白质序列库可以从 http://gi.cebitec.uni-bielefeld.de/comet/force/indexOld.html 上下载.表 3 描述的是文献[13]中的序列数据库 SDB1 包含的 12 个序列的特征.

Table 2 Characteristic of experimental data

表 2 实验数据集特征

序列数据库	数据集名称	序列类型	序列个数	总长度
SDB1	DNA 片段	DNA 片段	12	327 535
SDB2	WO02059377	DNA 序列	70	190 533
SDB3	ASTRAL95_1_161	蛋白质序列	507	91 875
SDB4	文献[26]	文本字符串	1	8 191

Table 3 Sequences of real biological data in SDB1

表 3 SDB1 真实生物数据片段

序号	位点	片段长度	序号	位点	片段长度
S_1	CY058560	844	S_7	CY058563	2 286
S_2	CY058557	982	S_8	CY058562	2 299
S_3	CY058558	1 418	S_9	AX829178	5 393
S_4	CY058559	1 516	S_{10}	AX829174	10 011
S_5	CY058556	1 720	S_{11}	AB038490	131 892
S_6	CY058561	2 169	S_{12}	AL158070	167 005

4.2 实验结果及分析

4.2.1 DNA 片段上算法的比较

DNA 控制 RNA 的转录以及遗传信息,因此,通过对 DNA 片段进行特定的模式匹配,找出匹配解的个数,对于致病基因以及遗传信息的检测、病毒传播的预防等起着重要的作用.一般间隙模式匹配有利于灵活地找出模式匹配个数的精确解,通过对匹配解的数目分析,对生物学中基因遗传的研究具有重要的价值.下面将具体介绍 MSAING 以及对比算法在 DNA 片段上的性能.为了分别对比测试 MSAING 算法在模式串不存在重复字符以及存在重复字符时的求解性能,选取了 P_1~P_4 模式为不存在重复字符的 4 种模式,P_5~P_9 模式为存在重复字符的 5 种模式,具体模式串在表 4 中给出.这 7 种算法的模式串的测试结果见表 5,其中给出了各种算法在不同模式下得到的出现数的总和;表 6 给出了各种算法在不同模式以及序列下的消耗时间总和.

Table 4 Patterns

表 4 模式串

序号	模式串
P_1	$a[-5,6]c[-4,7]g[-3,8]t$
P_2	$c[-1,2]a[-2,3]t[-3,4]g$
P_3	$g[1,2]t[0,3]c[-3,4]a$
P_4	$t[-2,2]c[-2,2]a$
P_5	$g[-1,5]t[0,6]a[-2,7]g[-3,9]t[-2,5]a[-4,9]g[-1,8]t[-2,9]a$
P_6	$g[-1,5]t[0,6]a[-2,7]g[-3,9]t[-2,5]a[-4,9]g[-1,8]t[-2,9]a[-1,9]g[-1,9]t$
P_7	$t[-1,7]t[-1,7]a[-1,7]g[-1,7]t[-1,7]a[-1,7]g$
P_8	$a[-1,7]a[-1,7]a[-1,7]a[-1,7]c$
P_9	$c[-1,7]a[-1,7]a[-1,7]a[-1,7]a$

Table 5 Comparison of the number of occurrences of P_1~P_9 between 7 algorithms

表 5 7 种算法在 P_1~P_9 上的出现个数

算法名称	模式串中不存在重复字符				模式串中存在重复字符				
	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
SAIL_Gen	36 327	23 885	13 257	31 132	10 961	8 289	17 790	18 192	18 401
RSAIL_Gen	36 327	23 885	13 257	31 132	10 961	8 289	17 790	18 192	18 342
SGSP_Gen	35 876	23 520	13 238	30 667	13 548	10 004	17 766	18 096	19 033
SBO_Gen	36 267	23 793	13 250	30 984	14 101	10 528	18 056	18 831	19 204
RBCT_Gen	36 253	23 763	13 213	30 947	14 125	10 593	18 052	18 822	19 433
DCNP	36 267	23 793	13 250	30 984	14 152	10 673	18 070	18 865	19 898
MSAING	36 327	23 885	13 257	31 132	14 158	10 786	18 094	20 136	20 166

