附件1

**参赛选题与作品要求**

1．本次竞赛采用两种选题方式：自主命题和指定选题。**参与吉林大学预选赛时只需初选题目，拟定设计思路，无需提交正式论文和代码等设计成果。**

2．自主命题方式

（1）采用开放式，但所有参赛题目须围绕“智能机器人”和“可穿戴设备”两个主题，专家组将对参赛作品进行主题审核，如果参赛队伍所报题目与上述主题无关或其内容违反赛事精神和章程，组委会有权要求参赛队伍进行修改或取消其参赛资格。竞赛只接受防御性的题目，不接受任何具有攻击性质或与国家有关法律、法规相违背的题目。

（2）参赛作品可以是软件或硬件。

（3）参赛作品要体现一定的智能性、创新性和实用性。

（4）竞赛现场为选手提供Windows 和 Linux平台，若参赛作品需要其它平台，请选手自带。

（5）凡已公开发布并已获得商业价值的产品不得参赛；凡有知识产权纠纷的作品不得参赛；与企业合作即将对外发布的产品不得参赛。

（6）选手所提交的作品应该是在本届参赛期间所完成的内容。评审时，专家只会对这部分内容进行评价。

3．指定选题方式

指定选题共10个，均为竞赛赞助方从实际出发拟定的与智能技术密切相关的题目，所有指定命题都与业界需求直接相关又能体现出智能特色，在一定程度上体现了业界对智能技术的需求。鼓励各参赛队伍积极选择此方式参赛。

