



(12) 发明专利

(10) 授权公告号 CN 102364473 B

(45) 授权公告日 2013. 11. 20

(21) 申请号 201110352002. 3

CN 102024056 A, 2011. 04. 20, 全文.

(22) 申请日 2011. 11. 09

审查员 刘洲俊

(73) 专利权人 中国科学院自动化研究所

地址 100190 中国北京市海淀区中关村东路
95 号

(72) 发明人 卢汉清 刘静 李泽超

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021

代理人 王波波

(51) Int. Cl.

G06F 17/30 (2006. 01)

(56) 对比文件

US 2004/0073646 A1, 2004. 04. 15, 全文.

CN 101714145 A, 2010. 05. 26, 全文.

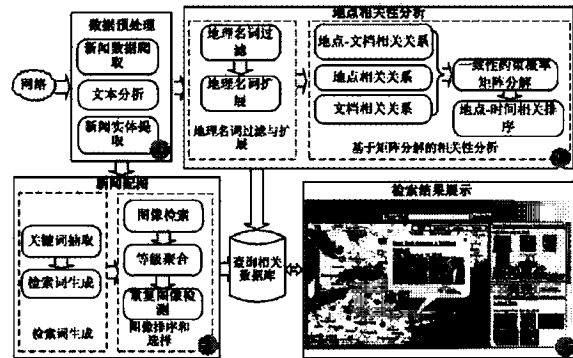
权利要求书3页 说明书11页 附图5页

(54) 发明名称

融合地理信息与视觉信息的网络新闻检索系
统及方法

(57) 摘要

本发明提出了一种融合地理信息与视觉信息的网络新闻检索系统及方法。该系统包括：数据预处理模块，用于爬取新闻数据并进行文本分析和信息数据提取，所述新闻数据包括人物、地点、时间和文本信息；地点相关性分析模块，用于执行新闻事件与新闻地点的相关性分析；新闻配图模块，用于为新闻选择合适的图像；检索结果展示模块，用于基于检索相关性排序展示检索到的新闻。本发明的系统及方法综合利用了地理位置信息和视觉信息对网络新闻进行描述与展示，为网络用户提供基于地理位置的多媒体新闻检索，同时综合了新闻地点-新闻事件的关系、新闻地点的相关性以及新闻事件之间的关系，从而提供给用户一个更生动、更富信息的新闻搜索结果。



1. 一种融合地理信息与视觉信息的网络新闻检索系统,该系统包括:

数据预处理模块,用于爬取新闻数据并进行文本分析和信息数据提取,所述新闻数据包括人物,地点、时间和文本信息;

地点相关性分析模块,用于执行新闻事件与新闻地点的相关性分析;

新闻配图模块,用于为新闻选择能够说明新闻内容的图像;

检索结果展示模块,用于基于检索相关性排序展示检索到的新闻;

所述地点相关性分析模块包括:

地理名词过滤和扩展模块,用于获取地理名词的地理位置信息;

基于矩阵分解的相关性分析模块,用于利用一致性约束概率矩阵分解方法分析新闻地点和新闻事件之间的关系,所述一致性约束概率矩阵分解方法基于下述规则分析新闻地点和新闻事件之间的关系:相似度较高的新闻事件很可能发生在同一个地方,以及相关性较高的多个地点与同一个新闻事件的关系是相近的,所述一致性约束概率矩阵分解方法的目标函数为:

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \delta_{ij} (R_{ij}^0 - R_{ij})^2 + \frac{\lambda_P}{2} \text{Tr}[\mathbf{P}^T \mathbf{P}] + \frac{\lambda_E}{2} \text{Tr}[\mathbf{E}^T \mathbf{E}] \\ + \frac{\lambda_C}{2} \text{Tr}[\mathbf{R}^T \mathbf{L}^C \mathbf{R}] + \frac{\lambda_S}{2} \text{Tr}[\mathbf{R} \mathbf{L}^S \mathbf{R}^T].$$

其中, M 为地点的数量; N 为事件的数量; δ 是标识矩阵,如果 i 和 j 的关系大于零,则 $\delta_{ij}=1$, 否则 $\delta_{ij}=0$; $\mathbf{R}^0 \in \mathcal{R}^{M \times N}$ 是初始的地点-事件的 0-1 关系矩阵; $\mathbf{R} \in \mathcal{R}^{M \times N}$ 表示期望的地点-事件关系矩阵; $\lambda_P = \sigma_R^2 / \sigma_P^2$; $\lambda_E = \sigma_R^2 / \sigma_E^2$; σ_R^2 为地点-事件的关系估计误差所服从的高斯分布的方差; σ_P^2 为潜在的 H 维地名特征矩阵服从高斯分布的方差; σ_E^2 为潜在的 H 维事件特征矩阵服从高斯分布的方差; $\mathbf{P} \in \mathcal{R}^{H \times M}$ 表示潜在的 H 维地名特征矩阵; $\mathbf{E} \in \mathcal{R}^{H \times N}$ 表示潜在的 H 维事件特征矩阵; λ_C 和 λ_S 是两个非负的权重系数; $\mathbf{L}^C = \mathbf{D}^C - \mathbf{C}$ 和 $\mathbf{L}^S = \mathbf{D}^S - \mathbf{S}$ 是拉布拉斯矩阵, \mathbf{D}^C 是对角矩阵,定义为 $D_{ii}^C = \sum_{j=1}^M C_{ij}$, \mathbf{D}^S 是对角矩阵,定义为 $D_{ii}^S = \sum_{j=1}^N S_{ij}$; $\mathbf{C} \in \mathcal{R}^{M \times M}$ 表示地点之间的相关性矩阵; $\mathbf{S} \in \mathcal{R}^{N \times N}$ 表示事件相似性矩阵; $\text{Tr}[\cdot]$ 是矩阵求迹运算; 基于上述模型求解得到 \mathbf{P} 和 \mathbf{E} , 然后采用 $\mathbf{P}^T \mathbf{E}$ 近似 \mathbf{R} ;

所述新闻配图模块包括:

检索词生成模块,用于从新闻数据中抽取一个或多个关键词,将其组合成检索词并提交给图像搜索引擎进行图像检索;

图像排序和选择模块,用于对检索到的图像进行排序和去重,并选择能够说明新闻内容的图像。

2. 根据权利要求 1 所述的系统,所述数据预处理模块包括:

新闻数据爬取模块,用于从新闻网站上爬取新闻文档和对应的新闻图像;

文本分析模块,用于提取出新闻数据的标题、时间、网站、摘要和正文以及对应的网址,提取出新闻图像的网址和图像对应的文本信息;

新闻实体提取模块,从新闻数据中提取出人物,地点和时间。

3. 根据权利要求 1 所述的系统,所述检索结果展示模块包括:
 地图视图模块,用于显示所选择的新闻在地图上的分布位置;
 新闻事件列表模块,用于按照预定的规则排序并显示检索到的新闻事件的列表。
4. 根据权利要求 1 所述的系统,其中
 所述检索词生成模块从新闻数据的多个部分中提取检索词进行图像检索;
 所述图像排序和选择模块采用基于等级聚合的方法对检索到的图像进行排序。
5. 根据权利要求 3 所述的系统,其中所述预定的规则排序包括下述中的一种或多种:
 新闻事件之间的相关性、新闻事件与检索地点之间的相关性以及新闻发生的时间信息。
6. 根据权利要求 3 所述的系统,其中所述地图视图模块,响应于用户输入的检索词或者点击地图上任何一个地点,显示出最相关新闻事件的标题及对应的图像。
7. 一种融合地理信息与视觉信息的网络新闻检索方法,该方法包括:
 数据预处理步骤,用于爬取新闻数据并进行文本分析和信息数据提取,所述新闻数据包括人物,地点、时间和文本信息;
 地点相关性分析步骤,用于执行新闻事件与新闻地点的相关性分析;
 新闻配图步骤,用于为新闻选择能够说明新闻内容的图像;
 检索结果展示步骤,用于基于检索相关性排序展示检索到的新闻;
 所述地点相关性分析步骤包括:
 地理名词过滤和扩展步骤,用于获取地理名词的地理位置信息;
 基于矩阵分解的相关性分析步骤,用于利用一致性约束概率矩阵分解方法分析新闻地点和新闻事件之间的关系,所述一致性约束概率矩阵分解方法基于下述规则分析新闻地点和新闻事件之间的关系:相似度较高的新闻事件很可能发生在同一个地方,以及相关性较高的多个地点与同一个新闻事件的关系是相近的,所述一致性约束概率矩阵分解方法的目标函数为:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \delta_{ij} (R_{ij}^0 - R_{ij})^2 + \frac{\lambda_P}{2} \text{Tr}[\mathbf{P}^T \mathbf{P}] + \frac{\lambda_E}{2} \text{Tr}[\mathbf{E}^T \mathbf{E}] \\ & + \frac{\lambda_C}{2} \text{Tr}[\mathbf{R}^T \mathbf{L}^C \mathbf{R}] + \frac{\lambda_S}{2} \text{Tr}[\mathbf{R} \mathbf{L}^S \mathbf{R}^T]. \end{aligned}$$

