

# 会议中群体交互语义的获取方法研究

樊祥超 於志文 马 荟

(西北工业大学计算机学院 西安 710129)

**摘 要** 会议是人们日常生活中不可缺少的重要活动,是解决问题、交换信息、共享和创造知识的重要途径,因此智能会议系统是当前学术界和产业界研究的热点之一。但当前的智能会议系统主要集中于物理交互的识别与可视化研究上,对会议中群体语义交互的研究相对较少。群体语义交互是指参会人员针对当前主题所做出的具备语义的交互活动。采用朴素贝叶斯模型,通过对会话中的头部动作、关注度、语气、说话时间长度、交互时机、上次会话角色类型和会话关键词等 7 个特征属性进行处理,设计并实现了一种会议中群体交互语义的获取方法。实验表明,利用该算法,群体交互语义的识别准确率可以达到 80.1%,该算法具有一定的有效性。

**关键词** 会议,朴素贝叶斯,群体交互语义

**中图分类号** TP301.6 **文献标识码** A

## Capturing Human Interaction Semantics in Meetings

FAN Xiang-chao YU Zhi-wen MA Hui

(School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China)

**Abstract** Meeting is an important and vital event in our daily life to solve questions, exchange information, share and create knowledge, so smart meeting system is one of the research hotspots in academia and industry. Current smart meeting systems mainly research on recognition and visualization of physical interaction, and less on human semantic interaction in meetings. Human semantic interaction is interaction activities with semantics which is done by participants with regard to current topic. We designed and implemented a method to capture human interaction semantics in meetings with Naïve Bayes model via dealing with contributes in session, including head gesture, attention from others, speech tone, speaking time, interaction occasion, type of previous interaction and keywords. Experiments show that the recognition accuracy rate of human interaction semantics can be up to 80.1% using this method, which is effective in some degree.

**Keywords** Meeting, Naïve bayes, Human interaction semantics

## 1 引言

会议是人们日常生活中的一项重要活动,用来解决问题、交换信息、共享和创造知识等<sup>[1]</sup>。智能会议系统也引起图像/视频处理、计算机视觉、人机交互和普适计算等诸多研究领域的广泛关注和研究。许多智能会议系统也已经被开发出来,像国外设计的“会议助手”<sup>[2]</sup>,其运用上下文感知技术和可穿戴计算技术,增进与会者与环境以及与会者之间的交互;又如“智能会议空间”<sup>[3]</sup>,通过多模态传感器网络监控参会者的动作并整理记录下来,用于实时的会议监控和会后回顾。智能会议系统也是智能空间的研究领域之一。国内在智能空间的研究方面起步较晚,我国于 2007 年度才正式立项国家“863 计划”,开始智能空间关键技术及原型系统的研究工作<sup>[4]</sup>。但当前的智能会议系统主要支持会议前的准备、会议中的信息交流和会后的内容回顾,更多集中于物理交互的识别与可视化,而对会议中的群体语义交互层面涉及不多。

所谓群体语义交互,是指参会人员针对当前主题所做出的具备语义的交互活动,如询问、提议、评论等。获取群体交互语义能在一定程度上帮助我们掌握各参会人员对会议主题或当前某个特定问题的看法、态度,并能了解一个结论获得的过程。本文针对会议中群体交互语义的获取提出了一种方法:采用朴素贝叶斯算法,处理已获得的会议中群体交互的会话记录,包括会话中的头部动作、关注度、语气、说话时间长度、说话时机、上次会话角色类型和会话关键词等 7 个特征属性,将处理后所得到的结果用于判定未分类的会话记录,获取该次会话中的交互语义。

## 2 朴素贝叶斯分类算法

贝叶斯算法<sup>[5]</sup>是一种基于概率的分类算法。朴素贝叶斯分类算法在贝叶斯算法基础上要求各分类属性之间条件独立,即分类属性之间无任何关联。朴素贝叶斯算法与其他方法相比所具有的优势,首先在于它在接受大数据量训练和查

到稿日期:2010-11-11 返修日期:2011-05-06 本文受国家自然科学基金(60903125),国家 863 高技术研究发展计划基金项目(2009AA011903),教育部“新世纪优秀人才支持计划”(NCET-09-0079),陕西省自然科学基金基础研究计划项目(2010JM8033)资助。

樊祥超(1987-),男,硕士生,主要研究领域为普适计算,E-mail:fanxiangchao@126.com;於志文(1977-),男,博士,教授,主要研究领域为普适计算;马 荟 女,硕士生,主要研究领域为普适计算。

