

基于 SVM 的离合词词义消歧

张振景 李新福 田学东 王 凯

(河北大学计算机科学与技术学院 保定 071000) (河北省机器学习与计算智能重点实验室 保定 071000)

摘 要 离合词词义消歧要解决如何让计算机理解离合词中的歧义词在具体上下文中的含义。针对离合词中歧义词在机器翻译中造成的对照翻译不准确以及在信息检索中无法匹配有效信息等问题,将词义消歧的方法应用于离合词中的歧义词,采用 SVM 模型建立分类器。为了提高离合词词义消歧的正确率,在提取特征时,结合离合词的特点,不仅提取了歧义词上下文中的局部词、局部词性、局部词及词性 3 类特征,还提取了“离”形式的歧义词的中间插入部分的特征;将文本特征转换为特征向量时,对布尔权重法进行了改进,依次固定某种类型特征权重,分别改变另外两种类型特征权重的消歧正确率来验证 3 类特征的消歧效果。实验结果表明,局部词特征、局部词及词性特征对消歧效果的影响高于局部词性特征,且采用不同类型的特征权重与采用相同的权重相比,消歧正确率提高了 1.03%~5.69%。

关键词 离合词, SVM, 词义消歧, 分类器

中图分类号 TP391.1

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2016.2.050

Liheci Word Sense Disambiguation Based on SVM

ZHANG Zhen-jing LI Xin-fu TIAN Xue-dong WANG Kai

(College of Computer Science and Technology, Hebei University, Baoding 071000, China)

(Hebei Key Laboratory of Machine Learning and Computational Intelligence, Baoding 071000, China)

Abstract The task of Liheci word sense disambiguation is to make computers choose the correct sense of a Liheci ambiguous word in a given context. For the problem that a Liheci ambiguous word in machine translation is not accurate and in the information retrieval is unable to match the useful information, a word sense disambiguation method was applied to the Liheci ambiguous words and a classifier model was established using SVM. In order to improve the accuracy of the Liheci word sense disambiguation, it extracts not only local word, local part of speech, local word and part of speech, but also the middle insert part of the separated form as disambiguation features according to the characteristics of Liheci. When the text characteristics was converted to feature vector, we could fixed feature weights of some type in turn and changed the feature weights of the other two types to verify the disambiguation effect of the three kinds of feature, respectively. The results show that the effect of local word feature, local word and part of speech features on disambiguation is higher than local part of speech, and using different types of feature weight compared with the same, disambiguation accuracy increases by 1.03%~5.69%.

Keywords Liheci, SVM, Word sense disambiguation, Classifier

1 引言

离合词是现代汉语中一种常见的特殊语言现象,其“可离可合”的特殊类型虽然保持了词义的整体性,却破坏了词结构的定型性^[1]。随着信息检索、机器翻译等方面的需求迅猛增加,离合词词义研究成为离合词研究的热点。然而,离合词中歧义词又给这一热点问题的解决带来了许多困难,如何在找到歧义词后确定其在特定的上下文语境中的意义,成为离合词词义研究的关键问题。

离合词中的歧义词和普通的歧义词只是结构不同,在相同的上下文语境中,表达的含义是相同的,例如“告状”在《现

代汉语词典》(第 6 版)中有两个不同的意义:①(当事人)请求司法机关审理某一案件;②向某人的上级或长辈诉说自己或别人受到这个人的欺负或不公正的待遇。选取北京大学 CCL 语料库中的两个例句如下:

1)她从不在领导面前告别人的状,只强调自己能力不够。

2)有的小朋友经常去老师那告状,说其他小朋友的缺点。

其中,“告状”是离合词,并且是歧义词。例句 1)表现了离合词“离”的特性;例句 2)表现了离合词“合”的特性,在两个例句中,“告状”的意义是相同的。计算机不仅应该把中间有插入成分的离合词检索出来,而且还应该确定离合词“告状”在这两句话中是两个意义中的哪一个,即对离合词进行词义消歧。

到稿日期:2015-07-16 返修日期:2015-11-30 本文受国家自然科学基金项目(61375075),河北省自然科学基金项目(F2013201134, F2012201020),保定市科学技术研究与发展指导计划项目(15ZR063)资助。

张振景(1990-),女,硕士,主要研究方向为智能图文信息处理, E-mail: 1028756007@qq.com;李新福(1970-),男,博士,教授,主要研究方向为智能图文信息处理;田学东(1963-),男,博士,教授,主要研究方向为模式识别与图像处理、中文信息处理;王 凯(1990-),男,硕士,主要研究方向为智能图文信息处理。