Table 6 Comparison of the running time on P_1~P_9 between 7 algorithms (s)

表 6 7 种算法在 P_1~P_9 上的运行时间 (s)

算法名称	模式串中不存在重复字符				模式串中存在重复字符				
	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
SAIL_Gen	0.98	0.93	0.85	0.82	4.17	5.54	2.67	2.03	2.05
RSAIL_Gen	1.26	1.17	1.98	1.89	4.21	6.36	3.23	3.18	3.56
SGSP_Gen	640.9	521.9	491.21	166.3	2 759.1	3 175.2	1 796.7	1 243.8	868.96
SBO_Gen	838.5	790.9	720.3	315.0	3 163.3	3 439	1 796.7	1 552.1	478.84
RBCT_Gen	0.82	0.78	0.83	0.89	3.25	4.12	1.05	1.14	1.39
DCNP	852	802.4	781.71	361.1	3 562.2	3 887.4	2 065.6	1 635.7	1 208.6
MSAING	0.93	0.83	0.81	0.8	7.2	9.69	3.14	2.1	3.37

(1) 在模式串中不存在重复字符的情况下,MSAING,SAIL_Gen,RSAIL_Gen 均属于完备算法.文献[8,24]证明了这种情况,从表 5 中可以看出,MSAING 能够达到完备的情况,与定理 2 的理论分析相一致.

(2) 在模式串中有重复字符的情况下,MSAING 算法性能最好,DCNP 算法性能次之,SAIL_Gen 算法性能最差.p_5~p_9 模式串中具有重复的字符,其中,p_8 为首部重复模式,p_9 为尾部重复模式,从表 5 可以清晰地看到,SAIL-Gen 不能在任何实例上取得最好的结果;而且模式串 p_5~p_9 在 12 个序列的出现总和中,SAIL-Gen 算法取得了最小值,占 MSAING 算法匹配数的 88%,而 DCNP 算法获得了次优匹配,占 MSAING 算法匹配数的 97%.这充分说明,对于 R 模式不宜采用 SAIL-Gen 算法进行求解.DCNP 算法在处理 RH 模式和 RT 模式上完备性较低,产生这种现象的原因是:DCNP 算法在 SBO_Gen 算法的基础上动态更新节点属性,采取每次选择出现相关数较少的策略,因而其可以很好地解决 one-off 条件下一般间隙的模式匹配问题.但是,由于只是局部最优的贪婪算法,因此未必能够达到全局最优,这种缺陷在连续重复字符的模式串,如 RH 和 RT 模式串上表现更为明显.而 MSAING 算法通过最左侧匹配以及回溯方法使每次匹配达到局部最优,同时采用内部检测机制避免内部重复,利用 Reverse 策略对模式串的结构以及符号集Σ中各个元素在序列串中的频度进行分析,使其达到最佳的匹配状态,保证匹配达到全局最优.

(3) 根据表 6 时间消耗可知,DCNP 算法消耗时间最长,RBCT_Gen 算法消耗时间最少;而在不存在重复字符的模式串中,MSAING 算法消耗的时间次之;在存在重复字符的模式串中,SAIL-Gen 算法消耗的时间次之.SAIL 是文献[8]提出的一种在线算法,具有最好的时间性能来解决非间隙模式匹配问题,SAIL-Gen 和 MSAING 算法继承了这种特性,当模式串不存在重复字符,进行一般间隙模式匹配时,SAIL-Gen 需要判断是否内部重复,消耗了大量的时间,而 MSAING 算法具有内部检查机制使时间消耗减少;在存在重复字符的模式串匹配中,MSAING 算法为了提高解的完备性,增加了回溯的功能,使消耗的时间变多.虽然 SAIL-Gen 耗时较少,但是基于一般性模式匹配问题时,求解性能很差,因此不宜采用 SAIL-Gen 算法进行求解.同理,RBCT_Gen 利用 CluTree 结构,在匹配的过程中减少了时间的消耗,但是相对于 MSAING 和 DCNP 算法,求解性能较差.DCNP 算法在 SBO_Gen 算法的基础上动态更新节点属性,因此所用的时间最长.综上所述,MSAING 算法解的性能优于 DCNP 和 SBO_Gen 算法.

