, an Android application is developed by integrating Tesseract OCR engine, Bing translator and phones' built-in speech out technology. Final deliverable is tested by various type of target end user from a different language background and concluded that the application benefits many users.

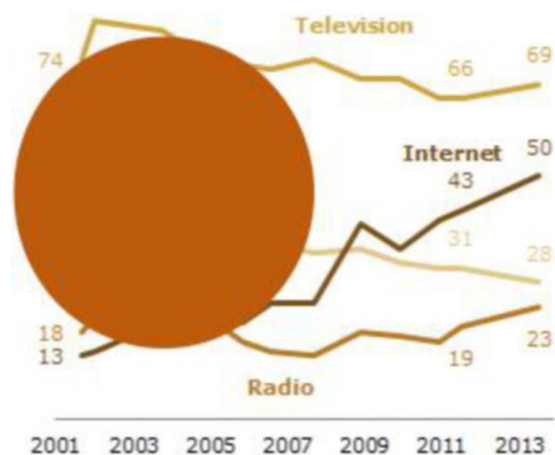


Fig. 2. The circle represents the content which is gathered from the television internet and radio.

Automatic detection and translation of text from natural scenes, Example based machine translation technology for sign translation and present a prototype system for Chinese

Manuscript received May 7, 2018; revised August 21, 2018.

The authors are with Sri Sairam Engineering College, India (e-mail: akshairamesh21@gmail.com).

sign translation. This system is capable of capturing images, automatically detecting and recognizing text, and translating the text into English. The translation can be displayed on a palm size PDA, or synthesized as a voice output message over the earphones [6].

Universal syllable tokeniser for language identification [3], this paper, we have compared two such LID systems namely Gaussian Mixture Model (GMM) tokeniser and syllable based LID system. The phonotactics of GMM and syllable based system are captured by GMM cluster indices and syllable tokens respectively. We propose the use of universal syllable models in building the LID systems and then deriving the uni-gram syllable statistics from this model. Experimental results on the OGI 1992 multilingual speech corpus show that syllable based LID system performs significantly better than the GMM Tokeniser system.

Unsupervised image translation, Bayesian technique for inferring the most likely output image. The prior on the output image $P(X)$ is a patch-based Markov random field obtained from the source image. The likelihood of the input $P(Y/splbsol/X)$ is a Bayesian network that can represent different rendering styles [7]. We describe a computationally efficient, probabilistic inference and learning algorithm for inferring the most likely output image and learning the rendering style. We also show that current techniques for image restoration or reconstruction proposed in the vision literature and image-based non photorealistic rendering could be seen as special cases of our model.

III. PROPOSED SYSTEM

The system makes use of 2 modules, Generator module and the Discriminator module. The Generator module is used to convert the received input message/voice to specified standard syntax. The output of the Generator module is given to the Discriminator module which is used to classify the given input and map the text to an index referring to the corresponding visual (maybe an image, graph or chart). This process is carried out in a feed-forward manner and thus generate a sequence of required visuals.

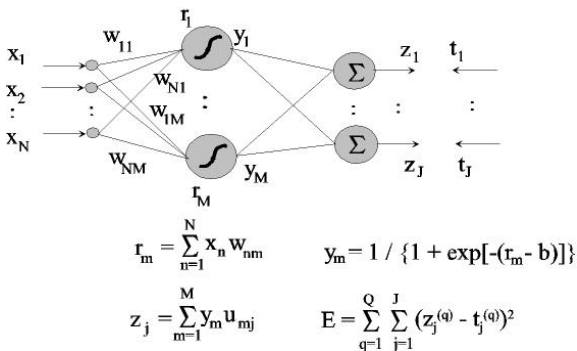


Fig. 3. Feed forward text(x_n) to image(y_n) mapping.

