

文章编号: 2095-2163(2020)01-0050-06

中图分类号: TP391

文献标志码: A

基于 Doc2Vec 和深度神经网络的战场态势智能推送研究

申 远, 黄志良, 胡 彪, 王适之

(空军预警学院, 武汉 430019)

摘要: 战场态势智能推送中的推荐算法是最核心、最关键的组成部分,在很大程度上决定了战场态势推送性能的优劣。深度神经网络推荐算法对用户和态势的需求度关系进行建模学习,能够学习到两者之间隐含的、更为复杂的非线性关系,缺点在于未利用到态势的额外信息。因此以深度神经网络推荐算法为基础,利用 Doc2Vec 方法提取态势文本内容信息来作为深度神经网络模型的辅助输入,设计了一种基于 Doc2Vec 和深度神经网络的战场态势智能推送方法。实验仿真结果表明,利用 Doc2Vec 算法来对态势的额外信息进行建模,能够提升算法推荐效果。

关键词: 战场态势; 智能推送; 推荐算法; 深度神经网络; Doc2Vec

Research on battlefield situation intelligent push based on Doc2Vec and Deep Neural Network

SHEN Yuan, HUANG Zhiliang, HU Biao, WANG Shizhi

(Air Force Early Warning Academy, Wuhan 430019, China)

[Abstract] The recommendation algorithm in the intelligent push of battlefield situation is the core and most important component, which largely determines the pros and cons of the battlefield situation. The deep neural network recommendation algorithm models the user and situational demand relationship, which could learn the implicit and more complex nonlinear relationship between the two. The disadvantage is that the extra information of the situation is not utilized. Therefore, based on the deep neural network recommendation algorithm, the Doc2Vec method is used to extract the text content information of the situation as the auxiliary input of the deep neural network model, and a battlefield situation intelligent push method based on Doc2Vec and Deep Neural Network is designed. The experimental simulation results show that using Doc2Vec algorithm to model the additional information of the situation can improve the algorithm recommendation effect.

[Key words] battlefield situation; intelligent push; recommendation algorithm; Deep Neural Network; Doc2Vec

0 引言

随着战场信息搜集手段、处理分析方法的快速发展,使得作战用户能够获得大量的不同类型、不同粒度、不同时空的战场态势资源,但是也会使得指挥员需要耗费很大的时间精力才能从海量的态势中发掘有效信息资源,会出现“态势泛滥”和“态势缺乏”的矛盾,即一方面战场态势呈指数级增长,另一方面作战用户可以利用到的合适态势资源却很少^[1]。目前解决该矛盾问题最直接有效的方法就是智能推送,态势智能推送是以推荐算法为核心,利用知识发现、数据挖掘、机器学习、人工智能等多种方法为用户提供合适的态势资源,态势智能推送能够根据用户个性化需求,主动为用户提供合适的态势,并且当用户需求变化时,态势智能推送也能自适应变化,及时准确地为用户提供个性化、智能化、动态化的态势资源。

智能推送的核心是推荐算法,常见的推荐算法有协同过滤推荐算法^[2]、基于内容的推荐算法^[3]和组合推荐算法^[4]。目前在军用领域,推荐算法理论研究也有一定的成果。胡旭等人^[5]针对协同过滤推荐算法的数据稀疏性、扩展性差问题,提出了初始聚类中心优化的 K-均值项目聚类推荐算法;余苗等人^[6]利用层次向量空间构建用户兴趣空间,根据情报用户少量的定制信息和朴素贝叶斯分类算法建立用户兴趣模型,提出了一种基于朴素贝叶斯分类算法的雷达情报按需分发技术;傅畅等人^[7]设计并实现了一个包括采集、处理、存储与检索的 Web 军事情报挖掘模型,提出了一种面向军事情报应用的文本聚类方法;袁仁进^[8]围绕军事新闻推荐模型构建与更新、融合地理情境的军事新闻推荐模型构建等方面进行了研究,提出了一种基于向量空间模型和 Bisecting K-means 聚类的军事新闻推荐方法,针对

