



专知

www.zhuanzhi.ai

长按二维码关注使用专知

## 自动文摘（ Automatic summarization ） 专知荟萃

---

- 自动文摘（ Automatic summarization ） 专知荟萃
  - 入门学习
  - 进阶论文
  - 代码
  - Tutorial
  - 数据集
  - 领域专家

### 入门学习

---

1. 自动文摘系列（1-13） [<http://rsarxiv.github.io/tags/seq2seq/>]
2. Text summarization with TensorFlow Google 官方发布  
[<https://research.googleblog.com/2016/08/text-summarization-with-tensorflow.html>]

3. Your tldr by an ai: a deep reinforced model for abstractive summarization 强化学习用于文档摘要 [<https://einstein.ai/research/your-tldr-by-an-ai-a-deep-reinforced-model-for-abstractive-summarization>]
4. 教机器学习摘要 [<https://zhuanlan.zhihu.com/p/21426100?refer=paperweekly>]

## 进阶论文

---

1. Rasim M Alguliev, Ramiz M Aliguliyev, Makrufa S Hajirahimova, and Chingiz A Mehdiyev. 2011. MCMR: Maximum coverage and minimum redundant text summarization model. Expert Systems with Applications 38, 12 (2011), 14514–14522. [<http://www.sciencedirect.com/science/article/pii/S0957417411008177>]
2. Rasim M Alguliev, Ramiz M Aliguliyev, and Nijat R Isazade. 2013. Multiple documents summarization based on evolutionary optimization algorithm. Expert Systems with Applications 40, 5 (2013), 1675–1689. [<http://www.sciencedirect.com/science/article/pii/S0957417412010688>]
3. M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut. 2017. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. ArXiv e-prints (2017). arXiv:1707.02919 [<https://arxiv.org/abs/1707.02919>]
4. Einat Amitay and Cécile Paris. 2000. Automatically summarising web sites: is there a way around it?. In Proceedings of the ninth international conference on Information and knowledge management. ACM, 173–179. [<https://dl.acm.org/citation.cfm?id=354756.354816>]
5. Elena Baralis, Luca Cagliero, Saima Jabeen, Alessandro Fiori, and Sajid Shah. 2013. Multi-document summarization based on the Yago ontology. Expert Systems with Applications 40, 17 (2013), 6976–6984. [<http://www.sciencedirect.com/science/article/pii/S0957417413004429>]
6. Taylor Berg-Kirkpatrick, Dan Gillick, and Dan Klein. 2011. Jointly learning to extract and compress. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 481–490. [<https://dl.acm.org/citation.cfm?id=2002534&amp%3bpreflayout=flat>]
7. Asli Celikyilmaz and Dilek Hakkani-Tur. 2010. A hybrid hierarchical model for multi-document summarization. In Proceedings of the 48th Annual Meeting of the

- Association for Computational Linguistics. Association for Computational Linguistics, 815–824. [<https://dl.acm.org/citation.cfm?id=1858765>]
8. Ping Chen and Rakesh Verma. 2006. A query-based medical information summarization system using ontology knowledge. In Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on. IEEE, 37–42. [<https://dl.acm.org/citation.cfm?id=1153019>]
  9. Freddy Chong Tat Chua and Sitaram Asur. 2013. Automatic Summarization of Events from Social Media.. In ICWSM. [<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6057/0>]
  10. John M Conroy and Dianne P O’leary. 2001. Text summarization via hidden markov models. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 406–407. [<http://pdfs.semanticscholar.org/1213/3cfc6688cc2cdea57595b045a28b94d98f1d.pdf>]
  11. Hal Daumé III and Daniel Marcu. 2006. Bayesian query-focused summarization. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 305–312. [<https://dl.acm.org/citation.cfm?id=1220214>]
  12. J-Y Delort, Bernadette Bouchon-Meunier, and Maria Rifqi. 2003. Enhanced web document summarization using hyperlinks. In Proceedings of the fourteenth ACM conference on Hypertext and hypermedia. ACM, 208–215. [<http://dl.acm.org/citation.cfm?id=900097>]
  13. Günes Erkan and Dragomir R Radev. 2004. LexRank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res.(JAIR) 22, 1 (2004), 457–479. [<https://arxiv.org/abs/1109.2128>]
  14. Yihong Gong and Xin Liu. 2001. Generic text summarization using relevance measure and latent semantic analysis. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 19–25. [<https://dl.acm.org/citation.cfm?doid=383952.383955>]
  15. Vishal Gupta and Gurpreet Singh Lehal. 2010. A survey of text summarization extractive techniques. Journal of Emerging Technologies in Web Intelligence 2, 3 (2010), 258–268. [<http://www.learnpunjabi.org/pdf/survey-paper.pdf>]
  16. Ben Hachey, Gabriel Murray, and David Reitter. 2006. Dimensionality reduction aids term co-occurrence based multi-document summarization. In Proceedings of arXiv, July 2017, USA Allahyari, M. et al the workshop on task-focused

