

ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC

**Phân giải đồng tham chiếu
trong bệnh án điện tử**

Hội đồng: Khoa Học Máy Tính

Giáo viên hướng dẫn:

GS. TS. Cao Hoàng Trụ

Giáo viên phản biện:

Sinh viên thực hiện:

Nguyễn Duy Hưng – 51101475

Vương Anh Tuấn – 51104040

TP. Hồ Chí Minh, 16/11/2015

Lời cam đoan

Lời cảm ơn

Tóm tắt luận văn

Mục lục

1	Tổng quan	11
1.1	Giới thiệu đề tài.....	11
1.2	Mục tiêu và phạm vi đề tài.....	11
1.3	Cấu trúc luận văn.....	11
2	Các công trình liên quan.....	11
2.1	Bệnh án điện tử.....	11
2.2	Các hướng khai thác dữ liệu bệnh án điện tử.....	11
3	Kiến thức nền tảng	11
3.1	Các định nghĩa và thuật ngữ	11
3.2	Học máy.....	11
3.3	Support Vector Machine	11
3.4	Xử lý ngôn ngữ tự nhiên.....	11
3.5	Phân giải đồng tham chiếu	11
4	Tổng quan hệ thống.....	11
4.1	Nội dung bài toán.....	11
4.2	Quy trình huấn luyện.....	12
4.3	Quy trình phân loại	12
5	Chi tiết hệ thống.....	12
5.1	Tiền xử lý	12
5.2	Xây dựng cặp khái niệm	12
5.3	Rút trích đặc trưng.....	12
5.4	Gom cụm và xây dựng chuỗi đồng tham chiếu	12
5.5	Đánh giá hiệu năng.....	12
6	Thí nghiệm đánh giá	13
6.1	Tập dữ liệu	13
6.2	Kết quả.....	13
7	Tổng kết	13
	Tài liệu tham khảo	13

1 Tổng quan

Từ 2 đến 3 trang

1.1 Giới thiệu đề tài

1.2 Mục tiêu và phạm vi đề tài

1.3 Cấu trúc luận văn

2 Các công trình liên quan

Từ 4 đến 5 trang

2.1 Bệnh án điện tử

[1]

2.2 Các hướng khai thác dữ liệu bệnh án điện tử

[2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]

3 Kiến thức nền tảng

Từ 9 đến 11 trang

3.1 Các định nghĩa và thuật ngữ

3.2 Học máy

3.3 Support Vector Machine

[15], [16]

3.4 Xử lý ngôn ngữ tự nhiên

3.5 Phân giải đồng tham chiếu

[17], [18], [19], [20]

4 Tổng quan hệ thống

Từ 4 đến 5 trang

4.1 Nội dung bài toán

[2]

4.2 Quy trình huấn luyện

[4]

4.3 Quy trình phân loại

[4]

5 Chi tiết hệ thống

Từ 11 đến 15 trang

5.1 Tiền xử lý

[4]

5.2 Xây dựng cặp khái niệm

[4], [19], [21]

5.3 Rút trích đặc trưng

[4]

Nhóm Person

[18]

Nhóm Patient

Nhóm Pronoun

[22] [23] [24]

Nhóm Problem/Treatment/Test

[23] [24] [25] [26] [27] [28] [29]

5.4 Gom cụm và xây dựng chuỗi đồng tham chiếu

[4], [19]

5.5 Đánh giá hiệu năng

Hệ đo MUC

[30]

Hệ đo B-CUBED

[31]

Hệ đo CEAF

[32]

6 Thí nghiệm đánh giá

Từ 4 đến 5 trang

6.1 Tập dữ liệu

[5]

6.2 Kết quả

7 Tổng kết

Từ 1 đến 2 trang

Tài liệu tham khảo

- [1] Hồ Tú Bảo, "Xây dựng và khai thác BADT: con đường mới trong khám chữa bệnh và nghiên cứu y học," *Khoa học & Công nghệ Việt Nam*, vol. 56, no. 3, pp. 16-20, 2015.
- [2] Özlem Uzuner et al., "Evaluating the state of the art in coreference resolution for electronic medical records," *Journal of the American Medical Informatics Association : JAMIA*, vol. 19, no. 5, pp. 786-791, 2012.
- [3] Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall, "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text," *Journal of the American Medical Informatics Association : JAMIA*, vol. 18, no. 5, pp. 552-556, 2011.
- [4] Yan Xu et al., "A classification approach to coreference in discharge summaries: 2011 i2b2 challenge," *Journal of the American Medical Informatics Association : JAMIA*, vol. 19, no. 5, pp. 897-905, 2012.
- [5] National Institutes of Health. Informatics for Integrating Biology and the Bedside. [Online]. <https://www.i2b2.org/>
- [6] Jason Roy, and Walter F. Stewart Jionglin Wu, "Prediction Modeling Using EHR," *Medical Care*, vol. 48, June 2010.
- [7] Robert S. Rahimi, Christopher Clark Amit G. Singal, "An Automated Model Using Electronic Medical Record Data Identifies Patients with Cirrhosis at High Risk for Readmission," *Clinical Gastroenterology and Hepatology*, 2013.
- [8] Dipankar Sengupta and Pradeep K Naik, "SN algorithm, analysis of temporal clinical," *Clinical Bioinformatics*, 2013.
- [9] Mezei, Tong Chung, "Adverse events in ambulatory surgery," April 1999.