询时所具备的高速度。由于该算法仅仅是针对特征属性的相关概率的数学运算,而且即使选用超大规模的训练集,记录的特征属性也相对有限,因此其能拥有相对较快的处理速度,尤其适用于对响应速度有严格要求的应用和训练量逐渐递增的场合等。其次,该算法在对训练集进行训练的过程中综合发挥了每个分类属性对于最终结果所起的判定作用,从而能够较为准确地为未知分类的样本做出分类。最后,该算法结合先验知识和后验知识,简单实用,同时结果也容易理解。

### 2.1 朴素贝叶斯分类模型

朴素贝叶斯算法的基本思想是根据训练集得到分类的先验概率,然后利用贝叶斯公式获得未知类别样本的后验概率,通过比较属于每一分类的概率大小得到未知类别样本的分类。贝叶斯分类模型如图1所示。

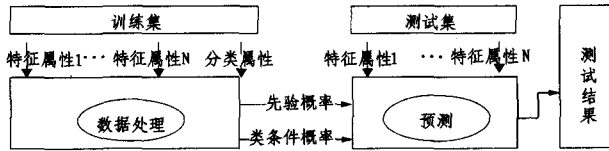


图1 贝叶斯分类模型

首先,该算法能够通过训练集进行训练,即通过学习可以获得每一分类的先验概率和类条件概率。

计算出先验概率和类条件概率之后,就可以利用贝叶斯公式,即式(2),计算出某一未知分类样本属于每一分类的后验概率;比较概率大小,将求出的最大概率对应的分类确定为该未知样本的分类。

朴素贝叶斯算法各分类属性之间条件独立,用数学公式可表述如下:

$$P(X|Y=y) = \prod_{i=1}^d P(X_i|Y=y) \quad (1)$$

式中, $y$ 为分类标号,每个属性集  $X = \{X_1, X_2, \dots, X_d\}$ 。

这样,朴素贝叶斯公式就变为如下形式:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)} \quad (2)$$

对于所有的  $Y$  而言,  $P(X)$  是固定的,故只要找出使式(2)分子最大的类就能够完成工作,从而确定未知类别样本的分类。

### 2.2 连续属性类条件概率的处理

#### (1) 连续属性离散化

把每个连续的属性离散化,然后用相应的离散区间替换连续属性值,通过计算类  $y$  的训练集中落入  $X_i$  对应区间中的比例来估计条件概率  $P(X_i|Y=y)$ 。

#### (2) 概率分布

假设连续变量服从某种概率分布,用训练集数据估计分布的参数。经常使用到的概率分布是高斯分布:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (3)$$

式中,  $P(X_i = x_i | Y = y_j)$  表示的是对每个分类  $y_j$ , 属性  $X_i$  的类条件概率,  $\mu_{ij}$  和  $\sigma_{ij}^2$  为对应的均值和方差。参数  $\mu_{ij}$  可以用训练集中的样本均值  $\bar{X}$  来表示,  $\sigma_{ij}^2$  可以用训练集的样本方差  $s^2$  来表示。

## 3 会议中群体交互语义的获取

会议中群体交互数据的获取依赖于会议环境中的相关设

备,比如摄像头、麦克风、运动传感器等,通过这些设备可获得会议过程中的会话记录。

本文选取了会话的7个特征属性——头部动作、关注度、语气、说话时间长度、说话时机、上次会话角色类型和会话关键词。通过会议环境中布设的摄像头、麦克风、运动传感器等设备,可以获得会议过程中会话的原始数据,包括视频信息、音频信息和动作信息。然后利用相关软件处理原始数据,提取出会话中特征属性的信息<sup>[1]</sup>;通过视频处理软件及相关算法,能够获取会话中的关注度信息和说话时机;利用语音识别引擎,可以获得每次会话中说话人的语气、说话时间长度以及会话的关键词;头部动作能够通过动作数据处理器提取得到,上次会话角色类型也可以通过上次会话交互的识别结果加以确定。会话角色类型分为提议、评论、确认、询问、同意、反对6种类型。利用朴素贝叶斯模型处理所获得的训练集中的会话记录来获得所需要的参数,进而预测未知语义会话记录的交互语义。

### 3.1 数据处理及算法思想

(1)首先是预处理训练集样本数据(见表1),特别是对“关键字”特征数据的处理,将其转换为算法容易处理的数据。

样本数据包含分类属性“角色类型”和7个特征属性。7个特征属性又分为2类:一类是连续的,包括关注度和说话时间长度(注:这里把关注度当作连续属性处理);一类是离散属性,包括头部动作、语气、说话时机、上次会话角色类型和会话关键词。由于会话关键词特征属性又区别于离散属性,因此会另作处理。