summarization and question answering. Association for Computational Linguistics, 1–7.

[<http://www.ltg.ed.ac.uk/np/publications/ltg/papers/Hachey2006Dimensionality.pdf>]

17. John Hannon, Kevin McCarthy, James Lynch, and Barry Smyth. 2011. Personalized and automatic social summarization of events in video. In Proceedings of the 16th international conference on Intelligent user interfaces. ACM, 335–338.  
[<https://dl.acm.org/citation.cfm?id=1943459>]
18. Sanda Harabagiu and Finley Lacatusu. 2005. Topic themes for multi-document summarization. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 202–209.  
[<https://dl.acm.org/citation.cfm?id=1076071>]
19. Leonhard Hennig, Winfried Umbrath, and Robert Wetzker. 2008. An ontologybased approach to text summarization. In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, Vol. 3. IEEE, 291–294. [<http://dl.acm.org/citation.cfm?id=1487345>]
20. Meishan Hu, Aixin Sun, and Ee-Peng Lim. 2007. Comments-oriented blog summarization by sentence extraction. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. ACM, 901–904.  
[<https://dl.acm.org/citation.cfm?id=1321571&CFID=824361189&CFTOKEN=11022411>]
21. Meishan Hu, Aixin Sun, and Ee-Peng Lim. 2008. Comments-oriented document summarization: understanding documents with readers' feedback. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 291–298.  
[<https://dl.acm.org/citation.cfm?id=1390385&CFID=824361189&CFTOKEN=11022411>]
22. Elena Lloret and Manuel Palomar. 2012. Text summarisation in progress: a literature review. Artificial Intelligence Review 37, 1 (2012), 1–41.  
[<https://link.springer.com/article/10.1007%2Fs10462-011-9216-z>]
23. Hans Peter Luhn. 1958. The automatic creation of literature abstracts. IBM Journal of research and development 2, 2 (1958), 159–165. [39] Inderjeet Mani and Eric Bloedorn. 1999. Summarizing similarities and differences among related documents. Information Retrieval 1, 1-2 (1999), 35–67.  
[[http://www.di.ubi.pt/~jpaulo/competence/general/1958\)Luhn.pdf](http://www.di.ubi.pt/~jpaulo/competence/general/1958)Luhn.pdf)]