其中, M 为地点的数量; N 为事件的数量; δ 是标识矩阵,如果 i 和 j 的关系大于零,则 $\delta_{ij}=1$, 否则 $\delta_{ij}=0$; $\mathbf{R}^0 \in \mathcal{R}^{M \times N}$ 是初始的地点 - 事件的 0-1 关系矩阵; $\mathbf{R} \in \mathcal{R}^{M \times N}$ 表示期望的地点 - 事件关系矩阵; $\lambda_P = \sigma_R^2 / \sigma_P^2$; $\lambda_E = \sigma_R^2 / \sigma_E^2$; σ_R^2 为地点 - 事件的关系估计误差所服从的高斯分布的方差; σ_P^2 为潜在的 H 维地名特征矩阵服从高斯分布的方差; σ_E^2 为潜在的 H 维事件特征矩阵服从高斯分布的方差; $\mathbf{P} \in \mathcal{R}^{H \times M}$ 表示潜在的 H 维地名特征矩阵; $\mathbf{E} \in \mathcal{R}^{H \times N}$ 表示潜在的 H 维事件特征矩阵; λ_C 和 λ_S 是两个非负的权重系数; $\mathbf{L}^C = \mathbf{D}^C - \mathbf{C}$ 和 $\mathbf{L}^S = \mathbf{D}^S - \mathbf{S}$ 是拉布拉斯矩阵, \mathbf{D}^C 是对角矩阵,定义为 $D_{ii}^C = \sum_{j=1}^M C_{ij}$, \mathbf{D}^S 是对角矩阵,定义为 $D_{ii}^S = \sum_{j=1}^N S_{ij}$; $\mathbf{C} \in \mathcal{R}^{M \times M}$ 表示地点之间的相关性矩阵; $\mathbf{S} \in \mathcal{R}^{N \times N}$ 表示事件相似性矩阵; $\text{Tr}[\cdot]$ 是矩阵求迹运算; 基于上述模型求解得到 \mathbf{P} 和 \mathbf{E} , 然后采用 $\mathbf{P}^T \mathbf{E}$ 近似 \mathbf{R} ;

所述新闻配图步骤包括：

检索词生成步骤，用于从新闻数据中抽取一个或多个关键词，将其组合成检索词并提交给图像搜索引擎进行图像检索；

图像排序和选择步骤，用于对检索到的图像进行排序和去重，并选择能够说明新闻内容的图像。

8. 根据权利要求 7 所述的方法，所述数据预处理步骤包括：

新闻数据爬取步骤，用于从新闻网站上爬取新闻文档和对应的新闻图像；

文本分析步骤，用于提取出新闻数据的标题、时间、网站、摘要和正文以及对应的网址，提取出新闻图像的网址和图像对应的文本信息；

新闻实体提取步骤，从新闻数据中提取出人物，地点和时间。

9. 根据权利要求 7 所述的方法，所述检索结果展示步骤包括：

地图视图步骤，用于显示所选择的新闻在地图上的分布位置；

新闻事件列表步骤，用于按照预定的规则排序并显示检索到的新闻事件的列表。

10. 根据权利要求 7 所述的方法，其中

所述检索词生成步骤从新闻数据的多个部分中提取检索词进行图像检索；

所述图像排序和选择步骤采用基于等级聚合的方法对检索到的图像进行排序。

11. 根据权利要求 9 所述的方法，其中所述预定的规则排序包括下述中的一种或多种：新闻事件之间的相关性、新闻事件与检索地点之间的相关性以及新闻发生的时间信息。

12. 根据权利要求 9 所述的方法，其中所述地图视图步骤，响应于用户输入的检索词或者点击地图上任何一个地点，显示出最相关新闻事件的标题及对应的图像。

融合地理信息与视觉信息的网络新闻检索系统及方法

技术领域

[0001] 本发明涉及网络新闻检索领域,特别的,涉及一种融合地理信息与视觉信息的网络新闻检索系统及方法。

背景技术

[0002] 随着信息技术的发展和网络的全球化,在线新闻越来越多并且也越来越受欢迎,日益变成了人们日常生活中获取信息的一种重要途径。人们可以通过一些主要的网络门户网站如雅虎、MSN 或者大型新闻网站如 CNN、AOL 和 MSNBC 获取和浏览新闻。

[0003] 但是,现有技术中的新闻展示方法具有若干不足。

[0004] 例如,现有的新闻展示方法缺乏以地理为基础的组织。有研究表明用户经常优先关注几个特定地点的新闻,比如家乡和工作地点。大部分的大型新闻网站可以根据相关的国家进行组织新闻。用户可以提交一个地点作为检索词检索新闻。但是文档中包含的地理名词经常存在噪声,因而降低了检索的性能。

[0005] 另外,现有的新闻展示方法不包含全面的视觉信息。

[0006] 图 1 显示了现有技术中一篇新闻文档中包含图片个数的分布情况。

[0007] 从图 1 中可以看到,现有技术中大部分的新闻文档没有图片或者包含很少的图片。例如,仅有不到 5% 的新闻文档包含超过一张的图片。

[0008] 通常来说,图片的表现效果胜过千言万语,作为新闻文本的补充,新闻图片能够使用户更快的获取信息。但是,如图 1 所示,现有的新闻文档中包含的图片数很少,因而远远不能满足用户获取信息的全面需求。

发明内容

[0009] 本发明的目的是提供一种融合地理与视觉信息的网络新闻检索系统及方法。根据本发明的系统及方法,能够为用户提供基于地理信息组织的新闻,使用户快速地浏览到所关心地区发明的新闻事件;进而,本发明采用图像信息对文本信息进行补充,使用户能够快速掌握新闻事件的内容。

[0010] 根据本发明的一个方面,提供了一种融合地理信息与视觉信息的网络新闻检索系统,该系统包括:数据预处理模块,用于爬取新闻数据并进行文本分析和信息数据提取,所述新闻数据包括人物,地点、时间和文本信息;地点相关性分析模块,用于执行新闻事件与新闻地点的相关性分析;新闻配图模块,用于为新闻选择合适的图像;检索结果展示模块,用于基于检索相关性排序展示检索到的新闻。

[0011] 其中,所述数据预处理模块包括:新闻数据爬取模块,用于从新闻网站上爬取新闻文档和对应的新闻图像;文本分析模块,用于提取出新闻数据的标题、时间、网站、摘要和正文以及对应的网址,提取出新闻图像的网址和图像对应的文本信息;新闻实体提取模块,从新闻数据中提取出人物,地点和时间。

[0012] 所述地点相关性分析模块包括:地理名词过滤和扩展模块,用于获取地理名词的

地理位置信息；基于矩阵分解的相关性分析模块，用于利用一致性约束概率矩阵分解方法分析新闻地点和新闻事件之间的关系。

[0013] 所述新闻配图模块包括：检索词生成模块，用于从新闻数据中抽取一个或多个关键词，将其组合成检索词并提交给图像搜索引擎进行图像检索；图像排序和选择模块，用于对检索到的图像进行排序和去重，并选择能够表达新闻文档内容的图像。