表1 数据预处理

属性	数据表示或转换
会话角色类型	提议—>1,评论—>2,确认—>3,询问—>4,同意—>5,反对—>6
头部动作	点头—>0,摇头—>1,正常—>2
关注度	整型数
语气	疑问—>0,非疑问—>1
说话时间长度	单位为秒
说话时机	回应—>0,自发—>1
上次会话角色类型	提议—>1,评论—>2,确认—>3,询问—>4,同意—>5,反对—>6

对关键词特征属性做如下处理,即凡含有如下单词或词组的会话记录,其关键词属性值就标注为对应的序号值;不能通过关键词确定的,用“0”作为其属性值,内容如表2所列。通过语音识别软件从原始会议数据中提取的是英文关键词(本文处理的是英文关键词)关于中文关键词的处理,需要在未来的工作中进行。

表2 关键词转换

序号	关键词
1	how about, think, let's, what about recommendation, seem, needless, isn't good, don't want, depend, don't know, maybe, make, only, although, simply, not particular, don't choose, once, can't, not, though, should, so that, you can, we can, he can, she can, better, not only, most
2	um, yes, ah, good, OK, uh, huh, oh, well, right
3	can I, do you, how, really, isn't it, what, where, what's
4	exactly, attractive, agree, all right, good, very good
5	no, but, however

注:不能通过关键词确定的,用“0”作为其属性值

(2)根据训练样本计算每种会话角色的先验概率,即属于每种会话角色的记录数占总记录的比例。

(3)针对某一特定会话角色,计算每个特征属性的每个值在该种会话角色中的比例,即类条件概率。针对离散属性,计算其均值即可;对于连续属性,则同时需要计算出每一分类中的均值和方差,并通过式(3)得到该连续属性相对于某一分类的类条件概率。

(4)确定未知分类的分类号。计算每种会话角色的后验概率,即用每种会话角色的先验概率乘以该种会话角色所对应的所有类条件概率,然后比较这些后验概率的大小,并将后验概率最大的那种会话角色确定为某一会话记录的分类。

### 3.2 算法流程

#### (1)算法中用到的变量

算法中用到的变量及其含义如表3所列。

表3 变量及其含义

变量	含义
trainingDataFilepath 和 testingDataFilepath	训练集和测试集文件路径
records	存储样本数据
numOfSession_role[]	统计每类会话角色的记录数
priorProbability[]	存储每类会话角色(session_role)的先验概率
head_actionCP[][]	存储头部动作属性的每个属性值相对于每类会话角色的类条件概率
toneCP[][]	存储语气属性的每个属性值相对于每类会话角色的类条件概率
occasion_of_speakingCP[][]	存储说话时机属性的每个属性值相对于每类会话角色的类条件概率
last_session_roleCP[][]	存储上次会话角色的每个属性值相对于每类会话角色的类条件概率
keywordsTypeCP[][]	存储关键词属性的每个属性值相对于每类会话角色的类条件概率
attention_degreeAVR[]和 attention_degreeVAR[]	存储关注度相对每类会话角色的均值和方差
length_of_speakingAVR[]和 length_of_speakingVAR[]	存储说话长度相对每类会话角色的均值和方差

#### (2)算法流程

首先,根据训练集文件路径 trainingDataFilepath 读取训练集文件,将读取的记录放入 records 中。其中要注意的是关键词属性的属性值变换,变换流程图如图2所示。

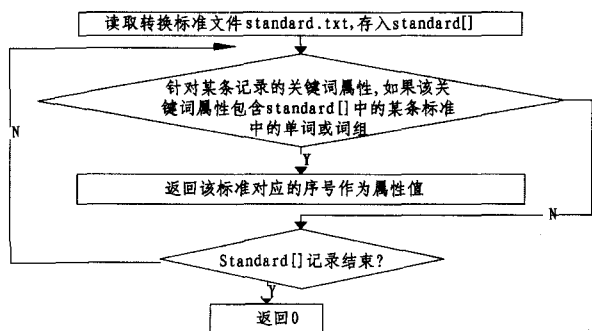


图2 关键词属性值变换图

然后,获取每种会话角色(session\_role)的先验概率。遍历 records 中的所有记录,并且随着遍历的进行, priorProbability[] 在每条记录中获取的该条记录所述会话角色类型]自增1,最后将 priorProbability[] 中的每个值除以记录总数,即得每种会话角色(session\_role)的先验概率。同时将每种角色类型所占数目存储到 numOfSession\_role[] 中。