当前词义消歧主要的研究方法有基于规则的方法、基于词典知识的方法、有指导的词义消歧方法和无指导的词义消歧方法^[2]。其中有指导的词义消歧是将词义消歧问题看作分类问题来考虑,将机器学习的分类算法用于词义消歧,而且支持向量机(Support Vector Machine, SVM)方法的分类效果最好^[3,4]。

近些年,许多学者对词义消歧进行了深入的研究。国内的朱靖波^[5]提出了基于对数模型的方法,构造了4种计算模型。但其仅对英语词汇进行了实验,在上下文特征选择时利用了词汇的形态变化、短信信息等特征,这些是汉语所不具备的,不能直接使用该模型对汉语进行词义消歧。苗海和张仰森^[6]将基于规则和基于统计两种词义消歧方法结合起来,解决了词义消歧的稀疏问题,同时提高了词义消歧的准确率。杨陟卓和黄河燕^[7]提出了一种基于异构网络图的词义消歧方法,同时设计并实现了一种基于模拟退火的自动估计各种知识类型关系权重的方法,有效地克服了数据稀疏和知识获取瓶颈的问题。同时,杨陟卓和黄河燕^[8]还采用语言模型对传统的有监督词义消歧模型进行了优化,利用两种模型的优势共同决策歧义词的词义。该模型在训练语料较少时,有效提高了词义消歧的效果。卢志茂^[9]采用了依存分析改进贝叶斯分类模型的有指导消歧方法,有效改善了词义分类的功能。李旭^[10]结合互信息和Z-测试的结果,提出了一种改进的全文无指导的词义消歧模型,有效地提高了汉语词义消歧的准确率。幸运^[11]采用支持向量机方法对汉语进行了词义消歧,填补了利用支持向量机方法对汉语进行词义消歧的空白。鹿文鹏和黄河燕^[12]提出了基于依存适配度的知识自动获取的词义消歧方法,利用了依存句法分析技术的优势,该方法使用WordNet作为语义词典,因此不需要任何的人工标注语料,消歧方法简单、有效。张仰森和黄改娟等人^[13]提出了采用隐最大熵原理构建汉语词义消歧模型,实现了基于义原搭配信息的文本隐性语义特征提取方法。文献^[14]结合《同义词词林》中的语义知识,将歧义词左边词和右边词的语义编码作为消歧特征,并利用贝叶斯分类器进行了词义消歧,消歧准确率有所提升。但以上几项研究只针对普通的歧义词,没有处理汉语中出现频率较高的离合词。国外的 Navigli^[15]利用多语联合分布来计算语义的相关性,该方法采用了词典信息和多语知识库中的语义信息。Navigli^[16]还利用多语知识库 BabbelNet 在多种语言之间进行了基于图的词义消歧。Mitesh^[17]提出了一种带有预测参数的多种语言的词义消歧方法。Cabezas^[18]首先介绍了美国马里兰大学的有监督的词义标记系统,该系统在 SENSEVAL-2 国际评测上对英语、西班牙语和瑞典语进行了测试。但是英文等语言与汉语相比,无论在结构还是在语法方面都有很大区别,Cabezas 并没有利用支持向量机对汉语词义消歧进行研究。Buscaldi^[19]采用支持向量机对多义动词进行消歧,首先提取一定窗口大小下的3类特征,包括多义动词上下文的词特征、词性特征、词和词性特征。其主要缺点是将3类特征看作同等重要,没有考虑特征权重对消歧效果的影响。

为了提高离合词词义消歧的正确率,本文采用 SVM 方法,提取一定窗口大小下的局部词、局部词性、局部词及词性3类特征,同时提取“离”形式的歧义词中间插入部分特征,然后将这些文本特征转换为特征向量。利用十折交叉验证方

法,首先对不同类型的特征在其权重相同的情况下进行实验;然后依次固定一种类型的特征的权重,在分别改变另外两种特征的权重的情况下进行实验;最后观察不同特征权重下消歧正确率的变化,同时比较特征权重相同时的消歧正确率和特征权重不同时的最高正确率。实验表明离合词词义消歧效果能有效提升。