作者简介: 申 远(1995-),男,硕士研究生,主要研究方向:预警情报组网;黄志良(1975-),男,副教授,主要研究方向:预警情报组网;胡彪(1985-),男,硕士研究生,主要研究方向:预警作战指挥;王适之(1995-),男,硕士研究生,主要研究方向:预警情报组网。

收稿日期: 2019-11-13

顾及时间上下文的用户兴趣模型更新问题,构造了一种基于时间的遗忘函数,最后针对地理推荐问题,提出了基于地理上下文的军事新闻推荐算法,并构建了军事新闻个性化推荐原型系统。

上述研究主要是利用用户和态势两者之间的历史交互关系进行学习建模,但该方法存在的问题是用于实现算法模型训练的用户历史行为记录数据是比较稀疏的,并且很难解释用户与这类态势产生交互行为(如定制、浏览、评价等)的原因,比如用户定制这类态势而不定制另外一类态势,很难用用户—态势需求度矩阵来解释。如果能够对用户产生交互行为的相对应态势内容进行分析,那么必将使得用户建模更为准确,并具有很强的解释性,能够描述用户历史行为记录背后隐藏着抽象的用户—态势需求关系。因此本文在神经网络推荐算法的基础上,利用自然语言学习中的 Doc2Vec 算法来提取态势内容特征信息,并训练成态势内容特征向量,再与用户—态势需求度矩阵中的用户—态势交互行为向量进行融合,这就使得算法模型能够学习到态势更多的特征信息,在一定程度上缓解数据稀疏性问题,从而提升推荐效果。

1 基于 Doc2Vec 的词嵌入模型

在对态势进行建模时,一般采用向量空间模型来表示。以预警情报为例,预警情报则可以表示为 n 维的特征向量 $\{(t_1, \omega_1), (t_2, \omega_2), \dots, (t_i, \omega_i), (t_n, \omega_n)\}$, t_i, ω_i 分别表示这些特征的关键词和相应的权重。这种表示方法的核心是对关键词进行提取,如文献[9]将预警情报特征提取为位置、高度、航向、速度、属性、机型等特征,这样一来该特征选取方法明显带有主观性,无法将情报更多的特征信息挖掘出来,会造成情报特征信息的流失,使得用户需求建模就会变得不准确。

在实际运用过程中,战场态势种类是多种多样的,如视频类、音频类、图像类、文本类等。其中,文本类态势特征明显,特征提取相对容易,而诸如视频类、音频类、图像类这些态势特征提取较为困难,因此需要对这些非结构化数据进行结构化处理,形成文本类态势。在得到同类型的态势文本数据后,可以利用相应的方法对态势内容进行表示。而文本内容一般采用词袋模型将文本转化成低维空间的稠密向量,这样的表示方法会使得文本所有重要的特征被提取出来,而不是仅仅取几个主要的文本特征^[10]。以“敌方/导弹/距离/我方/20km”和“敌方/飞机/在/我方/东北/方向”为例,基于以上文本内

容则可建立词典为:{"敌方":1,“导弹”:2,“飞机”:3,“距离”:4,“在”:5,“我方”:6,“20km”:7,“东北”:8,“方向”:9}。假设每一个词都有唯一的索引,根据各词的出现顺序及频率,则上述文本用词袋模型中9维的词向量分别表示成 $[1, 1, 0, 1, 0, 1, 1, 0, 0]$ 、 $[1, 0, 1, 0, 1, 1, 0, 1, 1]$ 。这种表示方式好处在于将文本内容的所有关键词都提取出来,态势信息主要信息被提取出来。缺点在于该方法忽略了各词语的上下文顺序,且当存在大量且特征繁多的战场态势时,态势特征的表示就会变得非常稀疏,即用词袋模型表征所有的态势特征时,对于单个态势而言,所拥有的特征相对有限,则词袋模型表示的大部分特征对单个态势是没有意义的。因此用向量空间模型来表示态势将具有很大的局限性。