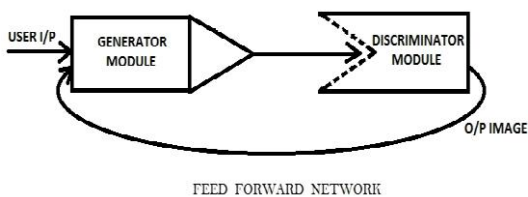


Fig. 4. Working of Generator and discriminator.

IV. IMPLEMENTATION

A. Language Translator

Multiple languages around the world require different character representations. Fortunately, all characters can be encoded into UTF-8 Unicode. UTF-8 is a variable byte sized encoding scheme that can represent up to 4 bytes or 4,294,967,296 characters and is the most widely used encoding scheme for Web pages. Additional character sets are used on web sites. In the Introduction, Fig. 1 shows a list of common character set encoding schemes found on the Internet. Language translation software prefers UTF-8 encoding and text should be converted into UTF-8 prior to translation. The encoding scheme being used may be detected by discovering the Web page's character set or encoding declaration. Most Web pages can easily be converted to UTF-8 using a Python library called BeautifulSoup. Addition programming is required for pages with missing character set declarations. Using publically available software over the Internet, here are the steps necessary to translate a Website from a foreign language into English. Find the Web site and download the HTML content. Remove the unnecessary HTML, script tags, and excess white space. If necessary, 43 capture the text and convert it into UTF-8.

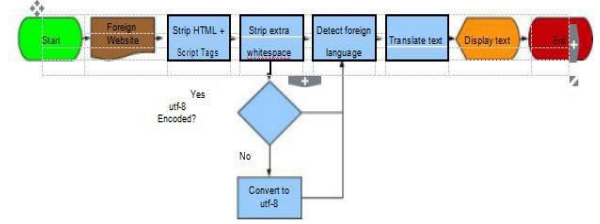


Fig. 5. Language translator design flow.

B. Tokeniser

The tokeniser receives stream of characters, breaks it up into individual tokens. Tokens here refer to each word in the content. This is done by removing punctuation marks from the reviews, removing stop words from the tokenised words and performing spell check on the filtered words by performing Edit Distance(with distance 2). Edit distance is a way of quantifying how dissimilar two strings are to one another by counting the minimum number of operations required to transform one string into the other [8].

The major elements in this approach are creating the syllable segmentation algorithm, universal syllable set and representative syllable model for all the languages. To create the set of universal syllables, we randomly select n syllables ($s_1, s_2, s_3, \dots, s_n$) from each of the languages. We assume that these syllables encompasses the frequently occurring syllables in the languages [9]. These syllables are then clustered using the syllable clustering algorithm. The clustering process will result in 'L' clusters which we refer to as the set of universal syllable models

C. Classifier

The term "classifier" refers to the mathematical function, implemented by a classification algorithm, that maps input data to a category.

How did we evaluate?

Step 1: Have built different classifiers by adopting different machine learning algorithms like K-nearest neighbors, Decision Trees, SVM, and Naïve bayes by changing the parameters for each classifier [10].

TABLE I: MULTIPLE MEANINGS OF THE WORD LIGHT

Noun	I had to light the candle
Verb	Please on the lights
Adjective	I had a light lunch

Step 2: Evaluated the performance of the classifiers using different generators.

a. HoldOut: 10-fold cross validation: It is a technique to evaluate the predictive models by partitioning the original samples into set of 9 training samples to train the models and a test set to evaluate it [11].

b. Resample: Generator: Ran these generators for 10 runs and had done reshuffling of data on each run. This ensures the different order of inputs into the feed forward system.

c. Sentiment Analysis: It is the process of extracting or identifying the subject through natural language processing. It aims to determine the attitude of the end user and the way that user understands the content [11].

SVM Classifier [12], SVM Classifier uses large margin for classification. It separates the tweets using a hyper plane. SVM uses the a discriminative function defined as,

$$g(X) = w^T \phi(X) + b \quad (1)$$

'X' is the feature vector, 'w' is the weights vector and 'b' is the bias vector. $\phi()$ is the non linear mapping from input space to high dimensional feature space. 'w' and 'b' are learned automatically on the training set.