2 SVM 模型

SVM 是建立在统计学习的 VC 维理论和结构风险最小化原则上的,在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势,是一种优秀的机器学习分类算法。支持向量机是典型的两类分类器,当离合词中歧义词有两个词义但两个词义的词性不同时,根据词性判断词义。这里讨论的是把支持向量机用于有两个词义且两个词义的词性相同的情况。

(1) 线性支持向量机

① 线性可分支持向量机

假设线性可分的样本集为 $(x_i, y_i), i=1, 2, \dots, n, x \in R^d, y \in \{+1, -1\}$ 。 x 为样本向量, y 为样本所属类别。假设线性判别函数为 $g(x) = w \cdot x + b$, 最优分类面为 $w \cdot x + b = 0$ 。最优分类面需要满足如下的约束条件:

$$\min \frac{1}{2} \|w\|^2 \quad (1)$$

$$\text{s. t. } y_i [(w \cdot x_i) + b] - 1 \geq 0, i=1, 2, \dots, n$$

为求解上述最优化问题,引入 Lagrange 函数:

$$L(w, b, \alpha) = \frac{1}{2} (w, w) - \sum_{i=1}^n \alpha_i \{y_i [(w \cdot x_i) + b] - 1\} \quad (2)$$

其中, $i=1, 2, \dots, n, \alpha_i$ 为 Lagrange 乘子。

利用 Lagrange 最优化理论,转化为对偶问题来求解:

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{s. t. } \sum_{i=1}^n y_i \alpha_i = 0 \quad (3)$$

$$\alpha_i \geq 0, i=1, 2, \dots, n$$

② 线性不可分支持向量机

对于线性不可分的情况,引入常数 $C > 0$ 和松弛变量 $\zeta_i \geq 0$, 常数 C 用来控制对错分样本的惩罚程度,此时应满足的约束条件为:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \zeta_i$$

$$\text{s. t. } y_i [(w \cdot x_i) + b] - 1 + \zeta_i \geq 0 \quad (4)$$

$$\zeta_i \geq 0, i=1, 2, \dots, n$$

根据对偶理论,得到上述问题的对偶问题为:

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{s. t. } \sum_{i=1}^n y_i \alpha_i = 0 \quad (5)$$

$$0 \leq \alpha_i \leq C, i=1, 2, \dots, n$$

(2) 非线性可分支持向量机

对于非线性可分的情况,采用特征映射的方法,将低维特征空间的数据通过非线性映射函数 $\varphi(x_i)$ 映射到高维特征空间,使样本数据在高维特征空间中可以用线性判别函数分类,但是空间变换会使得样本的维数增加。支持向量机通过采用

5 实验

5.1 实验数据

根据《现代汉语词典》(第6版)共统计出了3748个离合词,构建了离合词表和离合词词义表。其中474个离合词有两个词义,且两个词义的词性相同。虽然只占有离合词的12.65%,但其出现的频率非常高。所选语料全部来自于北京大学的CCL语料库^[23],从中选取了5个有两个词义且两个词义词性相同的离合词。遵循以下的选词原则:1)该离合词出现的频率较高,以保证样本容量较大;2)每个义项出现的最大频率不应高于90%,这样才能保证义项分布尽可能均衡。对获取的初始语料,首先进行离合词标注^[24],在此基础上进行训练测试。表3列出了这5个多义离合词以及选取的样本数量。

表3 5个多义离合词的样本数量

离合词	样本数量/个
告状	736
还原	1082
做梦	1725
赔钱	1080
摔跤	1385

5.2 实验结果与分析

交叉验证是一种重要的统计学方法,在数据集样本数量有限时,它可以用来评估分类学习方法的效果。交叉验证有多种方式,其中多折交叉验证(N-fold Cross Validation)最为常用。交叉验证的折数多少对结果影响并不大,但是通过大量的实验和数据集测试表明,十折交叉验证的效果最好。十折交叉验证的基本思想是将所有的数据集分成10份,轮流将其中的9份作为训练数据,1份作为测试数据进行实验。每次实验都会得出相应的正确率,10次结果的正确率的平均值作为最终词义消歧的正确率。

支持向量机的4种核函数中,最常用的是线性核函数和RBF核函数。线性核函数主要用于线性可分的情形,其参数少、速度快,能达到很好的分类效果。RBF核函数主要用于线性不可分的情形,其参数多,分类效果非常依赖参数。在对离合词进行词义消歧时,特征维数特别高,特征维数远大于样本数。综合复杂程度、速度、正确率等方面,线性核函数都表现出比较好的性能。因此这里选择线性核函数。