24. Inderjeet Mani, Gary Klein, David House, Lynette Hirschman, Therese Firmin, and Beth Sundheim. 2002. SUMMAC: a text summarization evaluation. *Natural Language Engineering* 8, 01 (2002), 43–68.
25. Qiaozhu Mei and ChengXiang Zhai. 2008. Generating Impact-Based Summaries for Scientific Literature.. In *ACL*, Vol. 8. Citeseer, 816–824.  
[[https://www.researchgate.net/publication/231901086\\_SUMMAC\\_a\\_text\\_summarization\\_evaluation](https://www.researchgate.net/publication/231901086_SUMMAC_a_text_summarization_evaluation)]
26. Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into texts. Association for Computational Linguistics.  
[<https://digital.library.unt.edu/ark:/67531/metadc30962/>]
27. Rada Mihalcea and Paul Tarau. 2005. A language independent algorithm for single and multiple document summarization. (2005).  
[[https://www.researchgate.net/publication/228340005\\_A\\_language\\_independent\\_algorithm\\_for\\_single\\_and\\_multiple\\_document\\_summarization](https://www.researchgate.net/publication/228340005_A_language_independent_algorithm_for_single_and_multiple_document_summarization)]
28. Liu Na, Li Ming-xia, Lu Ying, Tang Xiao-jun, Wang Hai-wen, and Xiao Peng. 2014. Mixture of topic model for multi-document summarization. In *Control and Decision Conference (2014 CCDC)*, The 26th Chinese. IEEE, 5168–5172.  
[<http://ieeexplore.ieee.org/document/6853102/metrics>]
29. Ani Nenkova and Amit Bagga. 2004. Facilitating email thread access by extractive summary generation. *Recent advances in natural language processing III: selected papers from RANLP 2003* (2004), 287.  
[[https://www.researchgate.net/publication/221303547\\_Facilitating\\_email\\_thread\\_access\\_by\\_extractive\\_summary\\_generation](https://www.researchgate.net/publication/221303547_Facilitating_email_thread_access_by_extractive_summary_generation)]
30. Ani Nenkova and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining Text Data*. Springer, 43–76  
[<https://www.mendeley.com/research-papers/survey-text-summarization-techniques/>]
31. Paula S Newman and John C Blitzer. 2003. Summarizing archived discussions: a beginning. In *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM, 273–276. [<https://dl.acm.org/citation.cfm?id=604097>]
32. You Ouyang, Wenjie Li, Sujian Li, and Qin Lu. 2011. Applying regression models to query-focused multi-document summarization. *Information Processing & Management* 47, 2 (2011), 227–237.  
[<http://www.sciencedirect.com/science/article/pii/S0306457310000257>]
33. Makbule Gulcin Ozsoy, Ilyas Cicekli, and Ferda Nur Alpaslan. 2010. Text summarization of turkish texts using latent semantic analysis. In *Proceedings of the*

- 23rd international conference on computational linguistics. Association for Computational Linguistics, 869–876. [<https://dl.acm.org/citation.cfm?id=1873879>]
34. Vahed Qazvinian and Dragomir R Radev. 2008. Scientific paper summarization using citation summary networks. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 689–696. [<https://dl.acm.org/citation.cfm?id=1599081.1599168>]
35. Vahed Qazvinian, Dragomir R Radev, Saif M Mohammad, Bonnie Dorr, David Zajic, Michael Whidby, and Taesun Moon. 2014. Generating extractive summaries of scientific paradigms. arXiv preprint arXiv:1402.0556 (2014). [[https://www.researchgate.net/publication/229534087\\_Generating\\_surveys\\_of\\_scientific\\_paradigms](https://www.researchgate.net/publication/229534087_Generating_surveys_of_scientific_paradigms)]
36. Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408. [<https://dl.acm.org/citation.cfm?id=638178.638179>]
37. Dragomir R Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroidbased summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization. Association for Computational Linguistics, 21–30. [<http://www.docin.com/p-853652484.html>]
38. Dragomir R Radev, Hongyan Jing, Małgorzata Styś, and Daniel Tam. 2004. Centroid-based summarization of multiple documents. Information Processing & Management 40, 6 (2004), 919–938. [<http://www.sciencedirect.com/science/article/pii/S0306457303000955>]
39. Owen Rambow, Lokesh Shrestha, John Chen, and Chirsty Lauridsen. 2004. Summarizing email threads. In Proceedings of HLT-NAACL 2004: Short Papers. Association for Computational Linguistics, 105–108. [<https://dl.acm.org/citation.cfm?id=1614011>]
40. Zhaochun Ren, Shangsong Liang, Edgar Meij, and Maarten de Rijke. 2013. Personalized time-aware tweets summarization. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 513–522. [<https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/ren-personalized-2013.pdf>]
41. Horacio Saggion and Thierry Poibeau. 2013. Automatic text summarization: Past, present and future. In Multi-source, Multilingual Information Extraction and