基于 Doc2Vec 的词嵌入模型是 Le 和 Mikolov 提出的一种深度文本学习算法。Doc2Vec 算法是从大型原始数据中以完全无监督的方式进行训练,而无需任何针对于特定任务的标记数据^[11]。该方法的优点在于应用场景广阔,可利用神经网络将任何长度的文本生成词向量嵌入到算法模型。Doc2Vec 方法的基本思想是通过神经网络学习训练,将文本中每一个词表示成低维稠密向量形式,最终的目标是生成可表示的词向量。根据文本领域的不同,该向量的维度会有所不同,通常的做法是取值 50 或 100,针对特定领域,向量维数固定的好处在于不会产生高维向量造成“维数灾难”问题。其中,向量的每一个维度包含着文本的某一种潜在特征,每个维度的取值范围为 0~1 的实数。对于一个文本而言,在经过预处理后,文本主要内容特征很容易人工提取,关键问题战场态势是海量的,不可能对每个文本都采用这样的方式,因此要根据文本内容的上下文顺序,利用 Doc2Vec 训练模型训练语料库来预测下一个词语,根据预测结果可以为每一个文本生成词向量,然后,Doc2Vec 方法具有 2 种模型。一种是分布式记忆模型(Distributed Memory, DM),另外一种是分布式词袋模型(Distributed Bag of Words, DBOW)。2 种方法的训练方式相同,都采用神经网络为基础进行训练。不同之处在于前者是设定一定阈值后,利用当前词语去预测阈值范围内的词语,从而形成词向量,而后者是根据词语之间的上下文关系来预测下一个词语,从而形成词向量。对于态势文本而言,文本中各类词语往往具有很大的关联性,采用分布式词袋模型来对文本建模更为准确,分布式词袋模型结构如图 1 所示。

分布式词袋模型的输入为一个随机段落,将文本段落映射到低维稠密形式的段落向量,分布式词袋模型可以通过一个段落向量来预测某个随机词的概率分布,而后再将出现概率高的词组成关键词^[12]。假设语料库由长度为 n 的词序列 $(x_1, x_2, \dots, x_i, \dots, x_n)$ 组成,文本组成的序列集合 A ,则 Doc2Vec 算法的训练目标是使每一个词在语言生成模型出现的概率最大,即 $\sum_A p(x_i | x_1, x_2, \dots, x_i, \dots, x_n)$ 最大,而当利用随机梯度下降法对算法模型训练后,在输出层将每一个词向量按顺序进行连接,则此时文本可以由多维词向量进行表示。

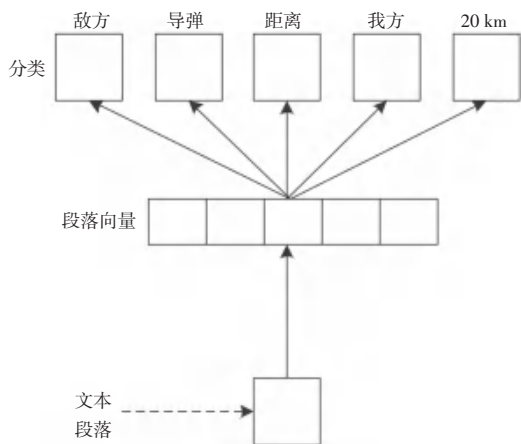


图1 分布式词袋模型结构

Fig. 1 Distributed word bag model structure

2 深度神经网络推荐算法

深度神经网络算法因其鲁棒性好、自适应能力强、处理非线性关系能力优异、并行计算速度快等优点而广泛应用于图像处理、模式识别、经济预测、业务决策等领域。文献[13]提出将深度学习方法—多层感知机应用于协同过滤推荐算法,该深度神经网络推荐算法(Deep Neural Network, DNN)通过建立多层前馈神经网络模型来学习用户—态势交互关系,使得算法模型更具备非线性建模能力。