首先在3种类型的特征对离合词中歧义词的影响相同的情况下进行实验,得到5个离合词的消歧正确率如表4所列。

表4 5个离合词的消歧结果

离合词	正确率/%
告状	71.55
还原	83.54
做梦	83.29
赔钱	94.08
摔跤	85.14

然后在3种类型的特征对离合词中歧义词的影响不同的情况下进行实验,在这个实验中,对于3种类型的特征权重 α 、 β 、 γ ,其权重之和为1。固定其中一个为0.1,以步长0.2逐渐改变剩余两个,依次进行实验,根据消歧正确率判定3种类型的特征的重要性。

固定局部词特征权重 α 为0.1,以步长0.2分别改变局部

词性特征权重 β 和局部词和词性特征权重 γ 的值,分别对5个离合词进行实验,其消歧正确率如表5所列。

表5 α 为0.1时的消歧正确率(%)

离合词	β		γ		β		γ	
	0.1	0.8	0.3	0.6	0.5	0.4	0.7	0.2
告状	75.94		71.88		67.53		63.97	
还原	87.50		85.30		82.07		77.76	
做梦	83.64		83.26		83.18		82.02	
赔钱	95.42		95.24		94.69		94.42	
摔跤	83.28		82.80		82.58		82.18	

从表5可以看出,固定局部词特征权重 α 为0.1,局部词性特征权重 β 以步长0.2从0.1开始逐渐增大,局部词和词性特征权重 γ 以步长0.2从0.8开始逐渐减小时,“告状”、“还原”、“做梦”、“赔钱”、“摔跤”5个离合词的消歧正确率均逐渐降低。说明在固定局部词特征权重 α 时,局部词性特征权重 β 越大,局部词及词性特征权重 γ 越小,消歧正确率越低,即局部词及词性特征对离合词词义的影响更大,局部词性特征对离合词词义的影响更小。

固定局部词性特征权重 β 为0.1,以步长0.2分别改变局部词特征权重 α 和局部词和词性特征权重 γ 的值,分别对5个离合词进行实验,其消歧正确率如表6所列。

表6 β 为0.1时的消歧正确率(%)

离合词	α		γ		α		γ	
	0.1	0.8	0.3	0.6	0.5	0.4	0.7	0.2
告状	75.94		74.76		74.64		74.19	
还原	87.50		88.98		86.91		86.39	
做梦	83.64		82.86		83.55		82.84	
赔钱	95.22		95.41		95.05		94.77	
摔跤	83.28		83.83		83.71		84.59	

从表6可以看出,固定局部词性特征权重 β 为0.1,局部词特征权重 α 以步长0.2从0.1开始逐渐增大,局部词和词性特征权重 γ 以步长0.2从0.8开始逐渐减小时,“告状”、“还原”、“做梦”、“赔钱”、“摔跤”5个离合词的消歧正确率的变化没有呈现明显的规律性,但是和表5相比可以看出,固定局部词性特征权重 β 为0.1时比固定局部词特征权重 α 为0.1时的消歧正确率更稳定。

固定局部词和词性特征权重 γ 为0.1,以步长0.2分别改变局部词特征权重 α 和局部词性特征权重 β 的值,分别对5个离合词进行实验,其消歧正确率如表7所列。

表7 γ 为0.1时的消歧正确率(%)

离合词	α		β		α		β	
	0.1	0.8	0.3	0.6	0.5	0.4	0.7	0.2
告状	60.40		67.42		70.66		74.68	
还原	73.96		79.87		82.58		85.50	
做梦	81.66		82.26		83.97		84.57	
赔钱	94.35		94.44		94.45		94.49	
摔跤	81.60		82.71		83.42		85.41	

从表7可以看出,固定局部词和词性特征权重 γ 为0.1,局部词特征权重 α 以步长0.2从0.1开始逐渐增大,局部词性特征权重 β 以步长0.2从0.8开始逐渐减小时,“告状”、“还原”、“做梦”、“赔钱”、“摔跤”5个离合词的消歧正确率均逐渐提高。说明在固定局部词及词性特征权重 γ 时,局部词特征权重 α 越大,局部词性特征权重 β 越小,消歧正确率越高,即局部词特征对离合词词义的影响更大,局部词性特征对离合词词义的影响更小。

