



(12)发明专利

(10)授权公告号 CN 105590623 B

(45)授权公告日 2019.07.30

(21)申请号 201610102345.7

G10L 25/30(2013.01)

(22)申请日 2016.02.24

(56)对比文件

(65)同一申请的已公布的文献号

申请公布号 CN 105590623 A

CN 1667699 A,2005.09.14,
CN 102880906 A,2013.01.16,
CN 1979637 A,2007.06.13,
EP 1184839 A2,2002.03.06,

(43)申请公布日 2016.05.18

US 2002/0046025 A1,2002.04.18,
Domokos Jozsef etc.Automated

(73)专利权人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号
百度大厦三层

Grapheme-to-Phoneme Conversion System for
Romanian.《2011 6th Conference on Speech
Technology and Human-Computer Dialogue》
.2011,第1-6页.

(72)发明人 陈志杰 康永国

(74)专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

代理人 宋合成

王阳.消除隐层冗余信息的改进神经网络及
应用.《中国优秀硕士学位论文全文数据库-信息
科技辑》.2012,(第06期),第5-8页.

(51)Int.Cl.

G10L 13/08(2013.01)

G10L 15/187(2013.01)

审查员 李春雨

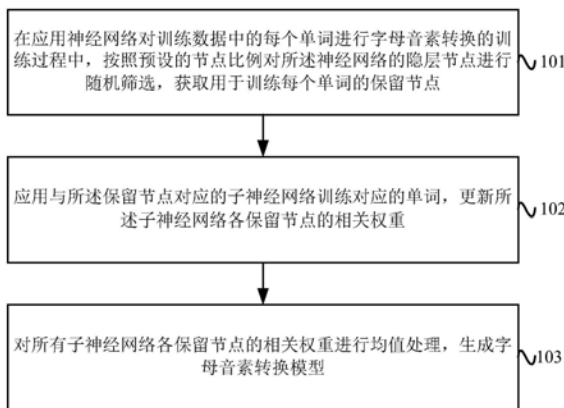
权利要求书1页 说明书6页 附图2页

(54)发明名称

基于人工智能的字母音素转换模型生成方法
及装置

(57)摘要

本申请提出一种基于人工智能的字母音素
转换模型生成方法及装置,其中,该方法包括:在
应用神经网络对训练数据中的每个单词进行字
母音素转换的每次训练过程中,按照预设的节
点比例对所述神经网络的隐层节点进行随机筛
选,获取用于训练每个单词的保留节点;应用与
所述保留节点对应的子神经网络训练对应的单
词,更新所述子神经网络各保留节点的相关权
重;最终对所有子神经网络各保留节点的相关
权重进行均值处理,生成字母音素转换模型。
通过本申请提供的基于人工智能的字母音素
转换模型生成方法和装置,提高了字母音素转
换模型的学习能力和泛化能力,从而避免了过
拟合现象。



1. 一种基于人工智能的字母音素转换模型生成方法,其特征在于,包括以下步骤:

在应用神经网络对训练数据中的每个单词进行字母音素转换的训练过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;

应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;

对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型;

其中,所述按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点,包括:

根据预设的节点隐藏比例,随机将所述神经网络中每层隐层节点中与所述节点隐藏比例对应数量的部分节点进行隐藏处理,以使每层隐层节点中没有隐藏的部分节点作为用于训练每个单词的保留节点。

2. 如权利要求1所述的方法,其特征在于,还包括:

根据所述训练数据中的单词数量调整所述神经网络的层数、每层节点数、以及所述预设的节点比例。

3. 一种基于人工智能的字母音素转换模型生成装置,其特征在于,包括:

处理模块,用于在应用神经网络对训练数据中的每个单词进行字母音素转换的训练过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;

训练模块,用于应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;

生成模块,用于对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型;

其中,所述处理模块用于:

根据预设的节点隐藏比例,随机将所述神经网络中每层隐层节点中与所述节点隐藏比例对应数量的部分节点进行隐藏处理,以使每层隐层节点中没有隐藏的部分节点作为用于训练每个单词的保留节点。

4. 如权利要求3所述的装置,其特征在于,所述处理模块还用于:

根据所述训练数据中的单词数量调整所述神经网络的层数、每层节点数、以及所述预设的节点比例。

基于人工智能的字母音素转换模型生成方法及装置

技术领域

[0001] 本申请涉及人工智能技术领域,尤其涉及一种基于人工智能的字母音素转换模型生成方法及装置。

背景技术

[0002] 人工智能(Artificial Intelligence),英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语音识别、图像识别、自然语言处理和专家系统等。其中,人工智能最重要的方面就是语音识别技术。

[0003] 字母到音素的转换模型(grapheme-to-phoneme,g2p)是英文语音合成中十分重要的处理模块,用于对接收到的单词中的字母转换成对应音素。现有技术中,通过深度神经网络技术来训练g2p模型,能够得到比基于统计语言模型更好的应用效果。