深度神经网络推荐算法是利用需求度的已知项来预测需求度未知项。深度神经网络推荐算法通过建立利用用户—态势交互函数 $f \in R^{X \times Y}$ 来表示用户历史反馈行为数据, R 为用户—态势需求度矩阵, X 和 Y 分别表示用户和态势的数目,定义如下:

$$f_{ui} = \begin{cases} 0, & \text{用户和态势无交互行为;} \\ 1, & \text{用户和态势有交互行为.} \end{cases} \quad (1)$$

研究中,利用用户行为数据来进行推荐的问题可以抽象表示为学习模型函数 $\tilde{f}_{ui} = h(u, i | \theta)$, 其中, \tilde{f}_{ui} 表示用户 u 对态势 i 的预测需求度, θ 表示模

型参数,此处模型为深度神经网络, h 为模型参数映射到预测需求度的函数。与矩阵分解推荐算法类似,假设 $P \in R^{X \times K}$ 和 $Q \in R^{Y \times K}$ 分别表示用户和态势的特征矩阵,则深度神经网络推荐模型预测需求度为:

$$\tilde{f}_{ui} = h(P^T S_u^U, Q^T S_i^I | P, Q, \theta_h) = \psi_{out}(\psi_Z(\dots \psi_2(\psi_1(P^T S_u^U, Q^T S_i^I)) \dots)), \quad (2)$$

其中, U 和 I 分别表示用户和态势集合; S_u^U 和 S_i^I 分别用来描述用户 u 和态势 i ; θ_h 表示映射函数 h 的模型参数; ψ_{out} 为输出层映射函数; ψ_Z 表示隐藏层第 Z 个映射函数。设置 Sigmoid 激活函数 $\sigma(x) = 1/(1 + e^{-x})$ 作用于输出层,定义权重为 O^T , 则预测需求度修正为:

$$\tilde{f}_{ui} = 1/(1 + \exp(-\psi_Z(\dots \psi_2(\psi_1(P^T S_u^U, Q^T S_i^I)) \dots)) \cdot O^T). \quad (3)$$

和矩阵分解推荐算法一样,在得到预测需求度后,可建立目标函数进行参数优化来找到最合适的参数 P 、 Q 和 O^T 。

综上所述,研发得到深度神经网络推荐算法的流程步骤分述如下:

Step 1 输入用户—态势需求度矩阵;

Step 2 对 P 、 Q 和 O^T 等模型参数进行初始化;

Step 3 通过相关处理,将用户和态势特征向量输入到深度神经网络来预测需求度;

Step 4 利用随机梯度下降法对模型进行训练,直到模型收敛;

Step 5 根据模型迭代结果,找到最合适的 P 、 Q 和 O^T , 计算此时的预测需求度 \tilde{f}_{ui} ;

Step 6 对各态势的预测需求度进行排序,将排序高的态势推送给用户。

3 基于 Doc2Vec 和深度神经网络的推荐算法

研究中利用 Doc2Vec 方法将态势的文本信息融入到深度神经网络模型后,文本中的内容信息被建模成词向量,与深度神经网络中的用户的潜在特征向量和态势潜在特征向量进行拼接融合,通过在嵌入层的特征叠加,则算法模型可以学习用户—态势建立关联关系隐含着的抽象特征。基于 Doc2Vec 和深度神经网络推荐算法结构如图 2 所示。

该模型主要分为输入层、嵌入层、隐藏层和输出层。其中,输入层是输入已经分布式处理后的用户特征向量、态势特征向量和态势内容特征向量;嵌入层用于将用户序列、态势序列以及态势内容序列转换成表征各自特征的低维稀疏向量,对于态势内容