在将文本特征转换为特征向量时,每种类型特征的权重不同消歧效果也不相同。本文结合离合词的特点对离合词进行词义消歧,将3种类型的特征的权重之和记为1,通过改进布尔权重加权方法,得到了较好的消歧效果。由于目前还没有开放的离合词语义标注语料,因此本文所用语料全部是人工词义标注(本实验室应用语言学硕士)。表8列出了均采用支持向量机模型时的消歧结果。

表8 支持向量机模型下的消歧结果

来源	离合词	加权方法	数据集来源	平均正确率/%
			SemEval-2007	70.5
文献[3]	否	布尔权重	北京大学现代汉语语义标注语料	81.7
文献[4]	否	布尔权重	SemEval-2	74.08
文献[11]	否	布尔权重	SemEval-3	64.91
本文	是	布尔权重	北京大学 CCL 语料	83.52
本文	是	改进布尔权重	北京大学 CCL 语料	86.06

从上述数据可以看出,本文利用改进布尔权重加权方法使离合词词义消歧的正确率提高了2.54%,但无论是离合词还是非离合词的词义消歧都还需深入细致的研究。

结束语 本文采用支持向量机方法对离合词进行词义消歧,对不同类型的特征赋予不同的权重。实验表明局部词特征、局部词及词性特征对离合词词义消歧的影响高于局部词性特征,词义消歧的正确率有所提升。

离合词是现代汉语里一类比较特殊的词,对离合词进行词义消歧还有很多问题有待探讨,例如利用多种分类器进行离合词词义消歧、如何更有效地选取特征等等。本文采用支持向量机方法对离合词词义消歧进行研究,对信息检索、机器翻译等有一定启示意义。

参考文献

[1] Wei Wei. Summary of Chinese Liheci studies[J]. Jinzhou Medical Journal(Social Sciences), 2006, 4(4): 80-83(in Chinese)
魏伟. 汉语离合词研究综述[J]. 锦州医学院学报(社会科学版), 2006, 4(4): 80-83

[2] Zhang Yang-sen. Oriented language resource development of Chinese word sense disambiguation and tagging research[R]. Beijing: Peking University, 2006(in Chinese)
张仰森. 面向语言资源建设的汉语词义消歧与标注方法研究[R]. 北京: 北京大学, 2006

[3] Wu Yun-fang, Wang Miao, Jin Peng, et al. Ensembles of Classifiers for Chinese Word Sense Disambiguation[J]. Journal of Computer Research and Development, 2008, 45(8): 1354-1361(in Chinese)
吴云芳, 王淼, 金澎, 等. 多分类器集成的汉语词义消歧研究[J]. 计算机研究与发展, 2008, 45(8): 1354-1361

[4] Joshi M, Pedersen T, Maclin R. A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain[C]// Proceedings of the Second Indian International Conference on Artificial Intelligence. 2005: 3449-3468

[5] Zhu Jing-bo, Li Hang, Zhang Yue, et al. Automatic word sense disambiguation based on logarithmic model[J]. Journal of Software, 2001, 12(9): 1405-1412(in Chinese)
朱靖波, 李珩, 张跃, 等. 基于对数模型的词义自动消歧[J]. 软件学报, 2001, 12(9): 1405-1412

[6] Miao Hai, Zhang Yang-sen. Research of word sense disambiguation based on combination of rules and statistics[J]. Computer Science, 2013, 40(12): 282-286(in Chinese)
苗海, 张仰森. 规则与统计相结合的词义消歧方法研究[J]. 计算机科学, 2013, 40(12): 282-286

[7] Yang Zhi-zhuo, Huang He-yan. The word sense disambiguation study based on heterogeneous networks[J]. Journal of Computer Research and Development, 2013, 50(2): 437-444(in Chinese)
杨陟卓, 黄河燕. 基于异构关系网络图的词义消歧研究[J]. 计算机研究与发展, 2013, 50(2): 437-444

[8] Yang Zhi-zhuo, Huang He-yan. Supervised WSD model optimization based on language model[J]. Journal of Chinese Information Processing, 2014, 28(1): 19-25(in Chinese)
杨陟卓, 黄河燕. 基于语言模型的有监督词义消歧模型优化研究[J]. 中文信息学报, 2014, 28(1): 19-25

[9] Lu Zhi-mao, Liu Ting, Zhang Gang, et al. Word sense disambiguation based on the dependence analysis improved bayesian[J]. High Technology Letters, 2003, 13(5): 1-7(in Chinese)
卢志茂, 刘挺, 张刚, 等. 基于依存分析改进贝叶斯模型的词义消歧[J]. 高技术通讯, 2003, 13(5): 1-7