通过将 MSAING 算法与其他各种算法在 SDB1 数据集上进行对比可知:MSAING 算法的匹配解比 DCNP 提高了 2%,比 SAIL_Gen 提高了 12%.MSAING 算法是通过 Reverse 策略提高匹配解的数目的,下面将通过实验进一步验证 Reverse 机制的有效性.

4.2.2 Reverse 机制有效性验证

为了验证 Reverse 机制的有效性,在数据集 SDB1 上选取需要转置的模式串 p_5,p_6,p_9,计算每个模式串在 12 条 DNA 片段上解的和.由图 7 可知,MSAING 算法转置后的解的完备性较高,说明了 Reverse 机制的有效性.这是由于 MSAING 算法通过对模式串结构的分析以及序列串中个字符元素的频度的计算,找出最佳的模式匹配形式,使解的完备性得到了提高.

本实验验证了 Reverse 策略的有效性,使匹配解的平均数目提高了 5.6%.下面将通过在序列数据库中比较算法的性能,证明 MSAING 算法能够在较大规模的数据库上进行有效的模式匹配.

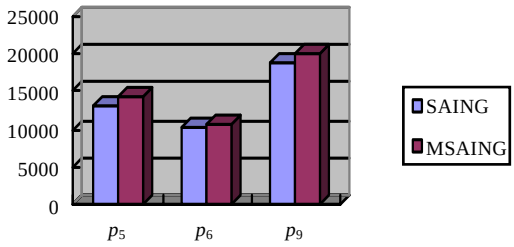


Fig.7 Validation of effectiveness of Reverse strategy

图 7 Reverse 机制有效性验证

4.2.3 序列库中算法性能的比较

随着蛋白质产品在人们日常生活中应用的普及,蛋白质检测在生物学领域得到了广泛的研究.食品中营养蛋白的检测、药物中蛋白酶以及胰岛素的检测以及蛋白质中氨基酸序列的检测等,都需要对蛋白质序列进行某些特定的模式串匹配.一般间隙的模式匹配能够灵活地找出匹配解的数量,在生物学中某种蛋白质含量的监测中有着重要的应用.下面的实验将比较 MSAING 与其他算法在 DNA 库以及蛋白质库中的性能.

为了更有效地对比各种算法匹配解的个数以及运行的速度,在 DNA 序列库 WO02059377、蛋白质序列库 ASTRAL95_1_161 上做了对比实验.

在 DNA 序列库上的模式匹配,模式串为表 3 中的 p_1~p_9,计算每个模式串在整个序列库中解的和.由图 8 可知:在 NR 模式串 p_1~p_4 的匹配中,SAIL_Gen,MSAING 取得了完备解;在 R 模式串 p_5~p_9 中,MSAING 取得了最优解,而 SAIL_Gen 算法解的完备性最差.这是由于 SAIL_Gen 算法只适合在线的模式匹配,在离线的情况下仅利用最左优先策略,而没有考虑回溯的情况,导致完备性较差.图 9 中可以看出:MSAING,SAIL_Gen 和 RBCT_Gen 算法的运行消耗的时间都比较少,而 DCNP 消耗的时间最多.这是由于 DCNP,SBO_Gen 都需要建立网树结构,消耗了大量的时间.