[0014] 所述检索结果展示模块包括：地图视图模块，用于显示所选择的新闻在地图上的分布位置；新闻事件列表模块，用于按照预定的规则排序并显示检索到的新闻事件的列表。

[0015] 在地点相关性分析模块中，所述一致性约束概率矩阵分解方法基于下述规则分析新闻地点和新闻事件之间的关系：相似度较高的新闻事件很可能发生在同一个地方，以及相关较高的多个地点与同一个新闻事件的关系是相近的。

[0016] 在新闻配图模块中，所述检索词生成模块是从新闻数据的多个部分中提取检索词进行图像检索；所述图像排序和选择模块采用基于等级聚合的方法对检索到的图像进行排序。

[0017] 在检索结果展示模块中，所述地图视图模块响应于用户输入的检索词或者点击地图上任何一个地点，显示出最相关新闻事件的标题及对应的图像；所述预定的规则排序包括下述中的一种或多种：新闻事件之间的相关性、新闻事件与检索地点之间的相关性以及新闻发生的时间信息。

[0018] 根据本发明的另一个方面，还提供了一种融合地理信息与视觉信息的网络新闻检索方法，该方法包括：数据预处理步骤，用于爬取新闻数据并进行文本分析和信息数据提取，所述新闻数据包括人物、地点、时间和文本信息；地点相关性分析步骤，用于执行新闻事件与新闻地点的相关性分析；新闻配图步骤，用于为新闻选择合适的图像；检索结果展示步骤，用于基于检索相关性排序展示检索到的新闻。

[0019] 如上所述，本发明的系统及方法，提供了基于地理位置的新闻检索，估计和细化了新闻文档与地理位置之间的关系，在此过程中综合考虑了新闻文档与地点的初始关系，地点之间的关系以及新闻文档之间的相似性。此外，通过提出给新闻文档配图的方法，使得使用户能够更直接和更快地获取新闻。此外，还提出了支持新闻地理名词的检索模块，以及通过点击地图上相应地理位置进行检索的用户检索与浏览界面。

[0020] 根据本发明提出的一致性约束概率矩阵分解模型，能够将地点 - 事件关系，地点之间相关性和事件之间的相似性融合起来，估计和细化地点与事件之间的相关性，能够去除噪声和估计出潜在的关系。

[0021] 根据本发明提出的从文本中抽取检索词进行网络图像检索和图像排序的方法，可以按照多种规则实现对新闻文档的精确配图。

[0022] 根据本发明提出的组合检索词的方法，能够避免目前网络搜索引擎不能够处理复杂检索的问题以及单个词作为检索不能表达文档内容的问题。

[0023] 此外，对于从网络搜索引擎中得到的不同的图像列表，本发明还提出了基于等级聚合的方法对这些图像列表进行融合排序的，从而选出最能表达新闻文档内容的图像。

[0024] 根据本发明提出的新闻文档排序方法，综合考虑了新闻的时效性、重要性以及检索相关性。该方法基于传统的马尔科夫随机游走模型，将前面分析得到的新闻事件 - 新闻地点相关性与新闻文档的时效性线性融合为该模型中的初始状态，并基于新闻文档之间的

相似性,实现新闻文档集合的检索相关性排序。

[0025] 本发明还提供便于用户检索和浏览新闻的交互界面。用户可以通过提交检索或者单击地图进行检索,同时为一个检索结果提供了标题、新闻图片和内容摘要,用户可以快速生动地获取需要的信息。如果用户想要了解更详细的信息,可通过点击界面的“更多”按钮来获取。

[0026] 综上所述,本发明以新闻的发生地点名称为搜索关键词或者通过点击地图上感兴趣的地理位置,提供给用户一个更生动、更富信息的新闻搜索结果,其结果展示界面包含两部分:其一,在真实地图上按照事件所发生地理位置来展示与检索地点最相关的新闻标题和图片信息;其二,具有新闻标题、图片和简短说明的多模态检索结果列表。

附图说明

[0027] 图 1 显示了现有技术中一篇新闻文档中包含图片个数的分布情况;

[0028] 图 2 显示了本发明的网络新闻检索系统的示意图;

[0029] 图 3 是本发明提出的一致性约束概率矩阵分解模型图;

[0030] 图 4 显示了本发明中一个新闻配图的实例;

[0031] 图 5 显示了本发明一个实施例的网络新闻检索和浏览界面;

[0032] 图 6 是 BM25 排序模型、概率矩阵分解模型和一致性约束概率矩阵分解模型的检索性能评价结果;

[0033] 图 7 给出了变化参数在 NDCG@50 规则下的结果;

[0034] 图 8 是本发明的新闻配图方法与现有技术的性能比较结果;

[0035] 图 9 显示了本发明的检索结果排序方法与现有技术的排序方法在检索相关性上的比较结果;

[0036] 图 10 显示了本发明的检索结果排序方法与现有技术的排序方法在时效性方面的比较结果。

具体实施方式

[0037] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实例,并参照附图,对本发明进一步详细说明。虽然本发明的实例是基于英文新闻提供的,但是本发明的方法不受语言种类的限制。

[0038] 本发明提出了一个利用计算机的基于多媒体分析的新闻检索系统,综合利用了地理信息和视觉信息。首先从文本中提取出新闻地点候选集,利用网络信息进行过滤和扩展,并获取其地理位置信息(经纬度)。通过本发明提出的基于一致性约束概率矩阵分解关系挖掘技术发现潜在的新闻地点与新闻事件之间的关系,综合考虑了新闻地点之间的相关性,新闻事件的相似性和新闻地点-新闻事件之间的初始关系。然后为了使用户快速生动地获取新闻,本发明提出了给新闻配图的方法。

[0039] 虽然目前的新闻文档中也包含了新闻图片,但是对应的新闻图片太少了甚至一半以上的文档还是没有图片的,如图 1 所示。本发明提出的相关方法能够给文档提供多张具有表现力的图片。对检索结果,本发明提出了考虑时间信息的基于网页排序的排序方法。设计了一个方便用户的新闻检索与浏览界面。

[0040] 图 2 显示了本发明的网络新闻检索系统的示意图。

[0041] 如图 2 所示,本发明的融合地理信息和视觉信息的新闻检索系统包括数据预处理模块、地点相关性分析模块、新闻配图模块及检索结果展示模块。

[0042] 数据预处理模块用于爬取新闻数据并进行文本分析和信息数据提取,所述新闻数据包括人物,地点、时间和文本信息。所述数据预处理模块包括新闻数据爬取模块、文本分析模块和新闻实体提取模块等子模块,其中:

[0043] 新闻数据爬取模块采用网络爬虫从新闻网站(例如,ABC、BBC、CNN 及谷歌等新闻网站)上爬取新闻文档和对应的新闻图像。

[0044] 文本分析模块采用自然语言处理技术提取出新闻文档的标题、时间、网站、摘要和正文以及对应的网址,提取出新闻图像的网址和图像对应的文本信息。

[0045] 新闻实体提取模块采用自然语言处理技术去除重复文档和从新闻文档中提取出人物,地点和时间。

[0046] 地点相关性分析模块,用于执行新闻事件与新闻地点的相关性分析。地点相关性分析模块包括地理名词过滤和扩展模块、基于矩阵分解的相关性分析模块等子模块,其中:

[0047] 地理名词过滤和扩展模块用于获取地理名词的地理位置信息(例如地理经纬度)。

[0048] 基于矩阵分解的相关性分析模块采用本发明的一致性约束概率矩阵分解方法分析新闻地点和新闻事件之间的关系。

[0049] 新闻配图模块用于为新闻选择能够说明新闻内容的图像。新闻配图模块包括检索词生成模块、图像排序和选择模块等子模块,其中:

[0050] 检索词生成模块,用于从新闻数据中抽取一个或多个关键词,将其组合成检索词并提交给图像搜索引擎进行图像检索,即从新闻文档中抽取一个或多个关键词,将其组合成长度不同的检索词,提交给图像搜索引擎(例如谷歌)进行图像检索。

[0051] 本实施例中,可以利用新闻数据的各个部分(标题、摘要和正文等)的不同重要性,从新闻中提取检索词进行网络图像检索,从而解决目前图像搜索引擎不能处理长检索词的问题以及单个检索词无法表达文档内容的问题。

[0052] 图像排序和选择模块,用于对检索到的图像进行排序和去重,并选择合适的图像。该模块执行图像排序和选择即考虑图像在返回列表中的位置以及与源文档包含图片的相似度,采用等级聚合方法学习不同长度的检索词对应列表的权重,再利用这些权重对图像进行排序,并去除重复图像,然后选择能够表达新闻内容的图像。

[0053] 检索结果展示模块用于基于检索相关性排序展示检索到的新闻。本发明提供了一个展示检索结果的用户界面。如图 2 所示,检索结果展示模块包括地图视图模块和新闻事件列表模块等子模块,其中:

[0054] 地图视图模块显示所选择的新闻在地图上的分布位置。

[0055] 新闻事件列表模块用于按照预定的规则排序并显示检索到的新闻事件的列表。

[0056] 如图 2 所示,用户可以在检索框中输入检索词进行检索,也可以浏览地图点击想要检索的地点在地图上对应的位置,系统自动返回相关结果。

[0057] 本发明的检索结果展示模块综合考虑了新闻的时效性、与检索的相关性以及新闻

的重要性。

[0058] 如图 2 所示,在地图上展示了最相关新闻的标题和前两幅图像。在右部分的列表中,每个新闻显示了标题、相关图像和简短摘要。更多的信息可通过点击“更多”按钮获取。

[0059] 以上介绍了本发明的新闻检索系统的结构组成,如图 2 所示,与该新闻检索系统的各个模块对应,本发明还提出了融合地理信息与视觉信息的网络新闻检索方法,该方法包括下述步骤:数据预处理步骤,用于爬取新闻数据并进行文本分析和信息数据提取,所述新闻数据包括人物,地点、时间和文本信息;地点相关性分析步骤,用于执行新闻事件与新闻地点的相关性分析;新闻配图步骤,用于为新闻选择合适的图像;检索结果展示步骤,用于基于检索相关性排序展示检索到的新闻。

[0060] 所述地点相关性分析步骤包括:地理名词过滤和扩展步骤,用于获取地理名词的地理位置信息;基于矩阵分解的相关性分析步骤,用于利用一致性约束概率矩阵分解方法分析新闻地点和新闻事件之间的关系。

[0061] 优选的,所述一致性约束概率矩阵分解方法基于下述规则分析新闻地点和新闻事件之间的关系:相似度较高的新闻事件很可能发生在同一个地方,以及相关性较高的多个地点与同一个新闻事件的关系是相近的。

[0062] 所述新闻配图步骤包括:检索词生成步骤,用于从新闻数据中抽取一个或多个关键词,将其组合成检索词并提交给图像搜索引擎进行图像检索;图像排序和选择步骤,用于对检索到的图像进行排序和去重,并选择合适的图像。

[0063] 优选的,所述检索词生成步骤从新闻数据的多个部分中提取检索词进行图像检索;所述图像排序和选择步骤采用基于等级聚合的方法对检索到的图像进行排序。

[0064] 所述检索结果展示步骤包括:地图视图步骤,用于显示所选择的新闻在地图上的分布位置;新闻事件列表步骤,用于按照预定的规则排序并显示检索到的新闻事件的列表。

[0065] 其中,所述预定的规则排序包括下述中的一种或多种:新闻事件之间的相关性、新闻事件与检索地点之间的相关性以及新闻发生的时间信息。

[0066] 其中,所述地图视图步骤,响应于用户输入的检索词或者点击地图上任何一个地点,显示出最相关新闻事件的标题及对应的图像。

[0067] 如上所述,该新闻检索系统采用了以下 4 个主要处理流程:(1) 基于一致性约束概率矩阵分解模型的地理位置相关性分析;(2) 新闻配图;(3) 检索结果的排序;(4) 检索结果浏览界面。

[0068] 下面以英文检索为例分别介绍上述子流程,主要包括新闻文档定位流程和配图流程,以及对用户检索结果的排序流程。但显然,本发明不限制于此,而是可以合理的应用于其他语言种类,例如中文等。

[0069] < 地理位置 - 新闻事件的相关性分析 >

[0070] 该流程包括四个步骤:(1) 候选地名提取;(2) 取出候选地名奇异性;(3) 得到地名与文档的初始关系;(4) 是对地名与文档的关系进行细化分析。

[0071] 首先,根据利用自然语言处理技术从数据库的新闻文档中抽取出新闻地名,得到了候选地名列表;然后将利用已得到的列表提交到维基百科,如果返回的页面中没有地理信息则认为这个候选地名是噪声并去除。

[0072] 然后,对于不同的地方对应同一个名字的情况,可以将过滤之后的列表再提交给

地理信息系统 (GeoNames) 进行扩展并爬取对应的地理信息 (经纬度)。

[0073] 随后,统计各地名在数据库新闻文档中的出现频率,从而得到初始的地名与文档的关系。

[0074] 根据上述方法得到的地名与文档的对应关系是包含噪声的,例如,一个关于名人婚礼的新闻会有婚礼举行地点以及新娘和新郎的家乡。婚礼地点才是该新闻真正的地点,也是最相关的,其他地点是不相关的。另外与新闻相关的地点可能没有出现在文章中,比如关于北京奥运会的新闻,介绍了相关的体育事件但没提及北京,然而与此新闻最相关的地点是北京。

[0075] 因此,为了更好地挖掘新闻地点与新闻事件之间的关系(新闻事件与的新闻文档是一一对应关系,即认为一个新闻文档描述了一个新闻事件),本发明基于传统的概率矩阵分解(Probabilistic Matrix Factorization, PMF)模型(参见 Ruslan Salakhutdinov and Andriy Mnih. “Probabilistic Matrix Factorization”, NIPS 2008.)提出了一致性约束概率矩阵分解模型(Consistent Constraints Probabilistic Matrix Factorization, 简称为 CCPMF),对新闻地点与新文档的相关性进行分析。相比传统的 PMF 模型,本发明的 CCPMF 模型引入了新闻文档与地点之间的一致相关性作为优化求解的约束条件(即相关的新闻文档需要对应相关的地点,反之亦然),从而能够更有效的分析二者之间的真实相关性。其中,地点相关性是利用搜索引擎(例如谷歌距离)计算各地名之间的统计共生相关性;新闻文档之间的相关性是通过考虑新闻标题、摘要和正文的不同重要性,进行线性组合计算而得的文本相似性。下面详细讲述前述第四个步骤:地名与文档的关系细化分析。

[0076] 图 3 是本发明提出的一致性约束概率矩阵分解模型图。

[0077] 如图 3 所示,考虑到非常相关的事件很有可能发生在同一个地点以及同一个新闻可能与非常相关的几个地名相关,本发明的一致性约束概率矩阵分解模型综合考虑了地名之间的相关性、文档之间的相似性以及地名与文档的关系。

[0078] 假设具有 M 个地点, N 个事件。 $\mathbf{R} \in \mathcal{R}^{M \times N}$ 、 $\mathbf{C} \in \mathcal{R}^{M \times M}$ 和 $\mathbf{S} \in \mathcal{R}^{N \times N}$ 分别表示地点-事件关系矩阵、地点之间的相关性矩阵和事件相似性矩阵。采用矩阵分解的思想挖掘潜在的高质量特征空间,即采用 $\mathbf{P}\mathbf{E}$ 近似 \mathbf{R} , 其中 $\mathbf{P} \in \mathcal{R}^{H \times M}$ 和 $\mathbf{E} \in \mathcal{R}^{H \times N}$ 表示潜在的 H 维地名和事件特征矩阵。 $\mathbf{R}^0 \in \mathcal{R}^{M \times N}$ 是初始的地点-事件的 0-1 关系矩阵。在概率矩阵分解模型中,假设对地点-事件的关系估计误差从均值为 0、方差为 σ_R^2 的高斯分布,则有:

[0079]

$$p(\mathbf{R}^0 | \mathbf{P}, \mathbf{E}, \sigma_R^2) = \prod_{i=1}^M \prod_{j=1}^N [\mathcal{N}(R_{ij}^0 | \mathbf{p}_i^T \mathbf{e}_j, \sigma_R^2)]^{\delta_{ij}}$$

[0080] 其中 $\mathcal{N}(x | \mu, \sigma^2)$ 表示均值为 μ 、方差为 σ^2 的高斯分布函数。 \mathbf{p}_i 和 \mathbf{e}_j 分别是矩阵 \mathbf{P} 和 \mathbf{E} 的第 i 列和第 j 列。 δ 是标识矩阵,如果 i 和 j 的关系大于零,则 $\delta_{ij} = 1$, 否则 $\delta_{ij} = 0$ 。另外,假设潜在特征空间和系数矩阵服从球形高斯分布,即:

[0081]

$$p(\mathbf{P} | \sigma_P^2) = \prod_{i=1}^M \mathcal{N}(\mathbf{p}_i | 0, \sigma_P^2 \mathbf{I})$$

[0082]

$$p(\mathbf{E}|\sigma_E^2) = \prod_{j=1}^N \mathcal{N}(\mathbf{e}_j|0, \sigma_E^2 \mathbf{I})$$

[0083] 其中 \mathbf{I} 是单位矩阵。经过简单的贝叶斯推导和取对数操作,得到目标函数:

[0084]

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \delta_{ij}^R (R_{ij}^0 - R_{ij})^2 + \frac{\lambda_P}{2} \text{Tr}[\mathbf{P}^T \mathbf{P}] + \frac{\lambda_E}{2} \text{Tr}[\mathbf{E}^T \mathbf{E}]$$

[0085] 其中 $\lambda_P = \sigma_R^2 / \sigma_P^2$, 以及 $\lambda_E = \sigma_R^2 / \sigma_E^2$ 。

[0086] 考虑到地名之间的关系以及文档之间的关系,一致性约束概率矩阵分解模型在概率矩阵分解模型基础上加了两个一致性约束,得到对应的目标函数:

[0087]

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \delta_{ij} (R_{ij}^0 - R_{ij})^2 + \frac{\lambda_P}{2} \text{Tr}[\mathbf{P}^T \mathbf{P}]$$

$$+ \frac{\lambda_E}{2} \text{Tr}[\mathbf{E}^T \mathbf{E}] + \frac{\lambda_C}{2} F_C(\mathbf{R}) + \frac{\lambda_S}{2} F_S(\mathbf{R})$$

[0089] 其中 λ_C 和 λ_S 是两个非负的权重系数,是表示地名相关性和文档相似性之间的均衡参数。 $F_C(\mathbf{R})$ 和 $F_S(\mathbf{R})$ 分别考虑了地名的相互关系和文档的相互关系,定义为:

$$[0090] \quad F_C(\mathbf{R}) = \frac{1}{2} \sum_{k=1}^N \sum_{i,j=1}^M (R_{ik} - R_{jk})^2 C_{ij} = \text{Tr}[\mathbf{R}^T \mathbf{L}^C \mathbf{R}]$$

$$[0091] \quad F_S(\mathbf{R}) = \frac{1}{2} \sum_{k=1}^M \sum_{i,j=1}^N (R_{ki} - R_{kj})^2 S_{ij} = \text{Tr}[\mathbf{R} \mathbf{L}^S \mathbf{R}^T]$$

[0092] 其中 $\mathbf{L}^C = \mathbf{D}^C - \mathbf{C}$ 和 $\mathbf{L}^S = \mathbf{D}^S - \mathbf{S}$ 是拉布拉斯矩阵, \mathbf{D}^C 是对角矩阵,定义为 $D_{ii}^C = \sum_{j=1}^M C_{ij}$, \mathbf{D}^S 是对角矩阵,定义为 $D_{ii}^S = \sum_{j=1}^N S_{ij}$ 。 $\text{Tr}[\cdot]$ 是矩阵求迹运算。于是,目标函数为:

[0093]

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \delta_{ij} (R_{ij}^0 - R_{ij})^2 + \frac{\lambda_P}{2} \text{Tr}[\mathbf{P}^T \mathbf{P}] + \frac{\lambda_E}{2} \text{Tr}[\mathbf{E}^T \mathbf{E}]$$

$$+ \frac{\lambda_C}{2} \text{Tr}[\mathbf{R}^T \mathbf{L}^C \mathbf{R}] + \frac{\lambda_S}{2} \text{Tr}[\mathbf{R} \mathbf{L}^S \mathbf{R}^T].$$

[0095] 采用梯度下降法求解目标函数,可以得到局部最优解。

[0096] < 新闻配图 >

[0097] 为了使用户快速生动地了解新闻内容,本发明提供了给新闻文档配图说明的方法。该流程包含两个步骤:图像检索词的生成以及检索图像的排序。

[0098] 图 4 显示了本发明中一个新闻配图的实例。

[0099] 如图 4 所示,用户看到美国电影明星朱莉娅·罗伯茨的新闻时想知道她是什么样子,通过本发明提供的图片很容易地获取需要的信息;用户看到冰壶比赛的新闻时想知道冰壶运动是什么样子,通过本发明提供的结果会快速地明白。

[0100] 首先,从新闻文档中提取出图像检索词。由于目前的网络图像搜索引擎不能处理复杂的检索,以及单个词的检索词不能够表达文档内容,本发明提供了一种有效的图像检

索词生成方法。

[0101] 由于一篇文章太长,从正文中抽取检索比较复杂,而经过人工编辑的文档标题是对文档内容很好的总结,因此本发明优先从标题中抽取组成检索的条目,而在标题太短的情况,再从正文中抽取。

[0102] 上述抽取方法需要对标题和正文的词语进行打分,本发明采用词频-反词频模型进行打分。例如,对每个文档,挑选出 c 个检索条目。通常情况下,如果采用太多检索条目进行检索图像,搜索引擎返回很少的结果甚至没有结果;而如果采用单个条目进行检索,返回的结果不能够表示文档的内容。因此,采用组合条目形成不同长度的检索进行检索会得到比较好的结果。据此,本发明提出组合这些检索条目以形成不同长度的检索词进行图像检索,然后将这些返回的结果列表进行融合排序。 c 个条目组合共有 $L = \sum_{k=1}^c \binom{c}{k} = 2^c - 1$ 个检索,提交给图像搜索引擎检索并保存相关图像

[0103] 然后,对所保存的图像进行排序。每个检索爬取前 h 幅图像组成一个列表,共有 L 个列表。本发明采用等级聚合的方法将这 L 个列表进行融合排序,从而为新闻文档选择合适的图像。由于有的文档中包含了人工编辑筛选的图片,这些图片能够很好地反应文档内容,因此与文档中图片视觉上越相似的图片越应排在前面。另外,图像在列表中的位置反应了与检索的文本相关性。因此,本发明提出的方法考虑单幅图像在列表中的位置以及与文档中包含图像的相似性给予其赋予初始的关系分数:

[0104]

$$s_j(x_i) = \mu \left(1 - \frac{k-1}{h}\right) + (1-\mu) \max_{x \in \mathcal{O}} \text{sim}(x, x_i).$$

[0105] 其中, x_i 是第 j 个列表中第 k 个位置的图像, \mathcal{O} 是文档中图像集合。本发明采用 1000 维的视觉词袋特征以及余弦相似度度量图像之间的相似性。