[10] Li Xu, Liu Guo-hua, Zhang Dong-ming. An improved unsupervised word sense disambiguation of Chinese[J]. Journal of Automatic, 2010, 36(1): 184-187(in Chinese)
李旭, 刘国华, 张东明. 一种改进的汉语全文无指导词义消歧方法[J]. 自动化学报, 2010, 36(1): 184-187

[11] Xing Yun. Chinese word sense disambiguation research based on support vector machine[D]. Beijing: Peking University, 2006(in Chinese)
幸运. 基于支持向量机的汉语词义消歧研究[D]. 北京: 北京大学, 2006

[12] Lu Wen-peng, Huang He-yan. Word sense disambiguation based on dependency fitness with automatic knowledge acquisition[J]. Journal of Software, 2014, 24(10): 2300-2311(in Chinese)
鹿文鹏, 黄河燕. 基于依存适配度的知识自动获取词义消歧方法[J]. 软件学报, 2014, 24(10): 2300-2311

[13] Zhang Yang-sen, Huang Gai-juan, Sun Wen-jie. Chinese word sense disambiguation based on latent maximum entropy principle[J]. Journal of Chinese Information Processing, 2012, 26(3): 72-78(in Chinese)
张仰森, 黄改娟, 苏文杰. 基于隐最大熵原理的汉语词义消歧方法[J]. 中文信息学报, 2012, 26(3): 72-78

[14] Zhang C X, Sun L R, Gao X Y, et al. Integrate Chinese semantic knowledge into word sense disambiguation[J]. International Journal of Hybrid Information Technology, 2015, 8(4): 105-116

[15] Navigli R, Ponzetto S P. Babel Relate! A joint multilingual approach to computing semantic relatedness[C]// Proceedings of the 26th AAAI Conference on Artificial Intelligence. 2012: 108-114

[16] Navigli R, Ponzetto S P. Joining forces pays off: multi-lingual joint word sense disambiguation[C]// Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012: 1399-1410

[17] Mitesh M. Projecting Parameters for Multilingual Word Sense Disambiguation[C]// Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore, August 2009: 49-53

[18] Cabezas C, Resnik P, Stevens J. Supervised Sense Tagging using Support Vector Machines[C]// Proceedings of Senseval-2, Second International Workshop on Evaluating Word Sense Disambiguation Systems. Toulouse, France: Association for Computational Linguistics, 2001; 59-62

[19] Buscaldi D, Rosso P, Pla F. Verb Sense Disambiguation Using Support Vector Machines; Impact of WordNet-Extracted Features[M]// Computational Linguistics and Intelligent Text Processing. 2006; 192-195

[20] Vapnik V N. 统计学习理论的本质[M]. 张学工, 译. 北京: 清华大学出版社, 2000

[21] Vapnik V N. 统计学习理论[M]. 许建华, 张学工, 译. 北京: 电子

工业出版社, 2004

[22] Li Gang, Kou Guang-zeng, Xia Chen-xi, et al. The optimal boundary issues of Chinese word sense disambiguation[J]. New Technology of Library and Information Service, 2009(2): 54-58 (in Chinese)

李纲, 寇广增, 夏晨曦, 等. 中文词义消歧上下文最优边界问题研究[J]. 现代图书情报技术, 2009(2): 54-58

[23] http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xian-dai

[24] Li X F, Liu B, Tian X D. Research on Automatic Recognition of Separable Words in Modern Chinese[J]. Applied Mechanics, and Manufacturing IV, 2014, 8(4): 1493-1498

(上接第 238 页)

敌方目标与我方目标群的位置关系, 图 4(b)为通过对敌方目标空间编群, 形成敌方 4 个空间群与我目标群的空间关系, 图 4(c)为敌方空间群对我方不同目标的进攻意图而形成的相互作用群。

