SeqTextVis

Jheng-Long Wu, Chinese Culture University, Taiwan
Mu-Hui Yu*, National Taiwan University, Taiwan
*Corresponded author. E-mail: b03902023@csie.ntu.edu.tw

A Distributed Word Representation Learning for Time Sequential Text Data Visualization

Background figure is from Hikone Castle Museum

Introduction

- **Explicit** and **Implicit** information can be helped to explaining overall **hidden relations among texts**.
  1) Explicit info: be easily observed and explored from co-occurrence matrix in text
  2) Implicit info: difficult to find on time sequential text data.
- **SeqTextVis**: to learn from time sequential text data, and visualize explicit and implicit information
  1) Based on distributed word representation learning methods: word2vec and node2vec.
  2) Using time-decay feature to make dynamic graphs of both explicit and implicit relations chronologically.
- We used case study to explain correctness and effectiveness for SeqTextVis system evaluation.

[Demo page]

Case 1 - Type of Relations Comparison

Finding top related nodes of **Ishida Mitsunari** (石田三成)

- **Co-occurrence**: Most of related nodes are person type.
- **word2vec & node2vec**: Not only person type.

Name	Freq	Name	Freq.	Name	Freq.	Name	Freq.
豊臣秀吉 (P) Toyotomi Hideyoshi	184	島左近 (P) Shima Sakon	172	徳川家康 (P) Tokugawa Ieyasu	57	初芽 (P) Hatsume	56
長政 (P) Nagamasa	47	近江 (L) Oumi	32	淀殿 (P) Yododono	24	佐和山 (L) Sawayama	21

Top	Word2vec (skip-gram)	Word2vec (CBOW)	Node2vec	Node2vec (time decay)
1	曲直瀬道三 (P) Manase Dousan	曲直瀬道三 (P) Manase Dousan	上野 (L) Ueno	島左近 (P) Shima Sakon
2	徳川家康 (P) Tokugawa Ieyasu	井伊直政 (P) Ii Naomasa	佐和山 (L) Sawayama	北庵 (P) Hokuan
3	井伊直政 (P) Ii Naomasa	朝鮮 (L) Korea	浅野家 (O) Asano Clan	熊本 (L) Kumamoto
4	京都 (L) Kyoto	上野 (L) Ueno	豊臣秀吉時代 (T) Toyotomi Reign	近江 (L) Oumi
5	伏見 (L) Fushimi	徳川家康 (P) Tokugawa Ieyasu	織田信長 (P) Oda Nobunaga	石見 (P) Ishimi

More implicit relations

Case 2 - Relation Extraction on Explicit and Implicit

Explicit Case: can be found in co-occurrence and word2vec related list, but not in node2vec list.
Implicit Case: can be found in node2vec and time decay node2vec list. (i.e. Toyotomi Hideyori*)
With time decay: time decay node2vec method can detect both two cases above. (i.e. Okurakyonotsubone*)
Time sequential case: to find other implicit relations in novel.

Sekigahara (関ヶ原)

Top	Co-occuren	Word2vec (skip-gram)	Word2vec (CBOW)	Node2vec	Node2vec (time decay)
1	豊臣秀吉 Toyotomi Hideyoshi	曲直瀬道三 Manase Dousan	曲直瀬道三 Manase Dousan	織田信長 Oda Nobunaga	島左近 Shima Sakon
2	島左近 Shima Sakon	徳川家康 Tokugawa Ieyasu	井伊直政 Ii Naomasa	豊臣秀頼 Toyotomi Hideyori*	北庵 Hokuan
2	徳川家康 Tokugawa Ieyasu	井伊直政 Ii Naomasa	徳川家康 Tokugawa Ieyasu	大野治長 Ono Harunaga	大蔵卿局 Okurakyonotsubone*
4	初芽 Hatsume	島左近 Shima Sakon	大野治長 Ono Harunaga	武田信玄 Takeda Shingen	徳川家康 Tokugawa Ieyasu
5	長政 Nagamasa	池田輝政 Ikeda Terumasa	浅野幸長 Asano Yoshinaga	黒田長政 Kuroda Nagamasa	豊臣秀頼 Toyotomi Hideyori*

Edge weight: Relation Intensity

Conclusion & Future Work

We have presented the SeqTextVis system for time sequential text data to help understanding relations. The SeqTextVis is a very powerful system for analyze and visualize relations. Therefore, SeqTextVis is strong way to find insights. In the future, we propose that using an advanced approach to visualize the mix relations (both explicit and implicit) on a dynamic network graph.

Reference

[1] T Mikolov, I Sutskever, K Chen, GS Corrado, J Dean. Distributed representations of words and phrases and their compositionality. Proceedings of the NIPS 2013, page 3111-3119. December 2013.
[2] J. Pennington, R. Socher, and C. Manning. GloVe: Global vectors for word representation. Proceedings of the EMNLLP 2014, pages 1523-1543, October 2014.
[3] M. Faruqui, J. Dodge, S.K. Jauhar, C. Dyer, E. Hovy, and N.A. Smith. Retrofitting word vectors to semantic lexicons. Proceedings of the NAACL 2015. pages 1606–1615, May 2015
[4] A. Grover and J. Leskovec. Node2Vec: Scalable feature learning for networks. Proceedings of the SIGKDD 2016. pages 855–864. ACM, August 2016.
[5] L. Lin, S. Chen, F. Hong, C. Lai, S. Chen, and X. Yuan. GraphLDA: Latent dirichlet allocation-based visual exploration of dynamic graphs. Proceedings of the IEEE PacificVis 2017, April 2017.

Poster Presentation on the 11th *IEEE Pacific Visualization Symposium* (*PacificVis* 2018) in Kobe, Japan, April 10 to 13, 2018.