[0106] 为了调节不同长度检索的作用,本发明对具有同等长度的检索赋予相同的权重,即有

[0107]

$$\theta = \theta_{1:L} = \left(\underbrace{\eta_1, \dots, \eta_1}_{\binom{c}{1}}, \dots, \underbrace{\eta_k, \dots, \eta_k}_{\binom{c}{k}}, \dots, \underbrace{\eta_c}_{\binom{c}{c}} \right)$$

[0108] 其中 η_k 是长度为 k 的 $\binom{c}{k}$ 个检索的权重。因此, x_i 的打分是

$$[0109] \quad s(x_i) = \sum_{j=1}^L \theta_j s_j(x_i)$$

[0110] c 个权重是根据训练数据集得到的。采用网格搜索使在训练集上第 10 位置上的归一化 nDCG(normalized Discounted Cumulative Gain) 最大。

[0111] 根据分数可以得到一个排序列表,先采用重复检测算法去除重复图像,然后从去重后的排序列表中为每篇文档选择 r 幅图像。其中,文档自带的图像需优先选择。

[0112] < 检索结果排序 >

[0113] 根据用户提交的检索,系统返回一系列的相关结果。针对新闻领域,用户关注的是新的、重要的以及与检索相关的新闻。本发明提出了一种综合考虑新闻的时效性、重要性以及检索相关性的新闻文档排序方法。检索结果排序流程包括下述步骤:时间信息量化、地名

文档相关性归一化、相关性排序（排序初始化以及排序）。

[0114] 首先进行新闻文档时间信息量化。时间是新闻重要的一个因素。首先要把新闻的时间量化，把时间表示成“年月日”的形式，例如把“九月 12 号，2010”表示成“20100912”。 $date_k$ 表示第 k 个文档的时间量化值，把它进行归一化：

$$[0115] \quad date_k = \frac{date_k - \min_j(date_j)}{\max_j(date_j) - \min_j(date_j)}$$

$$[0116] \quad date_k = \frac{date_k}{\sum_j date_j}$$

[0117] 然后执行归一化文档与地点的相关性。通过一致性约束概率矩阵分解，已得到了文档与地名之间的相关值，并对其进行归一化：

$$[0118] \quad score_k = \frac{score_k}{\sum_j score_j}$$

[0119] 最后执行新闻文档与地点的相关性排序。为了将最新的、热点的和最相关的新闻展示给用户，本发明提出了一种基于马尔科夫随机游走模型的新闻文档排序方法，综合考虑了新闻时间信息、重要性和与检索相关性。该模型可表示为：

$$[0120] \quad r_k^{iter} = y \times r_k^{iter-1} + (1 - y)r_k^0$$

[0121] 其中， r_k^{iter} 是第 k 个文档在第 iter 次迭代的值， r_k^0 是第 k 个文档的初始排序值。y 是权重系数，是非负常数。

[0122] 在随机游走模型中，需要给定一个初始排序值，本发明综合考虑新闻的时间信息以及与检索的相关性，将初始排序值设为

$$[0123] \quad r_k^0 = \frac{date_k + score_k}{2}$$

[0124] 在迭代过程中，考虑到新闻文档的重要性，本发明提出的迭代公式如下：

$$[0125] \quad r_k^{iter} = (1 - y)r_k^0 + y \sum_j \frac{S_{kj}}{\sum_m S_{mj}} r_j^{iter-1}$$

[0126] 其中， S_{kj} 表示两个文档的相似性，y 设为 0.85。重复以上迭代过程，最终达到一个稳定状态，就得到了排序结果。

[0127] 下面详细介绍用户操作界面。

[0128] < 检索结果浏览界面 >

[0129] 图 5 显示了本发明一个实施例的网络新闻检索和浏览界面。

[0130] 为了提供给用户生动快速的检索和浏览界面，本发明提供了一种新颖的用户检索和浏览界面。

[0131] 参见图 5，该界面包括地图视图和新闻事件列表这两个模块，分别对应图 5 中的左右两部分。左半部分是基于谷歌地图的地图视图。用户可以在上面的检索框中输入检索词进行检索，也可以通过浏览地图双击感兴趣的地点进行检索。系统返回检索结果，在地图上对应地点跳出一个窗口，展示排在第一位的新闻的标题以及前两幅最相关图像，用户可以点击“更多”按钮获取更多信息。右半部分按照新闻文档与检索的相关性降序排列，给出了与检索相关的所有事件列表。每一个列表条目对应一个新闻事件，给出了这个新闻事件对应的 5 幅新闻图片、标题和简单描述。如果用户想了解详细信息，可点击“更多”按钮获取

新闻的全文。用户也可以通过点击新闻标题到新闻的原始网页进行访问浏览。

[0132] < 优选实施例 >

[0133] 以下通过一个具体实施例来说明本发明所提供的算法和界面的技术效果。本实施例中,所有的数据是从 ABC、BBC 和 CNN 以及谷歌等新闻网站上爬取的。总共有 48,429 新闻文档和 20,862 个新闻图片。经过过滤和扩展,共得到 4,742 个地名,以及一些参数设定。对于一致性约束概率矩阵分解模型, $H = 100$, $\lambda_p = \lambda_e = 0.001$, $\lambda_c = 2^{-3}$ 和 $\lambda_s = 2^{-4}$ 。在新闻配图中, $h = 20$, $c = 5$, 每个文档抽取 5 个检索条目。

[0134] 为了评估用户对这整个系统的满意程度,本实施例设置了多个用户按照设定好的规则打分,然后采用 nDCG 准则度量评价结果。

[0135] 在评价当中,需要人为标定一些数据,包括地名与文档的相关性以及图片与文档的相关性。本实施例中设定了三种标准:非常相关、相关和不相关,并分别量化为 2,1,0。同时也邀请了 30 名年龄在 20 和 30 之间的用户进行用户研究,这些参与者来自两个国家并经常在线阅读英文新闻。

[0136] 首先对地点相关性分析进行实验评价。随机挑选了 500 个文档分别对进行一致性约束概率矩阵分解细化关系前后进行了评价,得到了两个平均值,分别为 0.492 和 0.954。这表明一致性约束概率矩阵分解对于去噪和细化地点-事件关系具有显著的效果。

[0137] 其次,通过检索检验地名相关性分析。随机选择了 100 个地名进行检索,并比较一致性约束概率矩阵分解、BM25 排序模型和传统的概率矩阵分解模型。

[0138] 图 6 是 BM25 排序模型、概率矩阵分解模型和一致性约束概率矩阵分解模型的检索性能评价结果。

[0139] 图 6 中,横坐标表示评价时选取的排在前面的文档数,纵坐标采用 nDCG 准则的度量值。如图 6 所示,本发明提出的一致性约束概率矩阵分解模型 (CCPMF) 的分析方法的 nDCG 度量值明显高于其他方法的相应值,这表明本发明的分析方法具有显著的技术效果。

[0140] 为了进一步观察模型中参数对系统的影响,本发明分别以变化 λ_c (固定 λ_s) 和变化 λ_s (固定 λ_c) 进行了结果评估。

[0141] 图 7 给出了变化参数在 NDCG@50 规则下的结果。图 7 所示结果说明了参数在很大范围内变化时,本发明提出的一致性约束概率矩阵分解模型 (CCPMF) 的效果明显优于其他两种方法。

[0142] 图 8 是本发明的新闻配图方法与现有技术的性能比较结果。

[0143] 在图 8 比较测试中,为了评价新闻配图,先随机挑出 300 个文档并标注其图像与文档的相关性,组成训练集学习权重系数。然后随机挑选出 1,000 个文档进行测试,并将本发明的新闻配图方法与简单搜索(把标题作为检索检索图像)和简单融合(把标题中的每个单词作为检索检索图像并融合)进行了比较。如图 8 所示,本发明的新闻配图方法明显优于其他两种方法,说明本发明方法的有效性。