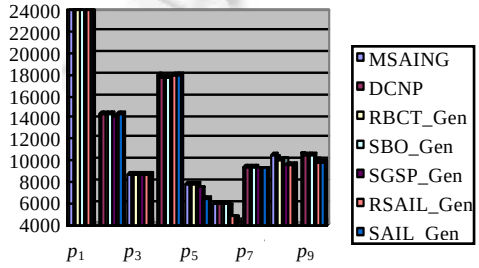


Fig.8 Comparison of the number of occurrences on DNA sequence database

图 8 DNA 序列库上出现的数目对比

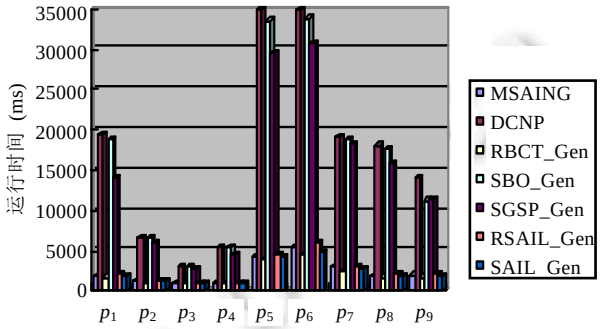
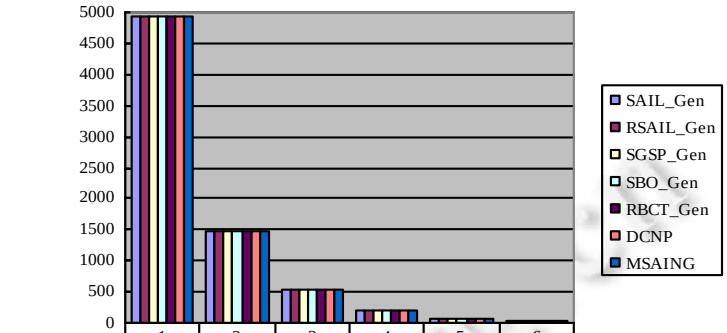


Fig.9 Comparison of the running time on DNA sequence database

图 9 DNA 序列库上运行时间对比

在蛋白质序列库上,分别选择长度为 1~6、间隙为[–2,7]的模式串,其中,长度为 1 模式串为{*a*,*c*,*d*,*e*,*f*},长度为 2 模式串是由长度为 1 的字符组合而成的 25 条模式串,模式长度为 3 的模式串 125 条,以此类推,模式串长度为 6 的 15 625 条模式串;然后,求相同长度的模式串在蛋白质序列库上出现解的和,并计算其平均值.通过图 10 可知:随着模式长度的变大,每个模式串出现的平均个数在减小.这是由于长度为(m+1)的模式串是在长度为 m 的模式串的基础上满足 p_m 的匹配.因此,模式越长,出现解的概率越低.其中,MSAING 算法的完备性最高,DCNP 算法的完备性次之.图 11 中比较的是各种算法的完备性,图中纵坐标是各种算法解的数目比上 MSAING 算法解的数目所获得的百分数.随着模式长度的增加,MSAING 算法的完备性相对于其对比算法的完备性越来越高.这是由于随着模式长度的增加,模式串中含有重复的字符越来越多,DCNP,RBCT_Gen,SBO_Gen,SGSP_Gen,RSAIL_

Gen,SAIL_Gen 算法只是利用局部最优策略,而没有考虑到模式串的结构与序列串中各个字符出现的频度之间的关系,导致模式串长度越大,解的完备性就越差.图 12 为每条模式串在蛋白质序列库上运行的平均时间,可以看出:随着模式越来越长,模式匹配消耗的时间越来越多.其中,RBCT_Gen,SAIL_Gen,MSAING 增长速度比较缓慢;而 DCNP,SBO_Gen,SGSP_Gen 增长速度较快,消耗时间较多.



	1	2	3	4	5	6
SAIL_Gen	4936.2	1467.32	524.408	198.0768	76.09248	29.44774
RSAIL_Gen	4936.2	1469.21	524.413	199.635	78.02596	30.02466
SGSP_Gen	4936.2	1470.12	524.696	200.8736	79.45216	31.88486
SBO_Gen	4936.2	1470.36	526.424	201.5344	79.64032	31.92928
RBCT_Gen	4936.2	1470.36	526.703	201.624	79.65146	31.94521
DCNP	4936.2	1470.36	526.728	201.6864	79.66912	31.96653
MSAING	4936.2	1470.72	527.088	202.024	80.35456	32.50893