[0004] 由于g2p模型的目标是将单词转换成对应的音素序列,而实际存在的单词数量也是相对固定的,也就是10万个单词左右。但是,g2p训练模型所需的数据量是相对固定的。通过深度神经网络来训练g2p模型的过程完全不同于声学模型的训练,具体如下:

[0005] 通过神经网络训练声学模型时,由于可以不断地增加训练数据,因此神经网络的层数以及每一层的单元数可以不断增大,以此带来更好的性能。但对于g2p模型的训练来讲,如果增大网络层数以及每一层的单元数,但训练数据量还是相对不变的话,很容易造成过拟合现象的发生。所谓过拟合现象就是指,训练出来的g2p模型在训练数据上表现很好,但在测试数据上却表现得不如训练数据那样好。

[0006] 然而,如果选择一个较小的网络来训练g2p模型,可以得到一个性能相对可以接受的g2p模型,但这样的网络的层数以及每一层的单元数都会相对较小,不如深层神经网络那样拥有更强的学习能力和泛化能力。

发明内容

[0007] 本申请旨在至少在一定程度上解决相关技术中的技术问题之一。

[0008] 为此,本申请的第一个目的在于提出一种基于人工智能的字母音素转换模型生成方法,该方法应用节点动态隐藏的深度学习神经网络训练数据,生成字母音素转换模型,提高了字母音素转换模型的学习能力和泛化能力,避免了过拟合现象。

[0009] 本申请的第二个目的在于提出一种基于人工智能的字母音素转换模型生成装置。

[0010] 为达上述目的,本申请第一方面实施例提出了一种基于人工智能的字母音素转换模型生成方法,包括:在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;对所有子神经网络各保留节点的相关权重进行均值处

理,生成字母音素转换模型。

[0011] 本申请实施例的基于人工智能的字母音素转换模型生成方法,通过在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。由此,实现了应用节点动态隐藏的深度学习神经网络训练数据,生成字母音素转换模型,提高了字母音素转换模型的学习能力和泛化能力,并且避免了过拟合现象。

[0012] 为达上述目的,本申请第二方面实施例提出了一种基于人工智能的字母音素转换模型生成装置,包括:处理模块,用于在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;训练模块,用于应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;生成模块,用于对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。

[0013] 本申请实施例的基于人工智能的字母音素转换模型生成装置,通过在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。由此,实现了应用节点动态隐藏的深度学习神经网络训练数据,生成字母音素转换模型,提高了字母音素转换模型的学习能力和泛化能力,并且避免了过拟合现象。

附图说明

[0014] 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0015] 图1是本申请一个实施例的基于人工智能的字母音素转换模型生成方法的流程图;

[0016] 图2是三层神经网络示意图;

[0017] 图3是加入防止过拟合技术训练的神经网络示意图;

[0018] 图4是本申请一个实施例的基于人工智能的字母音素转换模型生成装置的结构示意图。

具体实施方式

[0019] 下面详细描述本申请的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本申请,而不能理解为对本申请的限制。

[0020] 下面参考附图描述本申请实施例的基于人工智能的字母音素转换模型生成方法及装置。

[0021] 图1是本申请一个实施例的基于人工智能的字母音素转换模型生成方法的流程图

图。

[0022] 如图1所示,该基于人工智能的字母音素转换模型生成方法包括:

[0023] 步骤101,在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点。

[0024] 步骤102,应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重。

[0025] 步骤103,对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。

[0026] 具体地,为了使得训练出的g2p模型能够拥有更强的泛化能力,本发明采用了深度神经网络来训练得到g2p模型。并且在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,通过对网络中部分节点的隐藏来防止模型过拟合。

[0027] 需要解释的是,应用节点动态隐藏的神经网络训练数据,生成字母音素转换模型,并且防止生成的字母音素转换模型出现过拟合现象。其中,用于训练的神经网络为深度神经网络,在不同的训练场景中可以采用不同的模型,本实施例对此不作限制。例如:可以采用时间递归神经网络(Long-Short Term Memory, LSTM)等。

[0028] 深度神经网络包括三层:第一层为输入层、最后一层为输出层、中间的为隐层,有的神经网络具有一个隐层,有的神经网络具有多个隐层,每层都有节点。针对训练数据中单词量的大小,可以适应调整用于训练的深度神经网络的结构大小,具体包括调整深度神经网络结构的层数,以及每层的节点数量。

[0029] 为了防止模型过拟合,本发明在每一次对输入单词的训练过程中,按照预设的节点比例对训练的神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点。

[0030] 需要强调的是,节点比例反映出训练过程中对深度神经网络结构的简化程度,在训练之前,可以根据训练数据中的单词数量调整神经网络的层数、每层节点数、以及预设的节点比例。