特征向量,文本中每一个词语在嵌入层表示为多维向量 $\mathbf{D}_i = \{d_1, d_2, \dots, d_i, \dots, d_n\}$, 这样的表示方法使得文本中更多的特征被提取出来;在隐藏层,主要是对用户特征向量、态势特征向量以及态势内容特征向量进行连接得到一个用户—态势—态势内容特征向量,用于深度神经网络的多层感知机的输入,其中神经网络的结构采用是倒塔式结构,即下一层的神经元的数目为上一层的一半,此时可以利用多层感知机对用户—态势—态势内容特征向量进行特征学习并作为输出层的输入;输出层主要是根据学习到的用户—态势—态势内容特征,可对用户—态势—态势内容特征向量进行需求度预测,根据预测值与实际值的误差情况进行自适应调整,如果输出结果与实际结果相差太大,则不断调整各个参数,直到预测值与实际值的误差值小到一定程度后,即神经网络训练达到最优值,这时就不再调整参数,每次这样的深度神经网络学习成功的结果都会被神经网络中的神经元记忆储存,当作战用户提出态势需求时,深度神经网络都可以根据学习的结果向该用户主动推送合适的态势资源。

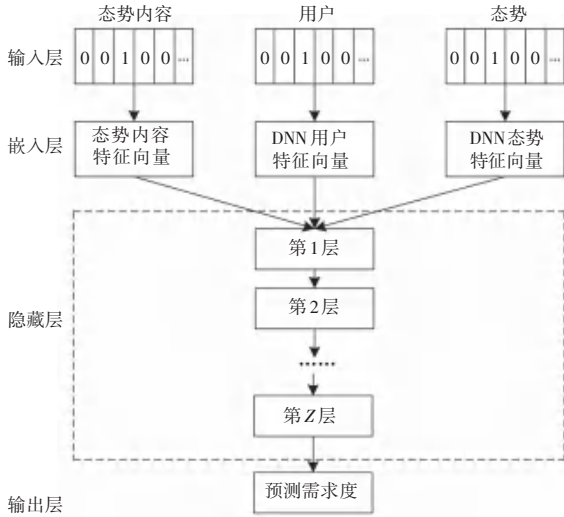


图2 基于 Doc2Vec 和深度神经网络推荐算法模型

Fig. 2 Recommendation algorithm model based on Doc2Vec and Deep Neural Network

假设态势内容特征向量记为 \mathbf{D}_i , Doc2Vec 向量预测已建模成态势潜在向量的堆积,假设权重向量为 \mathbf{G} , 偏置向量为 \mathbf{b} , Doc2Vec 预测向量为 $\tilde{\mathbf{D}}_i$, 则各向量间满足以下等式:

$$\tilde{\mathbf{D}}_i = \mathbf{G} \mathbf{D}_i + \mathbf{b}, \quad (4)$$

因为在深度神经网络算法中,态势内容向量也是在预测,因此也会存在预测上的误差(预测值与真实值),为防止训练过拟合,添加权重正则化

项 $\lambda_D \|\mathbf{G}\|^2$, 则此时基于态势内容向量预测的目标函数为:

$$F_{\text{Doc2Vec}} = \sum_{i \in I} \|\mathbf{D}_i - \tilde{\mathbf{D}}_i\|^2 - \lambda_D \|\mathbf{G}\|^2, \quad (5)$$

对于深度神经网络部分,定义 f^+ 表示用户和态势有交互行为, f^- 表示用户和态势无交互行为,将输出层输出值范围限定为 $[0, 1]$, 输出层激活函数保持不变,则定义似然函数为:

$$p(f^+, f^- | \mathbf{P}, \mathbf{Q}, \theta_h) = \prod_{r_{ui} \in f^+} \tilde{f}_{ui} \prod_{r_{ui} \in f^-} (1 - \tilde{f}_{ui}), \quad (6)$$