- Summarization. Springer, 3–21. [<https://hal.archives-ouvertes.fr/hal-00782442/document>]
42. Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management* 24, 5 (1988), 513–523. [<http://www.sciencedirect.com/science/article/pii/0306457388900210>]
  43. Yogesh Sankarasubramaniam, Krishnan Ramanathan, and Subhankar Ghosh. 2014. Text summarization using Wikipedia. *Information Processing & Management* 50, 3 (2014), 443–461.  
[<http://www.sciencedirect.com/science/article/pii/S0306457314000119>]
  44. Beaux P Sharifi, David I Inouye, and Jugal K Kalita. 2013. Summarization of Twitter Microblogs. *Comput. J.* (2013), bxt109.  
[<http://cs.uccs.edu/~jkalita/papers/2013/SharifiBeauxComputerJournal2013.pdf>]
  45. E. D. Trippe, J. B. Aguilar, Y. H. Yan, M. V. Nural, J. A. Brady, M. Assefi, S. Safaei, M. Allahyari, S. Pouriyeh, M. R. Galinski, J. C. Kissinger, and J. B. Gutierrez. 2017. A Vision for Health Informatics: Introducing the SKED Framework. An Extensible Architecture for Scientific Knowledge Extraction from Data. *ArXiv e-prints* (2017). arXiv:1706.07992 [<https://arxiv.org/abs/1706.07992>]
  46. Neural Summarization by Extracting Sentences and Words  
[<https://arxiv.org/pdf/1603.07252.pdf>]
  47. Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond  
[<https://arxiv.org/pdf/1602.06023.pdf>]
  48. A Neural Attention Model for Abstractive Sentence Summarization  
[<https://arxiv.org/pdf/1509.00685.pdf>]
  49. A Deep Reinforced Model for Abstractive Summarization  
[<https://arxiv.org/pdf/1705.04304.pdf>]
  50. Text summarization using Latent Semantic Analysis  
[[https://www.researchgate.net/publication/220195824\\_Text\\_summarization\\_using\\_Latent\\_Semantic\\_Analysis](https://www.researchgate.net/publication/220195824_Text_summarization_using_Latent_Semantic_Analysis)]
  51. TextRank: Bringing Order into Texts  
[<https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>]
  52. Sentence Extraction Based Single Document Summarization  
[[http://oldwww.iiit.ac.in/cgi-bin/techreports/display\\_detail.cgi?id=IIIT/TR/2008/97](http://oldwww.iiit.ac.in/cgi-bin/techreports/display_detail.cgi?id=IIIT/TR/2008/97)] ### 综述
  53. Automatic Summarization By Ani Nenkova and Kathleen McKeown  
[<https://www.cis.upenn.edu/enkova/1500000015-Nenkova.pdf>]



54. Text Summarization Techniques: A Brief Survey  
[<https://arxiv.org/pdf/1707.02268.pdf>]
55. A SURVEY OF TEXT SUMMARIZATION TECHNIQUES  
[<https://www.cs.bgu.ac.il/~elhadad/nlp16/nenkova-mckeown.pdf~>]
56. Recent automatic text summarization techniques: a survey  
[<https://link.springer.com/article/10.1007/s10462-016-9475-9>]
57. 近 70 年文本自动摘要研究综述 刘家益 邹益民  
[<http://210.76.106.46/qk/90051A/201707/672573777.html>]

## 代码

---

1. Sequence-to-Sequence with Attention Model for Text Summarization.  
[<https://github.com/tensorflow/models/tree/master/research/textsum>]
2. gensim.summarization offers TextRank summarization  
<https://radimrehurek.com/gensim/summarization/summariser.html>

## Tutorial

---

1. 文本自动摘要：现状与未来 万小军 北京大学 2016 年 10 月 16 日  
[<https://pan.baidu.com/s/1nuTUrSP>]
2. Tutorial on automatic summarization [<https://www.slideshare.net/dinel/orasan-ranlp2009>] [<https://pan.baidu.com/s/1o8bZJJk>]
3. How to Run Text Summarization with TensorFlow [<https://hackernoon.com/how-to-run-text-summarization-with-tensorflow-d4472587602d>]
4. Text Summarization with Gensim [<https://rare-technologies.com/text-summarization-with-gensim/>]

## 数据集

---

1. DUC 2004 [<http://www.cis.upenn.edu/~nlp/corpora/sumrepo.html>]
2. Opinosis Dataset - Topic related review sentences [<http://kavita-ganesan.com/opinosis-opinion-dataset>]
3. 17 Timelines [<http://kavita-ganesan.com/opinosis-opinion-dataset>]
4. Legal Case Reports Data Set  
[<http://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports>]



## 领域专家

---

1. 万小军 清华大学 [<https://sites.google.com/site/wanxiaojun1979/>]
2. 秦兵 哈工大  
[<https://m.weibo.cn/u/1880324342?sudaref=login.sina.com.cn&retcode=6102>]
3. 刘挺 [<http://homepage.hit.edu.cn/pages/liuting>]

汇总不全面，欢迎补全和提建议，敬请关注 <http://www.zhuanzhi.ai> 和关注专知公众号，获取第一手 AI 相关知识。