[0144] 另外,本实施例还采用用户研究比较了新闻配图前后的性能。每个用户自由浏览并比较这两个版本。对于较差的版本打 1 分,如果另外一个版本与它相比是好,更好或者相当,分别给 2,3 和 1 分。同时也做了双因子方差分析。该比较结果显示在如下的表 1 中。

[0145] 表 1:新闻配图前后的用户研究结果

[0146]

配图后与配图前的比较		排序主题因子		用户因子	
配图后	配图前	F 统计量	概率值	F 统计量	概率值
2.367±0.669	1.100 ± 0.305	58.486	1.970×10^{-8}	0.313	0.999

[0147] 表 1 显示了新闻配图前后的平均值和标准差以及方差分析,其中左边部分给出了平均分和标准差;右边给出了方差分析结果。由表 1 可以看出,用户更喜欢给新闻配图,并且本发明的新闻配图方法性能在统计结果上具有明显的有益效果。

[0148] 图 9 显示了本发明的检索结果排序方法与现有技术的排序方法在检索相关性上的比较结果。

[0149] 在图 9 所示测试中,比较了本发明的检索结果排序方法、PRT(采用时间信息作为静态排序的网页排序方法)方法,PRR(采用地名相关值作为静态排序的网页排序方法)方法以及 BM25 排序方法。随机挑选了 100 个检索进行了相关性评价,评价结果显示在图 9 中。如图 9 所示,BM25 取得了最差的性能,PRR 取得了最好的性能。本发明的检索结果排序方法仅仅略次于 PRR,但是在时效性方面明显优于 PRR 方法(参见图 10),即本发明能够将最新发生的新闻优先地展示给用户。

[0150] 图 10 显示了本发明的检索结果排序方法与现有技术的排序方法在时效性方面的比较结果。

[0151] 图 10 中,纵坐标表示发生在最近一周的新闻所占百分比。在该测试中,统计了 100 个检索返回列表中,在前 $d(d = 5, 10, 20, 50, 100)$ 个新闻中发生在最近一周的新闻的平均百分比。由图 10 可见,本发明的检索结果排序方法仅仅略次于 PRT 方法,但是如图 9 所示,在检索相关性方面明显优于 PRT 方法。

[0152] 综合图 9 和图 10 所示的测试结果,可以看到本发明的检索结果排序方法在相关性和时效性的综合效果明显优于现有技术中的排序方法,因而能够取得令人满意的性能。

[0153] 应当理解的是,本发明的上述具体实施方式仅仅用于示例性说明或解释本发明的原理,而不构成对本发明的限制。因此,在不偏离本发明的精神和范围的情况下所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。此外,本发明所附权利要求旨在涵盖落入所附权利要求范围和边界、或者这种范围和边界的等同形式内的全部变化和修改例。