Fig.10 Comparison of the number of occurrences on protein sequence database

图 10 在蛋白质序列库上出现的对比

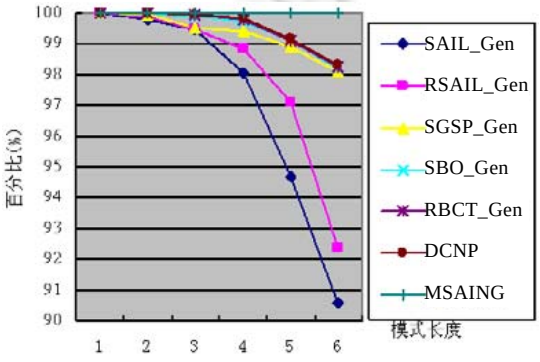


Fig.11 Percentage of each algorithm with MSAING on protein sequence database

图 11 在蛋白质序列库上各算法的出现与 MSAING 算法的百分比

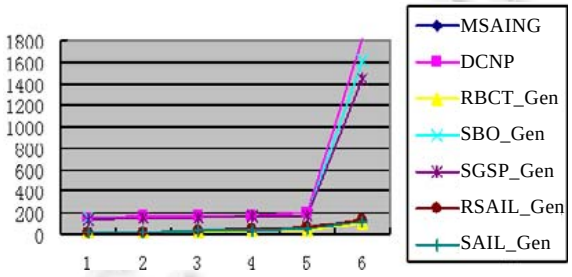


Fig.12 Comparison of the running time on protein sequence database

图 12 在蛋白质序列库中运行时间对比

在生物序列上的实验,验证了 MASING 算法进行较大规模数据库模式匹配时的性能.为了进一步证明该算法解决某些实际问题的有效性,本文把它应用在文本信息的检索中.

4.2.4 文本信息检索中的应用

在文本信息检索中考虑一般间隙更具有实际意义,例如在文献[26]中,出现 frequent closed subsequence 与带负间隔的模式 closed frequent subsequence 表示相同的意思.通过统计文本中模式出现的次数,可以实现文本中关键词的抽取.带一般间隙的模式匹配将文本中单词、短语视为模式串,通过提高文本中模式出现解的完备性,

从而可以进一步提高关键词抽取的精确度,在文本信息检索的过程中获得更多有价值的信息.

以文献[26]作为信息检索文本,分别检索 data mining,closed subsequence,frequent closed subsequence 在文本中出现的次数,由表 7 可知:一般间隙的模式匹配出现的次数更多,更具有灵活性.为了更好地验证一般间隙的性质,计算在文献[26]中所有长度为 2 的模式串出现的次数之和,非负间隙将给定的间隙的最小值设为 0.从图 15 可以看出:随着模式串的间隙的增大,解的个数不断增多,一般间隙比非负间隙获得更多的解.这是由于一般间隙在非负间隙的基础上又考虑了负间隙的情况,即模式串中字符的顺序发生颠倒的情况;同时,间隙越大,满足匹配出现的概率也越大,解的个数也就越多.

Table 7 Comparison of the number of occurrences about no-negative gap with general gap

表 7 非负间隙与一般间隙出现个数比较

模式	非负间隙	一般间隙
Data mining	2	3
Closed subsequence	3	4
Frequent closed subsequence	4	9

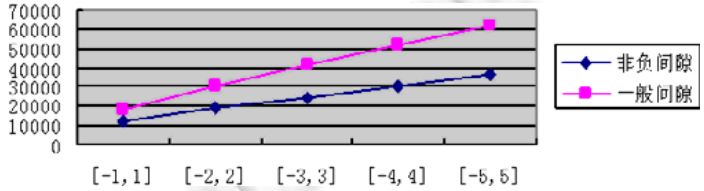


Fig.13 Comparison of the number of occurrences about no-negative gap with general gap

图 13 非负间隙与一般间隙解的出现对比

5 结 论

本文提出了一般间隙与 one-off 条件约束的模式匹配问题——SPMGOO 问题.该问题的研究允许用户更加灵活地设定模式串.该问题具有如下特点:间隙可以为负;同时,序列串中任何字符最多只能使用 1 次.本文把线性表应用于该问题的求解,并基于线性表的结构提出了 MSAING 算法.算法首先采用 Reverse 策略使模式与序列达到最佳的匹配状态,克服了某些算法容易陷入局部最优的缺点;其次,利用线性表的结构使匹配过程中的时间和空间消耗大为减少,并利用回溯的方法提高匹配的成功率;最后,根据 *inside_Checking* 机制判断模式串内部是否会产生重复现象,有效提高算法运行效率.最后,本文从理论和实验两个方面验证了 MSANIG 算法匹配的有效性.