Maximum Entropy Classifier, In Maximum Entropy Classifier, no assumptions are taken regarding the relationship between features. This classifier always tries to maximize the entropy of the system by estimating the conditional distribution of the class label.

The conditional distribution is defined as

$$P_X(y|X) = 1/Z(X) \exp \{ \sum \lambda_i f_i(X, y) \} \quad (2)$$

'X' is the feature vector and 'y' is the class label. $Z(X)$ is the normalization factor and λ_i is the weight coefficient. $f_i(X, y)$ is the feature function which is defined as $f_i(X, y) = (1, X=x_i \text{ and } y=y_i, \text{ otherwise})$.

D. Text to Image Translation

Deep convolutional generative adversarial networks have begun to generate highly compelling images of specific categories, such as faces, album covers, and room interiors. To solve this challenging problem requires solving two sub problems: first, learn a text feature representation that captures the important visual details; and second, use these to Image Synthesis to synthesize a compelling image that a human might mistake for real.

In real world applications, however, images rarely appear in isolation as they are often accompanied by unstructured textual descriptions, such as on web pages and in books. The

additional information from these descriptions could be used to simplify the image modeling task.

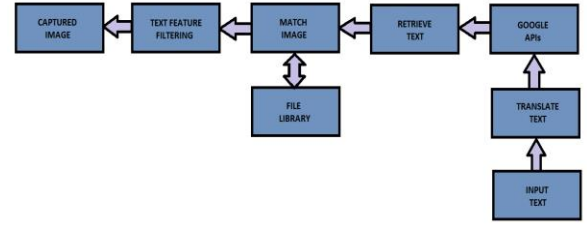


Fig. 6. Text to image translation workflow.

E. Sentence Interpolation

Although there is no ground-truth text for the intervening points, the generated images appear plausible. Since we keep the noise distribution the same, the only changing factor within each row is the text embedding that we use. Note that interpolations can accurately reflect color information, such as a background changing from blue to red while the pose and background are invariant. Here, we sample two random noise vectors. By keeping the text encoding fixed, we interpolate between these two noise vectors and generate images with a smooth transition between two styles by keeping the content fixed [13].

F. Algorithm Used to Produce Visuals from Text

The algorithm makes use of recursion concept to implement the feed forward of text to image mapping after every successful mapping until the entire content is mapped to its corresponding visual [14].

Input: mini batch images x , matching text t , mismatching t^* , number of training batch steps S .

TABLE II: TABLE SHOWS THE STEPS TO IMPLEMENT THE DISCRIMINATOR AND GENERATOR NETWORKS.

Step 1	$h \leftarrow \phi(t)$ {Encode matching text description}
Step 2	$h^* \leftarrow \phi(t^*)$ {Encode mis-matching text description}
Step 3	$Z \leftarrow N(0, 1)^2$ {Draw sample of random noise} $straight \leftarrow D(x, h)$ {real image, right text}
Step 4	$left \leftarrow D(x, h^*)$ {real image, wrong text}
Step 5	$right \leftarrow D(x, h)$ {fake image, right text}
Step 6	$LD \leftarrow \log(straight) + (\log(1 - left) + \log(1 - right))/2$ $LG \leftarrow \log(sf)$
Step 7	$func(D \leftarrow D - \alpha \partial LD / \partial D, G \leftarrow G - \alpha \partial LG / \partial G)$ //recursive feed forward for next image

Algorithm: func(D,G) {Discriminator & Generator Networks}

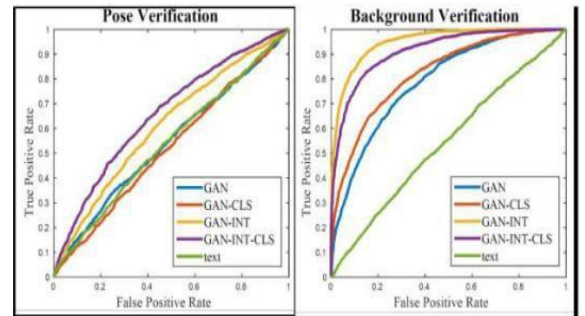


Fig. 7. ROC curves using cosine distance between predicted style vector on same vs. different style image pairs. Left: image pairs reflect same or different pose. Right: image pairs reflect same or different average background color.