随后,计算类条件概率。由于样本数据中的数据与所定义的数组变量的下标一致,因此可以用处理后的样本中的数据作为数组变量的下标,以便于统计和计算类条件概率。具

体处理流程如图3所示。

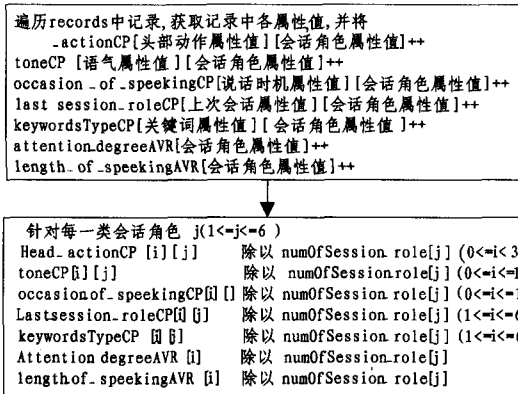


图3 类条件概率处理流程

这样,就得到离散属性的类条件概率以及连续属性的均值。

其次,通过遍历 records 数组和计算方差的公式,可得到关注度和说话时间长度两个特征属性的方差,即 attention\_degreeVAR[j]和 length\_of\_speakingVAR[j](1<=j<=6)。

得到连续属性的均值和方差之后,就可以运用式(1),得到连续分类属性的类条件概率。

接着,针对某一未知分类的记录,根据式(2)计算其属于每种会话角色的概率并进行比较,将所得概率最大的会话角色类型作为该会话记录的分类。

最后,就可以利用通过训练集得到的以上结果(主要为先验概率和类条件概率)对测试集进行测试,并计算测试准确率。

## 4 实验结果分析与讨论

本文选取 161 个会话记录作为训练集,另外 282 个会话记录作为测试集。经过全部特征属性的提取,算法的识别正确率达到 80.1%。

### 4.1 针对特定会话角色的识别结果

针对每种会话角色,计算该会话角色对应的会话记录的正确识别率和错误识别率,错误识别率具体到错误识别为哪种会话角色的概率,如表4所列。

表4 针对不同会话角色的识别结果(%)

	提议	评论	确认	询问	同意	反对
提议	10	90	0	0	0	0
评论	2.2	87.1	5	0	0.7	5
确认	0	14.3	82.9	1.4	1.4	0
询问	5.9	5.9	0	88.2	0	0
同意	5.9	0	29.4	0	58.8	5.9
反对	0	41.7	0	0	8.3	50

从表4可以看出,每种会话角色的测试准确率从高到低依次为询问、评论、确认、同意、反对、提议。测试集中属于“询问”会话角色的会话记录中有 88.2% 测试正确,即通过算法测试确定为“询问”会话角色,识别率最高,但也有 5.9% 的测试为“提议”会话角色、5.9% 的测试为“评论”会话角色;测试集中属于“提议”会话角色的会话记录的识别率最低,只有 10% 的会话记录测试正确,其他 90% 均被错误确定为“评论”会话角色。

### 4.2 单个特征属性的识别结果

仅用一个特征属性来对测试集进行测试,并记录识别准

(下转第 258 页)

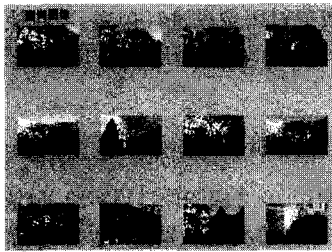


图3 基于 SVM 反馈的检索

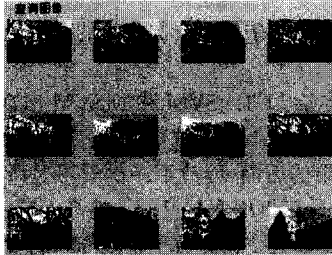


图4 基于 MSVM 反馈的检索

通过对比图3和图4可以看出,经过3次反馈后,MSVM方法比SVM方法有更好的检索效果,图像检索的精确度得到了很大的提高,从而验证了MSVM方法的有效性。

**结束语** SVM方法并不假设样本的分布条件,它基于结构风险最小化原则,对小样本情况下的学习问题给出最优解,并且在样本趋于无穷时可以保持良好的一致收敛性。

MSVM方法是在SVM的基础上,通过记忆功能,用历次反馈的累积样本代替一次反馈样本,从而增加了学习样本数量,减小了查准率的振荡,提高了检索精度。同时为了减轻用户负担,本文提出了记忆性标注。实验证明,MSVM方法可以避免因训练样本集过小而出现的局部最小化的问题,能较为准确地分类图像库中的图像,同时有效地减轻了用户的负担。