对式(6)取负对数, $f(\lambda_{\text{DNN}})$ 表示深度神经网络的正则化项,用于防止模型过拟合,则此时深度神经网络算法需要进行优化的目标函数为:

$$\begin{aligned} F_{\text{DNN}} = & - \sum_{r_{ui} \in f^+} \lg \tilde{f}_{ui} - \sum_{r_{ui} \in f^-} - \lg(1 - \tilde{f}_{ui}) - f(\lambda_{\text{DNN}}) = \\ & \sum_{r_{ui} \in f^+ \cup f^-} f_{ui} \lg \tilde{f}_{ui} + (1 - f_{ui}) \lg(1 - \tilde{f}_{ui}) - f(\lambda_{\text{DNN}}) = \\ & \sum_{r_{ui} \in f^+ \cup f^-} f_{ui} \lg \tilde{f}_{ui} + (1 - f_{ui}) \lg(1 - \tilde{f}_{ui}) - \lambda_{\text{DNN}} (\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2 + \|\mathbf{O}^T\|^2), \end{aligned} \quad (7)$$

根据式(5)~式(7)可知基于 Doc2Vec 和深度神经网络的推荐算法的目标函数为:

$$\begin{aligned} F_{\text{DNN+Doc2Vec}} = & F_D + F_{\text{DNN}} = \\ & \sum_{i \in I} \|\mathbf{D}_i - \tilde{\mathbf{D}}_i\|^2 - \lambda_D \|\mathbf{G}\|^2 + \\ & \sum_{r_{ui} \in f^+ \cup f^-} f_{ui} \lg \tilde{f}_{ui} + (1 - f_{ui}) \lg(1 - \tilde{f}_{ui}) - \\ & \lambda_{\text{DNN}} (\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2 + \|\mathbf{O}^T\|^2). \end{aligned} \quad (8)$$

在得到目标函数后,利用随机梯度下降法^[14]进行学习训练,根据训练结果,将一组预测需求度高的相应态势推送给用户。

综上所述,则可以得到基于 Doc2Vec 和深度神经网络的推荐算法的算法步骤详述如下:

Step 1 输入用户—态势需求度矩阵;

Step 2 对态势原始数据进行预处理,包括数据清洗,预训练等;

Step 3 利用 Doc2Vec 方法训练词向量,并将文本内容特征耦合到用户—态势潜在特征中去;

Step 4 对推荐模型参数初始化;

Step 5 利用随机梯度下降法对模型进行训练,直到模型收敛;

Step 6 根据模型迭代结果,找到最合适的用户特征矩阵、态势特征矩阵和各参数的权重值,计算此时的预测需求度 \tilde{f}_{ui} ;

Step 7 对各态势的预测需求度进行排序,将排序高的态势推送给用户。

4 实验结果及分析

本文实验数据采用雷达航迹仿真得到各类情报750 000条,其中用户定制的情报为1 500条,将数据划分为训练集和测试集。将各类情报数据转化成文本类型情报后,首先要对文本进行清洗,包括剔除文本中无用的标点符号信息;其次需要对文本进行分词及同义词合并,在分词的结果中剔除对语义表达没有影响的停用词;最后训练词向量,实现更加精确的词向量表示。

为评估算法性能,需要设置相应的评价指标来衡量算法优劣。本文采用均方根误差、准确率、召回率和 $F1$ 这四个指标,各指标详细解析参见文献[15],此处不再赘述。实验编程语言用Python3.5,采用的机器学习库包括Numpy、Pandas、Scikit-Learn、Gensim、Jieba和CPU版本的TensorFlow。

在参数设置上,最大迭代次数为40次、批处理参数为1 000、隐因子个数为8、正则化参数为 $\lambda_D = 0.05$ 、学习速率为0.001、网络层数分布为[64, 32, 16, 8, 1]。本文算法性能受迭代次数($epoch$)、隐因子个数($factor$)、正则化参数($regulation$)、学习率($learning\ rate$)、神经网络层数($layer$)、推荐数目(k)等参数的影响。其它参数条件不变,来比较分析单一参数对算法性能的影响。对比实验的研究分析内容具体如下。

(1)对比实验1。其它参数不变,不同迭代次数条件下,对比本文算法(DNN+Doc2Vec)与神经网络推荐算法(DNN)、矩阵分解推荐算法^[16](MF)的均方根误差。实验结果如图3所示。