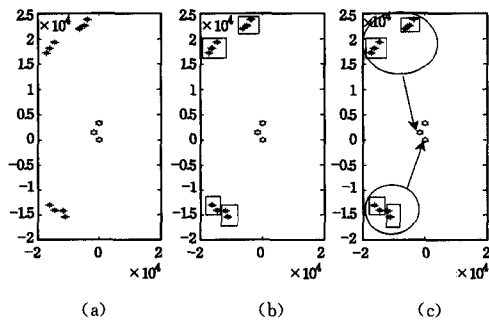


图 4 敌方目标分群过程

按照文献[10]的算法计算, 在同样场景下仿真, 可以得出空间编群的结果一致, 而相互作用群划分时, 按照其功能类型互补、视线夹角小于 130° 等划分原则, 对敌相互作用群的划分时, 敌同一空间群在我方不同空间群时, 其处于不同相互作用群。显然这样的结果无法有效地在敌多目标群对我多目标群进攻时实现对敌目标编群, 无法满足我海空编队作战时对敌方态势的有效估计。

结束语 针对敌方多目标对我多目标进攻时, 对敌目标分群的问题, 本文采用两步来实现。首先改进 chameleon 算法实现空间群的聚类, 不但具有发现和构造任意大小空间群的能力, 实际目标属性的约束条件也使得聚类准确, 运算更加简化。其次根据敌空间群的几何要素, 通过计算进攻要素优势函数, 对各进攻要素优势函数值线性加权, 获得进攻矩阵, 通过对进攻矩阵的分析, 实现对敌目标相互作用群的划分。最后通过场景设置和仿真对比分析, 验证了本算法的有效性。

参 考 文 献

[1] Wang Xiao-fan. Research on situation assessment techniques in information fusion[D]. Xi'an: Xidian University, 2012 (in Chinese)

王晓帆. 信息融合中的态势评估技术研究[D]. 西安: 西安电子科技大学, 2012

[2] Long Zhen-zhen, Zhang Ce, Wang Wei-ping, et al. An unsupervised method for multi-attribute target grouping based on clus-

tering analysis[J]. Mathematics in practice and theory, 2009, 39(19): 157-161 (in Chinese)

龙真真, 张策, 王维平, 等. 基于聚类分析的非监督多属性目标分群方法[J]. 数学的实践与认识, 2009, 39(19): 157-161

[3] Zhang Ming-yuan. Research of technologies for situation assessment in multi-source data fusion[D]. Xi'an: Xidian University, 2004 (in Chinese)

张明远. 多源数据融合系统中态势评估技术的研究[D]. 西安: 西安电子科技大学, 2004

[4] Long Zhen-zhen, Zhang Ce, Wang Wei-ping. Algorithm for operational object grouping in situation assessment based on hierarchical clustering[J]. Journal of projectiles, rockets, missiles and guidance, 2009, 29(3): 209-221 (in Chinese)

龙真真, 张策, 王维平. 基于层次聚类态势估计中的目标分群算法[J]. 弹箭与制导学报, 2009, 29(3): 209-221

[5] Li Yun, Lao Song-yang, Liu Gang, et al. Maximum entropy object grouping algorithm based on dissipative structure theory [J]. System engineering-theory and practice, 2012, 32(12): 2816-2824 (in Chinese)

李贽, 老松杨, 刘钢, 等. 基于耗散结构理论的极大熵目标分群算法[J]. 系统工程理论与实践, 2012, 32(12): 2816-2824

[6] Han Jia-wei, Kamber M. 数据挖掘概念与技术[M]. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2001

[7] Huang Lei, Guo Lei. Method of clustering about situation assessment [J]. Computer Applications, 2006, 26(5): 26-29 (in Chinese)

黄雷, 郭雷. 一种面向态势估计中分群问题聚类方法[J]. 计算机应用, 2006, 26(5): 26-29

[8] Lin Qi-ning. Decision analysis[M]. Beijing: Beijing University of Posts and Telecommunications Press, 2005 (in Chinese)

林齐宁. 决策分析[M]. 北京: 北京邮电大学出版社, 2005

[9] Wang Chang-qing, Zhang Yi-nong, Xu Wan-li. Research of determining index weights based on Least squares method in post-evaluation process[J]. Journal of Jilin University (Information Science Edition), 2010, 28(5): 513-518 (in Chinese)

王长青, 张一农, 许万里. 运用最小二乘法确定后评估指标权重的方法[J]. 吉林大学学报(信息科学版), 2010, 28(5): 513-518

[10] Wang Xin-wei, Yang Shao-qing, Lin Hong-wen, et al. Target Clustering Technology in Sea Battlefield [J]. Ship Electronic Engineering, 2013, 33(11): 25-27 (in Chinese)

王新为, 杨绍清, 林洪文, 等. 海战场目标分群技术研究[J]. 舰船电子工程, 2013, 33(11): 25-27