V. SCOPE

A. Newspapers & Social Media Content

The news from various sources like television, the additional information about the incident that is posted on the internet and the radio are merged. This information is compressed using compression algorithm and made as a summary that can be visualized.

B. Crime Case Investigation

In case of any crime case the police can get the visualization of the entire crime scene by arranging the evidence they got at the crime spot.



Fig. 8. Data to information mapping.

C. Template & Content

This technique gives the developers a template to create interactive data and story lines quickly. Relationships between various objects, methods and their interactions in a realistic way. By doing so, it provides a focused, interactive and more concise reading experience for the end reader.

D. Learning

This can be used while reading story book contents. The whole story scenario is given as an input and the story is analyzed using sentiment analysis. The analyzed data is compressed and represented in pictorial form. For instance if the story revolves around some place say restaurant then the real time experience of the exact restaurant is shown visually to the end user [15].

E. Process Summary

The input message is first run through the language translator. The translated content is then run through a tokeniser and spellchecker. The tokenised contents are run through the classifier to express the right insight of the message, finally the objects are mapped to corresponding visual representation connected using the connectives defined in the tokeniser. The final image results are mapped to its corresponding data using the hashmap table [16].

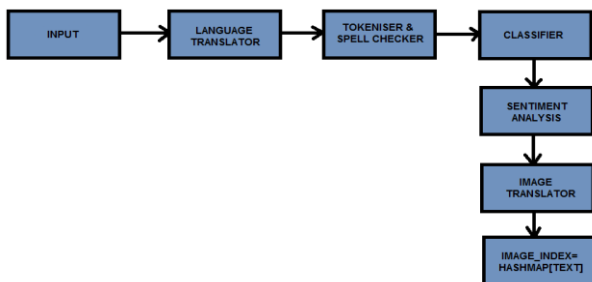


Fig. 9. The work flow of data driven story telling.

Doesn't require more than uploading two spreadsheets of data through a customized uploader: one that focuses on the data while the other on the storytelling [17]. They can easily upload their dataset and link their narrative to certain data

points. The data visualization changes as the readers scroll through the article which essentially transform the reading experience.

VI. CONCLUSION

In this work we developed a simple and effective model for generating images based on detailed visual descriptions to take perspective of seeing things to a new level and make the process of learning simple. We express that the model can synthesize many plausible visual interpretations of a given text caption. Then match it with data in the visual descriptions using sentence interpolation to increase the accuracy of the model. Our manifold interpolation regularizer substantially improved the text to image synthesis. The model can be generalized based on our approach to generating images with multiple objects and variable backgrounds.