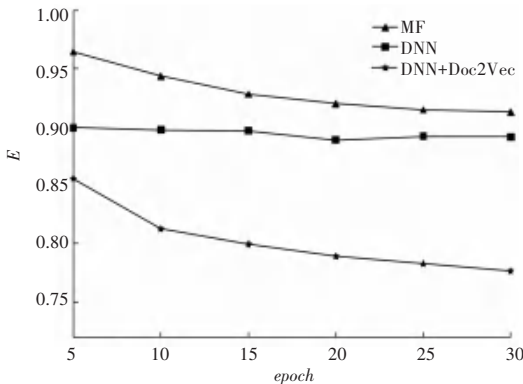


图3 3种算法均方根误差比较

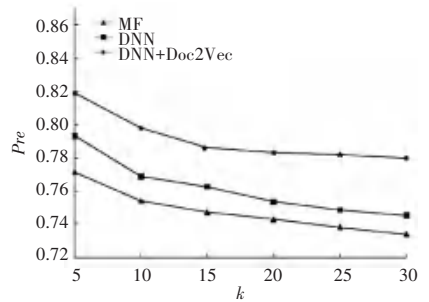
Fig. 3 Comparison of root mean square error of three algorithms

由图3可知,分析得到的研究结论可阐释表述如下:

① 本文算法(DNN+Doc2Vec)均方根误差随迭代次数增加而减少,并且要小于神经网络推荐算法和矩阵分解推荐算法,本文算法在神经网络推荐算法的基础上,利用Doc2Vec方法学习文本内容特征,能够提升模型精度;

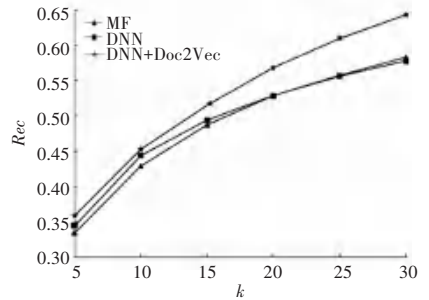
② DNN+Doc2Vec算法均方根误差随迭代次数增加而减少,但减小速率变慢,这是因为此时DNN+Doc2Vec算法训练正逐渐逼近最优解,故而算法推荐性能有所提升。

(2)对比实验2。其它参数不变,迭代次数为30次,不同推荐数目下,本文算法与神经网络推荐算法(DNN)、矩阵分解推荐算法(MF)的准确率、召回率和 $F1$ 指标比较。实验结果如图4所示。



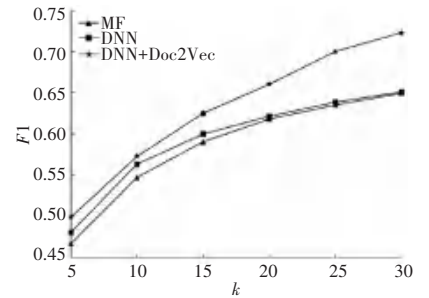
(a) 准确率

(a) Accuracy



(b) 召回率

(b) Recall



(c) $F1$

(c) $F1$

图4 3种算法准确率、召回率、 $F1$ 指标比较

Fig. 4 Comparison of the precision, recall and $F1$ of the three algorithms

由图4(a)、(b)、(c)可知,分析得到的研究结论可阐释表述如下:

(1)本文算法(DNN+Doc2Vec)与神经网络推荐算法和矩阵分解推荐算法一样,准确率均随推送数目增大而变小,3种算法的召回率均随推送数目增大而变大,3种算法的F1指标均随推送数目增大而变大;

(2)DNN+Doc2Vec算法与DNN算法、MF算法相比而言,前者的均方根误差要小于后两种算法,说明本文算法利用Doc2Vec方法学习文本内容特征后,算法的推荐准确性有一定的提升。