- [10] Tianxi Cai Katherine P. Liao, "Electronic Medical Records for Discovery Research in Rheumatoid Arthritis," *Arthritis Care & Research*, vol. 62, August 2010.
- [11] Zhenming Fu, Anushi Shah Hua Xu, "Extracting and Integrating Data from Entire," *AMIA Annu Symp Pro*, October 2011.
- [12] Ahn, Halwan Chu, "A decision support system to facilitate management of patients with acute gastrointestinal bleeding," *Artif Intell Med*, Mar 2008.
- [13] Georgehripcsak, Hua Xu Elizabeths, "Automated Acquisition of Disease-Drug Knowledge from Biomedical and Clinical Documents: An Initial Study," *America Medical Informatics Association*, vol. 15, February 2008.
- [14] Peter B. Jensen, Henriette Schmock Francisco S. Roque, "Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts," *PLoS Comput Biol*, August 2011.
- [15] Stephen Marsland, *Machine Learning: An Algorithmic Perspective.*: Chapman & Hall/CRC, 2009.
- [16] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1--27:27, 2011, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [17] Vincent Ng, "Supervised noun phrase coreference research: the first fifteen years," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala, Sweden: Association for Computational Linguistics, 2010, pp. 1396-1441.
- [18] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim, "A machine learning approach to coreference resolution of noun phrases," *Computational Linguistics*, vol. 27, no. 4, pp. 521-544, December 2001.
- [19] Vincent Ng and Claire Cardie, "Improving machine learning approaches to coreference resolution," in *Proceedings of 40th Annual Meeting on Association for Computational Linguistics*. Philadelphia, Pennsylvania: Association for Computational Linguistics, 2002, pp. 104-111.
- [20] Xiaofeng Yang, Jian Su, and Chew Lim Tan, "A Twin-Candidate Model for Learning-Based Anaphora Resolution," *Computational Linguistics*, vol. 34, no. 3, pp. 327-356, 2008.
- [21] Vincent Ng and Claire Cardie, "Combining Sample Selection and Error-driven Pruning for Machine Learning of Coreference Rules," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10.*: Association for Computational Linguistics, 2002, pp. 55-62.
- [22] Universiteit Antwerpen. Computational Linguistics & Psycholinguistics Research Center. [Online]. <http://www.clips.ua.ac.be/pages/mbasp-tags>

- [23] Sergey Tihon. (2014) OpenNLP.NET. [Online]. <https://github.com/sergey-tihon/OpenNLP.NET>
- [24] Apache. (2010) OpenNLP. [Online]. <https://opennlp.apache.org/>
- [25] Randolph A. Miller, Kevin B. Johnson, Anderson Spickard Joshua C. Denny, "Development and Evaluation of a Clinical Note Section Header Terminology," *Journal of the American Medical Informatics Association : JAMIA*, 2008.
- [26] (2013, November) Open Health Natural Language Processing. [Online]. http://ohnlp.org/index.php/OHNLP_Tool_List
- [27] U.S. National Library of Medicine. Unified Medical Language System Terminology Services. [Online]. <http://uts.nlm.nih.gov/>
- [28] Dr. Alan (Lan) Aronson. (2015, June) MetaMap. [Online]. <https://metamap.nlm.nih.gov/>
- [29] University of Waikato. Wikipedia Miner. [Online]. <http://wikipedia-miner.cms.waikato.ac.nz/>
- [30] Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman, "A Model-theoretic Coreference Scoring Scheme," in *Proceedings of the 6th Conference on Message Understanding*: Association for Computational Linguistics, 1995, pp. 45-52.
- [31] Amit Bagga and Breck Baldwin, "Algorithms for Scoring Coreference Chains," in *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*., 1998, pp. 563-566.
- [32] Xiaoqiang Luo, "On Coreference Resolution Performance Metrics," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*: Association for Computational Linguistics, 2005, pp. 25-32.
- [33] Daniel Chung Yong Lim, Hwee Tou Ng Wee Meng Soon, "A Machine Learning Approach to Coreference Resolution of Noun Phrases," p. 525, 2001.