REFERENCES

- [1] A. Gleed, *Research Director*, Booktrust Reading Habits Survey 2013.
- [2] F. Bond (2006). Machine translation introduction. [Online]. Available: <http://www.csse.unimelb.edu.au/research/lt/nlp06/materials/Bond/mt-intro.pdf>
- [3] A. Dosovitskiy, J. T. Springenberg, and T. Brox, "Learning to generate chairs with convolutional neural networks," in *Proc. CVPR*, 2015.
- [4] J. D. Ze-ya, Z. Han-fen, Z. Quan, M. Jian-ming, and C. Yuhuan, "Automatic Machine Translation Evaluation Based on Sentence Structure Information," in *Proc. International Conference on Asian Language Processing*, IEEE Computer Society, pp. 162-166, 2009.
- [5] S. Reed, Z. Akata, X. C. Yan, L. L. B. Schiele, and H. Lee "Generative adversarial text to image synthesis," in *Proc. the 33rd International Conference on International Conference on Machine Learning*, New York, NY, USA, Jun. 2016.
- [6] K. Macherey, O. Bender, and H. Ney, "Applications of statistical machine translation approaches to spoken language understanding," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 803-818, May 2009.
- [7] C. F. Tsai and C. Huang, "Automatic annotating images with keywords: A review of image annotation systems," *Recent Patterns on Computer Science*, vol. 1, pp. 55-68, 2008.
- [8] A. Pipitone, M. C. Campisi, and R. Pirrone, "An A* based semantic tokeniser for increasing the performance of semantic applications," *IEEE Seventh International Conference on Semantic Computing*, vol. 13, no. 4, pp. 393-394, 2013.
- [9] J. A. Bakar, K. Omar, M. F. Nasrudin, and M. Z. Murah, "Tokeniser for the Malay language using pattern matching," in *Proc. 14th International Conference on Intelligent Systems Design and Applications*, 2014.
- [10] S. Das and A. K. Kolya, "Sense GST: Text mining & sentiment analysis of GST tweets by naive bayes algorithm," in *Proc. Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2017.
- [11] P. Fabian, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, Oct. 2011.
- [12] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267-307, 2011.
- [13] G. Nasierding and A. Z. Kouzani, "Image to text translation by multi-label classification advanced intelligent computing theories and applications. With aspects of artificial intelligence," *Lecture Notes in Computer Science*, vol. 6216. Springer, Berlin, Heidelberg.
- [14] S. Jasper, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, 2012.
- [15] M. L. Zhang and Z. H. Zhou, "ML - KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038-2048, 2007.
- [16] Y. Bengio, G. Mesnil, Y. Dauphin, and S. Rifai, "Better synthesis mixing via deep representations," in *Proc. ICML*, 2013.
- [17] E. Mansimov, E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Generating images from captions with attention," in *Proc. ICLR*, 2016.



R. Akshai was born on 21st September 1996, in Chennai. He did his schooling in Kendriya Vidyalaya A.F.S, Tambaram, Chennai. He is doing his UG (B.E computer science and engineering) in Sri Sairam Engineering College.

He won 1st prize and got an intern opportunity at ZOHO Pvt. Ltd. held in MIT College, Chennai. His project “Look Out Lantern” was approved by TNSSTC in the year 2016. Currently, He is doing my Internship at ZOHO Pvt. Ltd., as a software developer. And He is a member of Computer Society of India.



S. Meenakshi was born on 31st July 1997 in Villupuram. He completed my higher secondary education in Sri Sairam Sankara Vidyalaya Matriculation Higher Secondary School, Pammal, Chennai. She is doing her under graduate degree in B.E. in computer science and engineering in Sri Sairam Engineering College, Chennai. She is a member of Computer Society of India. And she is Zoho Creator certified associate.



S. Shweta was born on 3rd June 1998, in Tirunelveli. She did my schooling in Kendriya Vidyalaya O.C.F, Avadi, Chennai. Currently, She is pursuing her bachelor of engineering in computer science and engineering in Sri Sairam Engineering College (Anna University), Tambaram, Chennai. She is a member of Computer Society of India.



S. Rohit Krishnan was born on 5th September 1996, in Chennai. He did my schooling in Kendriya Vidyalaya A. F. S, Tambaram, Chennai. He is doing his UG (B.E computer science and engineering) in Sri Sairam Engineering College.

He won 1st prize and got an intern opportunity at ZOHO Pvt. Ltd. held in MIT College, Chennai. He is a member of Computer Society of India.



B. P. Venkatesh was born in Chennai on 24th Seotember 1996. He completed 10th standard at the Senior Secondary School and completed 12th standard at Sir Sivaswami Kalalaya Higher Secondary School. He is currently pursuing his bachelor degree in Sri Sairam Engineering College, Anna University in Computer Science and Engineering discipline. And he is a member of Computer Society of India.