综上所述,神经网络是推荐算法的基础,引入Doc2Vec方法来学习态势文本内容,能够挖掘出用户需求背后隐藏的用户和态势更加复杂的交互关系,使得态势特征表征更加明确,推荐效果得以提升,证明本文算法(DNN+Doc2Vec)确实能够提升推荐质量。

5 结束语

本文在军事运用背景条件下,提出了一种基于Doc2Vec和深度神经网络的推荐算法。该算法是在神经网络推荐算法对用户—交互关系进行建模的基础上,引入基于Doc2Vec的词嵌入模型来对文本态势实现建模,以期挖掘到态势更多的特征信息,使得用户需求建模更为准确。实验表明,基于Doc2Vec和深度神经网络的推荐算法与神经网络推荐算法、矩阵分解推荐算法等算法相比,本文算法得到的推荐效果更为优异。

参考文献

- [1] 熊家军,兰旭辉,代科学,等. 战略预警情报服务[M]. 北京:军事科学出版社,2015.
- [2] PAGARE R, PATIL S A. Study of collaborative filtering recommendation algorithm - scalability issue [J]. International Journal of Computer Applications, 2013, 67(25):10-15.
- [3] SHU Jiangbo, SHEN Xiaoxuan, LIU Hai, et al. A content-based recommendation algorithm for learning resources [J]. Multimedia

- Systems, 2018, 24(2):163-173.
- [4] ZHEN Yanggai. The application of hybrid recommendation algorithm in education cloud [J]. Applied Mechanics & Materials, 2014, 551:670-674.
- [5] 胡旭,鲁汉榕,陈新,等. 初始聚类中心优化的K-均值项目聚类推荐算法[J]. 空军预警学院学报, 2014, 28(3):203-207.
- [6] 余苗,杨瑞娟,程红斌,等. 基于Naive Bayes算法的雷达情报分发技术[J]. 现代雷达, 2014, 36(7):46-50, 53.
- [7] 傅畅,宋佳庆. 一种基于文本聚类的web军事情报挖掘系统设计与实现[J]. 中国电子科学研究院学报, 2015, 10(5):541-545.
- [8] 袁仁进. 军事新闻个性化推荐关键技术研究[D]. 郑州:战略支援部队信息工程大学, 2018.
- [9] 古秦弋,杨瑞娟,黄美荣. 基于内容相似度的雷达情报筛选技术[J]. 空军预警学院学报, 2017, 31(3):190-193.
- [10] PASSALIS N, TEFAS A. Learning bag-of-embedded-words representations for textual information retrieval [J]. Pattern Recognition, 2018, 81:254-267.
- [11] LE Q V, MIKOLOV T. Distributed representations of sentences and documents [C]//Proceedings of the 31st International Conference on Machine Learning (ICML-14). Beijing: Springer, 2014, II: 1188-1196.
- [12] WANG Ronggui, DING Kai, YANG Juan, et al. A novel method for image classification based on bag of visual words [J]. Journal of Visual Communication & Image Representation, 2016, 40(A): 24-33.
- [13] Dziugaite G K, Roy D M. Neural network matrix factorization [EB/OL]. (2015-12-15) [2018-12-31]. https://www.researchgate.net/publication/284476555_Neural_Network_Matrix_Factorization/amp.
- [14] JIN Jing, LAI Siyan, HU Su, et al. GPUSGD: A GPU-accelerated stochastic gradient descent algorithm for matrix factorization [J]. Concurrency & Computation Practice & Experience, 2015, 28(14):3844-3865.
- [15] KATSONIS P, OLIVIER L. A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein-coding variations on fitness [J]. Genome Research, 2014, 24(12):2050-2058.
- [16] YIN Hang, CHANG Guiran, WANG Xingwei. A cold-start recommendation algorithm based on new user's implicit information and multi-attribute rating matrix [C]//9th International Conference on Hybrid Intelligent Systems (HIS 2009). Shenyang, China: IEEE, 2009:353-358.