指定命题如下：**（A类：社交媒体应用； B类：手机智能应用；C类： 计算机视觉和自然语言处理应用）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **类别** | **序号** | **题目** | **题目描述** | **要求** | **具体规格说明** | **作品格式** | **评选标准** |
| **A** | 1 | 社交媒体的事件检测 | 对社交媒体中的爆发性事件实现自然语言理解、语义识别、事件关键词识别和跟踪、事件内容和相关人物的聚类 | 1. 能够实现社交媒体中爆发性事件的检测和跟踪 2. 对事件内容、人物进行聚类，根据不同的方面描述事件的发展和演化 3. 能够展示检测和跟踪结果（通过可视化工具对事件的演化进行描述） | 1．典型场景应用说明 2. 关键技术分析 | 1.设计论文  2.算法原型  3.原型系统 | 1.方案设计清晰明确，算法优秀，效率高。 2.事件的识别率高，支持交互式展示。 |
| **A** | 2 | 社交地点分类 | 在社交媒体中，用户之间会有互动，用户也往往会在其访问过的地点进行签到。通过用户带时间的签到信息，用户之间的互动，用户之间的社交网络以及少量的带标注样本，对用户签到过的地点进行分类，例如餐厅、商店等。 | 1. 对社交数据进行抓取和分析，包括用户的关系网络，社交行为数据，签到数据，以及某些地点的类别数据 2. 对多种数据信息进行混合建模，得出准确的分类模型 3. 对未标注的地点进行分类，并分析实验结果 | 1. 提供典型场景的解决方案 2. 提供算法说明与性能评估   参考数据集：  微博API接口  (<http://open.weibo.com/wiki/%E5%BE%AE%E5%8D%9AAPI>) | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确，算法优秀，效率高。  2.如方法新颖，有创新思想和参考价值，作为加分项。  3. 评价指标：分类正确率 |
| **A** | 3 | 网络中VIP发现 | 在社交网络，通信网络中，发现特定条件下的高价值用户；例如进行社交营销的种子用户，有可能使用增值业务的手机用户等。 | 1. 对社交数据进行抓取和分析，包括用户的关系网络，社交行为数据等 2. 通过不同的维度对用户进行排序 3. 对选择的用户进行模拟评估 | 1.提供典型场景的解决方案  2.提供算法说明与性能评估  参考数据集：  异网高端用户发现  (<http://www.shumo.com/home/html/1504.html>) | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确，算法优秀。  2.能够提出恰当的评估方法。  3. 评价指标：F1-Measure (<http://en.wikipedia.org/wiki/F1_score>) |
| **B** | 4 | 手机旅游翻译助手 | 通过手机拍图(印刷体)进行OCR，对相应的文本信息进行翻译 | 1. 图片OCR，接近实时的文档分析 2. 对某一类问题（如菜单）进行翻译（如中翻英） 3. 对翻译结果使用手机进行展示 | 1.提供典型场景的解决方案 2.提供算法说明与性能评估 | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确，算法优秀，效率高。 2.能够实时的展示结果。(实测数据需要考虑现实内容,具体训练和测试样例要同步提交) |
| **B** | 5 | 通过geo-tag照片研究环境变化 | 通过对某一景区的游览照片的分析来判断近年来环境的变化。 | 1. 从Filckr等网站获取带有时间和地点信息的照片（参考mapping the world’s photos论文） 2. 将同一地点的照片按时间排序以后（例如天安门前）观察环境的变化。总体变化和季节性变化。 |  | 1.设计论文 2.算法原型 3.原型系统 | 1.收集的数据量是否能够支持分析  2. 主要以中国城市为主  3. 变化和发现 |
| **B** | 6 | 个性化聊天机器人 | 基于特定人语料库以及公共语料库，构建出模拟特定人聊天模式的机器人；例如利用某一用户的大量微博信息以及聊天记录，并基于已有的公共语料，构建出与此用户聊天模式类似的机器人。 | 1. 能够实现简单的对话，比如问候 2. 能够表现出一定人类的情绪 3. 能够体现出特定人的聊天特点，例如用词，表情使用等 | 1．提供应用场景  2. 关键技术分析  参考数据集：   1. 新闻数据   (<http://www.datatang.com/data/43724>)   1. 词语搭配数据   (<http://www.datatang.com/data/43725>) | 1.设计论文  2.算法原型  3.原型系统 | 1.方案设计清晰明确，算法优秀，效率高。  2.与特定人应答相似度高，支持交互式展示。 |
| **B** | 7 | 预测出租车目的地和抵达时间 | 类似滴滴打车手机软件需要根据每个出租车的起点和中间的一段路程，来预测终点和抵达终点所需的时长，从而可以动态调度出租车资源。 | 1. 收集出租车GPS轨迹数据 2. 地图匹配得到起点和终点 3. 通过起点和路程的GPS轨迹预测终点位置 4. 预测抵达终点位置的时长 | 1. 提供建模算法 2. 提供算法评估结果   参考数据集：  <https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i/rules> | 1.设计方案书  2.算法流程图  3.原型系统 | 1.方案简单可行。  2.预测精度高。 |
| **C** | 8 | 物体自动识别系统 （life-long learning, never ending learning） | 只考虑识别一类物体，比如鼠标。给定一张物体图片（可以只含有一个物体），系统判断该物体是不是目标物体（比如鼠标），用户以语音对话的方式把结果（正确或错误）反馈给系统，系统根据用户反馈更新识别模型。 | 系统通过用户不断的反馈变得越来越智能。 | 关键技术分析 | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确。  2.能够实施的展示结果。  3.训练数据,测试数据需要同步提交。 |
| **C** | 9 | 自动对话系统 | 设计一个自动对话系统，要求可以对用户的输入的任意中文语句给出适当的回复，对内容不做限制，但是要求回复的多样性和适当性（语法和语义上的） | 1. 爬对话的数据，或使用我们发布微博-回复数据 <http://data.noahlab.com.hk/conversation/> 2. 参赛者也可以跳过第一步，实现一个完全基于规则的系统 3. 如果使用基于机器学习的系统，用户需要自己设计算法来从对话数据中训练模型并建立一个简单的对话系统 | 关键技术分析 | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确。  2.能够实施的展示结果。 |
| **C** | 10 | 英文电子邮件写作助手 | 设计一个outlook插件，在用户输入英文的同时，能自动建议若干个提示，提示可以是下一个词，下两个词，下三个词，以及一句完整的英文。提示的结果可以有多个，并且提示的结果可以根据用户的使用情况进行调整，使得排在前面的提示结果被用户选用的可能性大。 | 1. 3到5个提示结果； 2. 用户不能因为使用该插件感觉到明显的滞后； 3. 在我们提供的测试样本上，前三个提示的点击率大于30%。 | 可以使用任何互联网资源 | 1.设计论文 2.算法原型 3.原型系统 | 1.方案设计清晰明确。  2.能够实施的展示结果。 |