## 参考文献

- [1] Rui Y, Huang T S. Optimizing learning in image retrieval[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island; IEEE, 2000; 236-243
- [2] 周建新, 高科, 李锦涛, 等. 图像检索中一种有效的 SVM 相关反馈算法[J]. 计算机辅助设计与图形学学报, 2007, 19(4): 535-540
- [3] Zhon X S, Huang T S. Relevance feedback in image retrieval: a comprehensive review[J]. Multimedia Systems Journal, 2003, 8(6): 536-544
- [4] 正涛, 樊孝忠, 郭剑毅. 基于支持向量机的汉语问句分类研究[J]. 华南理工大学学报: 自然科学版, 2005, 33(9): 25-29
- [5] Hong P, Tian Q, Huang T S. Incorporate support vector machine to content-based image retrieval with relevance feedback [C]// Proceedings of International Conference on Image Processing. Vancouver; IEEE, 2000; 750-753
- [6] <http://sipi.usc.edu/services/database/index.html>

(上接第242页)

-确率,如表5所列。

表5 单个特征属性的识别准确率

特征属性	识别率
头部动作	0.571
关注度	0.521
说话语气	0.603
说话时间长度	0.613
交互时机	0.493
上次会话角色类型	0.390
关键词	0.709

从表5可知,不同特征属性对最后的识别率的影响不同,仅用“关键词”特征属性进行测试所得的测试准确率最高,为70.9%;而仅用“上次会话角色类型”特征属性进行测试所得的测试准确率最低,为39%。识别准确率从高到低依次为关键词、说话时间长度、说话语气、头部动作、关注度、交互时机、上次会话角色类型。

### 4.3 不同特征属性组合的识别结果

表6 不同特征属性组合的识别结果

特征属性组合	识别率
{f1, f2}	57.1%
{f3, f4}	71.6%
{f5, f6}	42.9%
{f1, f2, f3, f4}	72.0%
{f1, f2, f5, f6}	48.2%
{f3, f4, f5, f6}	65.6%
{f1, f2, f7}	70.9%
{f3, f4, f7}	79.1%
{f5, f6, f7}	67.4%
{f1, f2, f3, f4, f5, f6}	67.7%
{f1, f2, f3, f4, f5, f6, f7}	80.1%

注: f1: 头部动作, f2: 关注度, f3: 说话语气, f4: 说话时间长度, f5: 交互时机, f6: 上次会话角色类型, f7: 关键词

将不同的特征属性带入朴素贝叶斯模型,对测试集进行测试,得到的实验结果如表6所列。

从表6可以得到如下结论,不同特征属性的组合所得到的识别准确率有一定的差异,而选取7个特征属性的组合所计算出的识别准确率最高,为80.1%,证明所选取特征属性的合理性。**结束语** 本文从获取会议中群体交互语义的角度出发,针对当前智能会议系统只集中于物理交互研究的情况,采用朴素贝叶斯模型,设计并实现了群体交互语义的获取方法。实验结果表明,该方法在群体交互语义的识别方面能够取得不错的效果,识别准确率达到80.1%。但该方法在处理数据上还有诸多不足之处,有待进一步优化,如会话记录中的关键字处理,只是将训练样本集中的关键字数据做简单分类,并没有涵盖全部。总体而言,该方法基本能达到预期的效果,具有一定的适用性和有效性。

## 参考文献

- [1] Yu Zhi-wen, Yu Zhi-yong, Aoyama H, et al. Capture, Recognition, and Visualization of Human Semantic Interactions in Meetings[C]// Pervasive Computing and Communications 2010 (PerCom 2010). 2010; 107-115
- [2] Dey A K, Salber D, Abowd G D, et al. The Conference Assistant: Combining Context-awareness with Wearable Computing [C]// ISWC'99. 1999; 21-28
- [3] Mikic I, Huang K, Trivedi M. Activity Monitoring and Summarization for an Intelligent Meeting Room[C]// IEEE Workshop on Human Motion. 2000; 107-112
- [4] 余意, 易建强, 赵冬斌. 智能空间研究综述[J]. 计算机科学, 2008, 35(8): 1-3
- [5] (美) Tan Pang-ning, Steinbach M, Kumar V. 数据挖掘导论 [M]. 范明, 范宏建, 等译. 北京: 人民邮电出版社, 2006; 141-145