本文只是对一般间隙与 one-off 条件的模式匹配问题进行了研究,而模式匹配是序列模式挖掘的基础,因此,下一步将对一般间隙的序列模式挖掘问题进行研究.该研究将有助于挖掘更多有价值的频繁模式.此外,在实际应用中有很多模式匹配是近似模式匹配,这不但具有实际意义,而且研究难度也会更大,这些问题均是未来研究的方向.

References:

[1] Wu XD, Zhu XQ, He Y, Arslan AN. PMBC: Pattern mining from biological sequences with wildcard constraints. Computers in Biologyand Medicine, 2013,43(5):481−492. [doi: 10.1016/j.compbiomed.2013.02.006]

[2] Zhang C, Zheng Y, Ma XL, Han JW. Assembler: Efficient discovery of spatial co-evolving patterns in massive geo-sensory data. In: Proc. of the 21th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2015. 1415−1424. [doi: 10.1145/2783258.2783394]

[3] Shokoohi-Yekta M, Chen YP, Campana B, Hu B, Zakaria J, Keogh E. Discovery of meaningful rules in time series. In: Proc. of the ACM SIGKDD Int'l Conf. 2015. 1085−1094. [doi: 10.1145/2783258.2783306]

[4] Chou CP, Jea KF, Liao HH. A syntactic approach to twig-query matching on XML streams. Journal of Systems and Software, 2011, 84(6):993−1007. [doi: 10.1016/j.jss.2011.01.033]

[5] Manber U, Baeza-Yates R. An algorithm for string matching with a sequence of don't cares. Information Processing Leters, 1991, 37(3):133−136. [doi: 10.1016/0020-0190(91)90032-d]

[6] Bille P, Gørtz IL, Vildhøj HW, Wind DK. String matchingwith variable length gaps. In: Chavez E, *et al*., eds. Proc. of the 17th Int'l Conf. on String Processing and Information Retrieval. Berlin: Springer-Verlag, 2010. 385−394. [doi: 10.1007/978-3-642-16321-0_40]

[7] Fischer MJ, Paterson MS. String matching and other products. In: Proc. of the String-Matching and Other Products. Cambridge: Massachusetts Institute of Technology, 1973. 113−125.

[8] Chen G, Wu XD, Zhu XQ, Arslan AN, He Y. Efficient string matching with wildcards and length constraints. Knowledge and Information Systems, 2006,10(4):399−419. [doi: 10.1007/s10115-006-0016-8]

[9] Zhu XQ, Wu XD. Discovering relational patternsacross multiple databases. In: Proc. of the IEEE 23rd Int'l Conf. on Data Engineering. 2007. 726−735. [doi: 10.1109/ICDE.2007.367918]

[10] Liu YL, Wu XD, Hu XG, Gao J. A matching algorithm in PMWL based on CluTree. New Generation Computing, 2014,32(2): 95−122. [doi: 10.1007/s00354-014-0201-3]

[11] Fredriksson K, Grabowski S. Efficient algorithms for pattern matching with general gaps and character classed. In: Crestani F, *et al*., eds. Proc. of the Int'l Conf. on String Processing and Information Retrieval. Berlin: Springer-Verlag, 2006. 267−278. [doi: 10.1007/11880561_22]

[12] Fredriksson K, Grabowski S. Efficient algorithms for pattern matching with general gaps, character classes, and transposition invariance. Information Retrieval, 2008,11(4):335−357. [doi: 10.1007/s10791-008-9054-z]

[13] Chai X, Jia XF, Wu YX, Jiang H, Wu XD. Strict pattern matching with general gaps and one-off condition. Ruan Jian Xue Bao/ Journal of Software, 2015,26(5):1096−1112 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4707.htm [doi: 10.13328/j.cnki.jos.004707]

[14] Wu YX, Fu S, Jiang H, Wu XD. Strict approximate pattern matching with general gaps. Applied Intelligence, 2014,42(3): 1−15. [doi: 10.1007/s10489-014-0612-3]

[15] Wu YX, Liu YW, Guo L, Wu XD. Subnettrees for strict pattern matching with general gaps and length constraints. Ruan Jian Xue Bao/Journal of Software, 2013,24(5):915−932 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4381.htm [doi: 10.3724/SP.J.1001.2013.04381]

[16] Kalai A. Efficient pattern-matching with don't cares. In: Proc. of the 13th ACM-SIAM Symp. on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2002. 655−656.