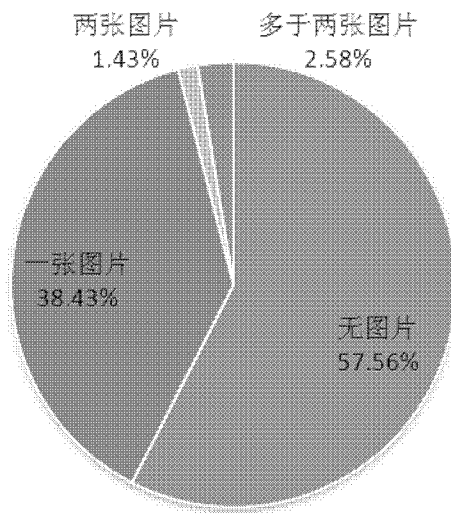


图 1

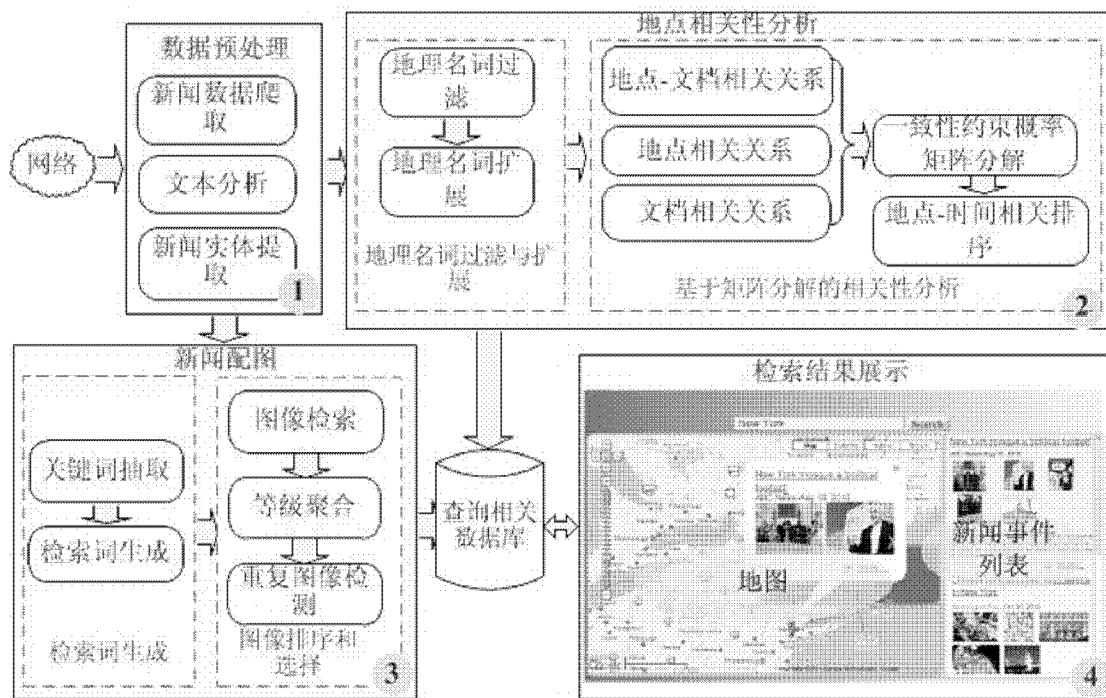


图 2

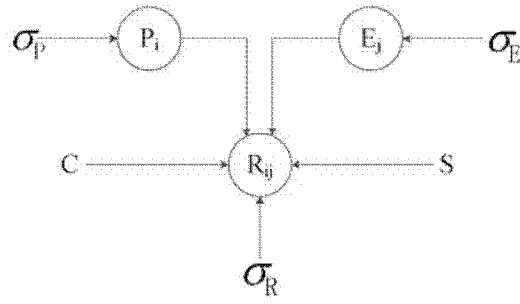


图 3

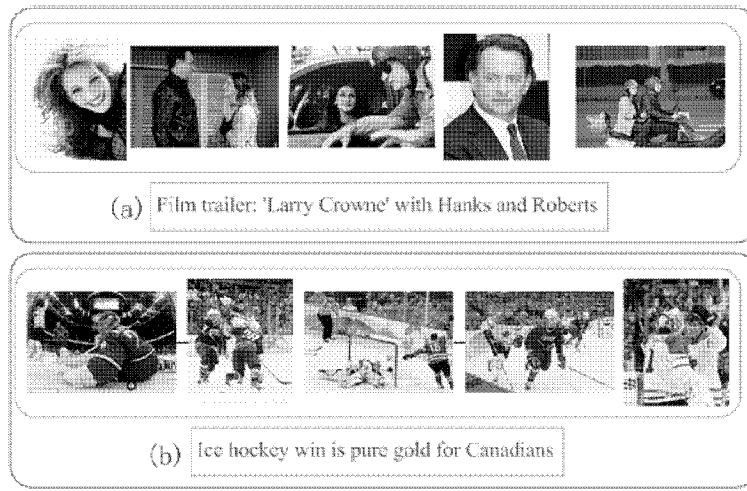


图 4

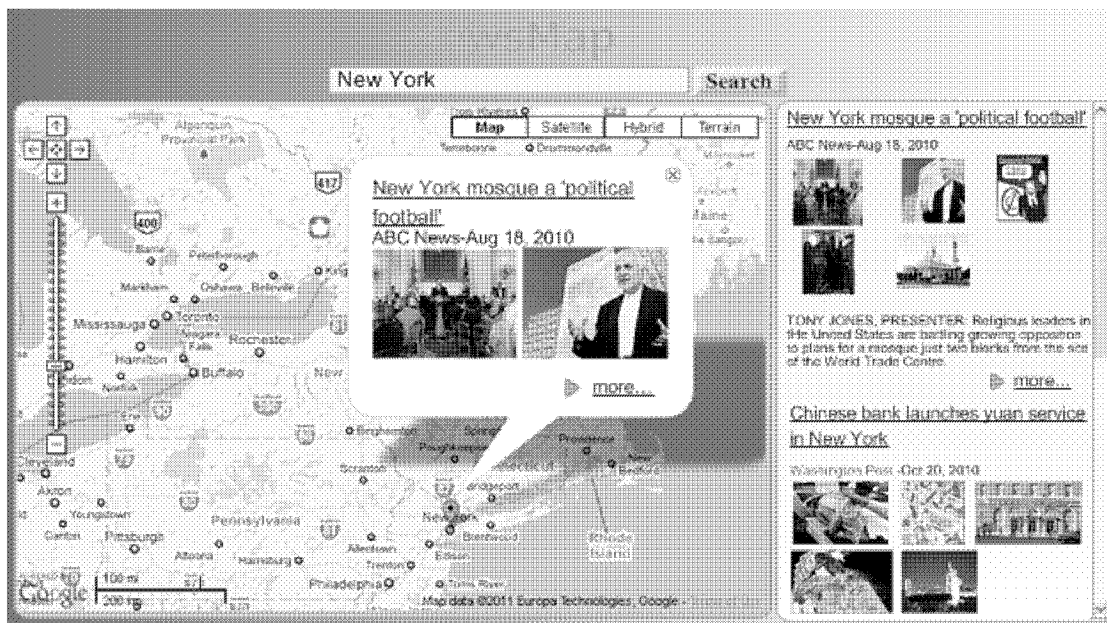


图 5

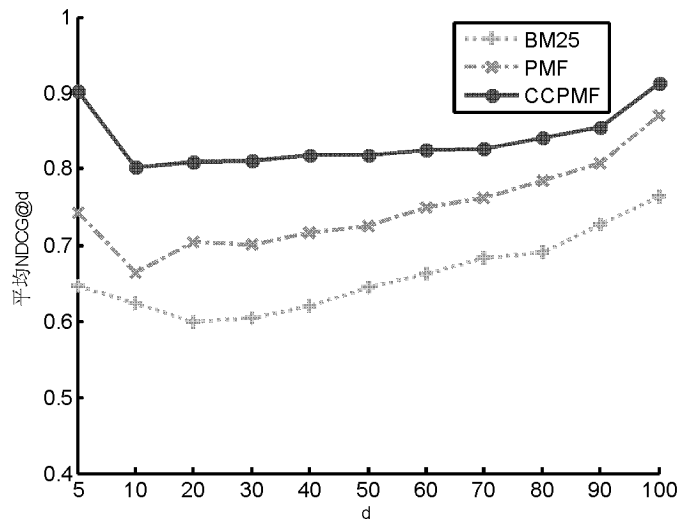


图 6

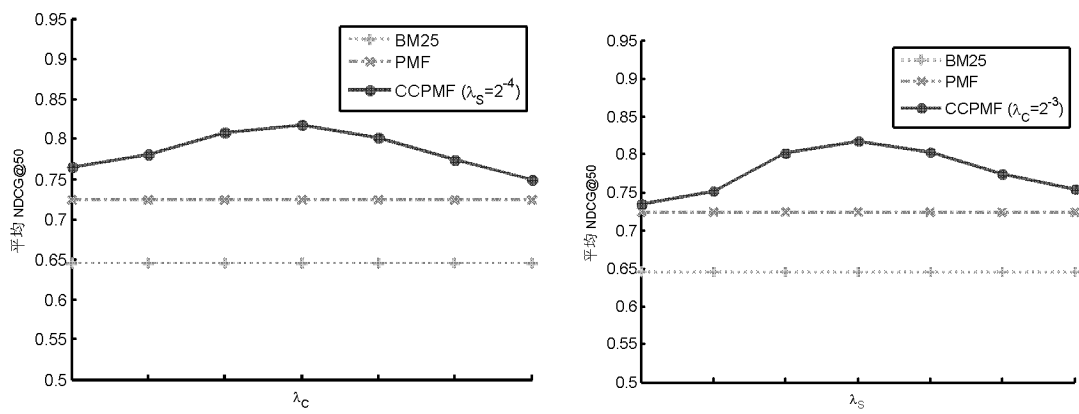


图 7

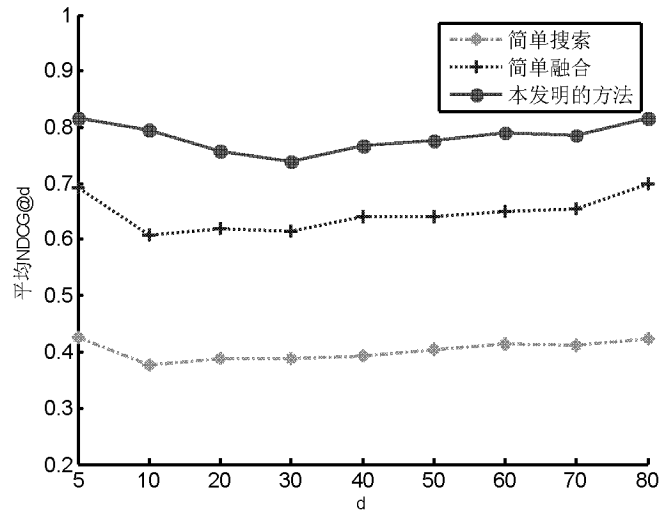


图 8

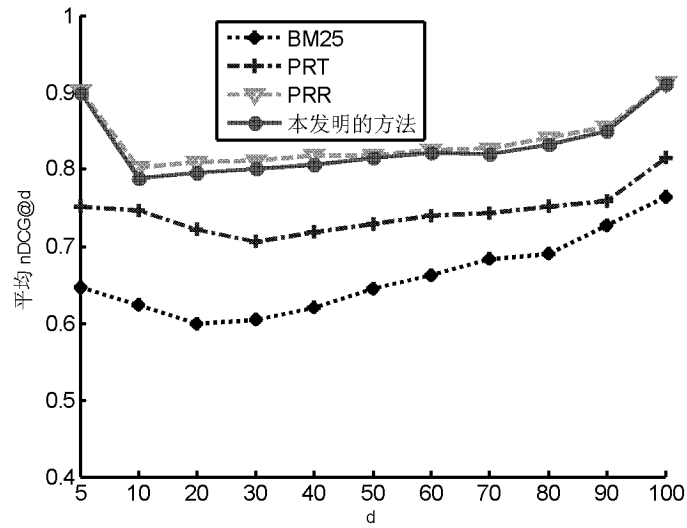


图 9

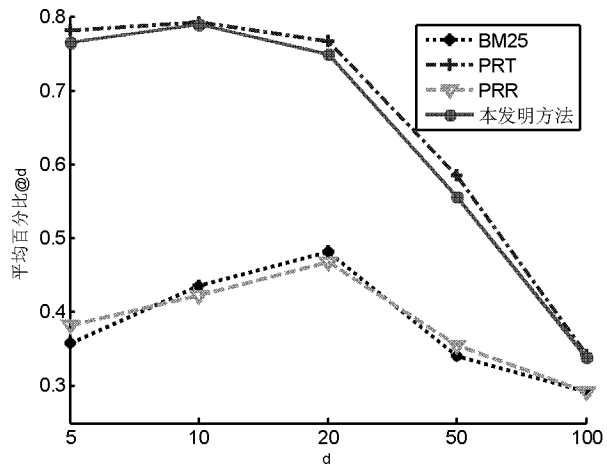


图 10