[17] Kucherov G, Rusinowitch M. Matching a set of strings with variable length don't cares. In: Nivat M, *et al*., eds. Proc. of the Theoretical Computer Science. Berlin: Springer-Verlag, 1995. 230−247. [doi: 10.1007/3-540-60044-2_46]

[18] Wu YX, Wang LL, Ren JD, Ding W, Wu XD. Mining sequential patterns with periodic wildcard gaps. Applied Intelligence, 2014, 41(1):99−116. [doi: 10.1007/s10489-013-0499-4]

[19] Myers EW. Approximate matching of network expressions with spacers. Journal of Computational Biology, 2009,3(1):33−51. [doi: 10.1089/cmb.1996.3.33]

[20] Wu YX, Wu XD, Jiang H, Min F. A heuristic algorithm for MPMGOOC. Chinese Journal of Computers, 2011,34(8):1452−1462 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2011.01452]

[21] Lam H, Morchen F, Fradkin D, Calders T. Mining compressing sequential patterns. Statistical Analysis and Data Mining, 2012,7(1): 34−52. [doi: 10.1002/sam.11192]

[22] He D, Wu XD, Zhu XQ. SAIL-APPROX: An efficient on-line algorithm for approximate pattern matching with wildcards and length constraints. In: Proc. of the 2007 IEEE Int'l Conf. on Bioinformatics and Biomedicine. Washington: IEEE Computer Society, 2007. 151−158. [doi:10.1109/BIBM.2007.48]

[23] Ding B, Lo D, Han JW, Khoo SC. Efficient mining of closed repetitive gapped subsequences from a sequence database. In: Proc. of the 2009 IEEE Int'l Conf. on Data Engineering. Washington: IEEE Computer Society, 2009. 1024−1035. [doi: 10.1109/ICDE.2009.104]

[24] Wang HP, Xie F, Hu XG, Li PP, Wu XD. Pattern matching with flexible wildcards and recurring characters. In: Proc. of the 2010 IEEE Int'l Conf. on Granular Computing. Washington: IEEE Computer Society, 2010. 782−786. [doi: 10.1109/GrC.2010.156]

[25] Zhou K. Mining sequential patterns with periodic general gap constraints [MS. Thesis]. Shijiazhuang: Hebei University of Technology, 2014 (in Chinese with English abstract).

[26] Li C, Wang JY. Efficiently mining closed subsequences with gap constraints. In: Proc. of the SIAM Int'l Conf. on Data Mining. 2008. 313−322. [doi: 10.1137/1.9781611972788.28]

附中文参考文献:

[13] 柴欣,贾晓菲,武优西,江贺,吴信东.一般间隙及一次性条件的严格模式匹配.软件学报,2015,26(5):1096−1112. http://www.jos.org.cn/1000-9825/4707.htm [doi: 10.13328/j.cnki.jos.004707]

[15] 武优西,刘亚伟,郭磊,吴信东.子网树求解一般间隙和长度约束严格模式匹配.软件学报,2013,24(5):915−932. http://www.jos.org.cn/1000-9825/4381.htm [doi: 10.3724/SP.J.1001.2013.04381]

[20] 武优西,吴信东,江贺,闵帆.一种求解 MPMGOOC 问题的启发式算法.计算机学报,2011,34(8):1452−1462. [doi: 10.3724/SP.J.1016.2011.01452]

[25] 周坤.一般周期间隙约束的序列模式挖掘[硕士学位论文].石家庄:河北工业大学,2014.

刘慧婷(1978－),女,安徽阜阳人,博士,副教授,CCF 专业会员,主要研究领域为数据挖掘,机器学习.

黄厚柱(1991－),男,硕士,主要研究领域为模式匹配,序列挖掘.

刘志中(1990－),男,硕士,主要研究领域为模式匹配,序列挖掘.

吴信东(1963－),男,博士,教授,博士生导师,主要研究领域为数据挖掘,基于知识的系统,万维网络信息探索.