[0031] 需要注意的是,在不同的应用场景中可以采用不同的节点比例设置方式随机的对训练的神经网络的隐层节点进行筛选,获取用于训练每个单词的保留节点,举例说明如下:

[0032] 第一种示例:节点隐藏比例

[0033] 具体地,根据预设的节点隐藏比例,随机将神经网络中每层隐层节点中与节点隐藏比例对应数量的部分节点进行隐藏处理。进而,将每层隐层节点中没有隐藏的部分节点作为用于训练每个单词的保留节点。

[0034] 第二种示例:节点保留比例

[0035] 根据预设的节点保留比例,随机将神经网络中每层隐层节点中与节点保留比例对应数量的部分节点进行保留处理,作为用于训练每个单词的保留节点。

[0036] 进而,应用与保留节点对应的子神经网络训练对应的单词,更新子神经网络各保留节点的相关权重。

[0037] 也就是说,在模型训练中,不使用隐藏的节点(非保留节点),但是会保留它们所对应的权重,只是在这次训练中不再更新这些权重。当下次训练时,又重新随机地按照一定比例隐藏隐层的一些节点。如此周而复始。这种策略之所以能够达到抗过拟合的原因,就是因

为,每次训练的时候,在隐藏一些节点后,实际上剩下的网络是一个相对原网络较小的网络,这样的网络基本不会过拟合。

[0038] 最后,对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。

[0039] 可以理解的是,而每次训练时都随机地隐藏不同的节点实际上就是得到了很多小的网络,最后的结果可以认为是对这些小网络取平均的方式,而这是一种很有效的抗过拟合的方式。需要注意的是:上面所述的隐藏节点,不是真的删除,只是在当前训练时不使用这些节点,而在下次训练时,还会使用。

[0040] 为了更加清楚的说明上述过程,通过图2和图3为例说明如下:

[0041] 图2是三层神经网络示意图,如图2所示,输入层为3个节点,隐层为6个节点,输出层为2个节点。

[0042] 图3是加入防止过拟合技术训练的神经网络示意图。参见图3,图3中用虚线画的节点表示的是该次训练中被隐藏的节点,也就是该次训练不考虑也不更新这些节点的相关权重,因此它们对应的权重也用虚线表示。实线画的节点表示的是该次训练中被保留的节点,也就是使用保留节点对数据进行训练,并更新这些节点的相关权重。

[0043] 在利用神经网络训练数据,生成字母音素转换模型的过程中,为了更加清楚的描述使用部分节点隐藏技术生成模拟以防止过拟合,以及使用全部的节点生成模型的对比结果,通过表1的实验结果进行说明:

[0044] 表1

[0045]

模型	网络节点是否隐藏	Phoneme错误率
256-BLSTM	否	过拟合
256-BLSTM	是	9.37%
128-BLSTM+64-BLSTM+64-BLSTM	否	过拟合
128-BLSTM+64-BLSTM+64-BLSTM	是	8.81%
128-BLSTM+128-BLSTM+64-BLSTM	否	过拟合
128-BLSTM+128-BLSTM+64-BLSTM	是	8.38%

[0046] 参见表1,只给出了这些模型的隐层网络的描述,比如,针对模型128-BLSTM+64-BLSTM+64-BLSTM,就是说明一个隐层拥有3个BLSTM的神经网络模型。

[0047] 对于256-BLSTM模型,参数为64万,对于128-BLSTM+64-BLSTM+64-BLSTM模型,参数为44万,对于128-BLSTM+128-BLSTM+64-BLSTM模型,参数为74万。这些模型的参数是指节点边的权重,参数的数量对于我们的训练数据来讲,显然是多了。如果不在训练这个模型的时候隐藏部分节点,那么模型肯定会过拟合。通过隐藏神经网络中的部分节点来进行模型训练,能够很好地防止模型过拟合,并且训练出性能很好的模型。

[0048] 从上面的描述以及实验已经看出,本发明的优点如下:

[0049] 通过隐藏神经网络的部分节点训练g2p模型,使得在训练模型时可以适当增加网络层数以及每层的节点数,以此来提高模型的泛化能力并且不会使得过拟合现象的发生,从而使得将较大规模的神经网络运用在实际的g2p模型的训练上成为现实。

[0050] 本申请实施例的基于人工智能的字母音素转换模型生成方法,通过应用神经网络

络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行筛选,获取用于训练每个单词的保留节点;应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。由此,实现了应用节点动态隐藏的深度学习神经网络训练数据,生成字母音素转换模型,提高了字母音素转换模型的学习能力和泛化能力,并且避免了过拟合现象。

[0051] 为了实现上述实施例,本申请还提出一种基于人工智能的字母音素转换模型生成装置。

[0052] 图4是本申请一个实施例的基于人工智能的字母音素转换模型生成装置的结构示意图。

[0053] 如图4所示,该基于人工智能的字母音素转换模型生成装置包括:

[0054] 处理模块11,用于在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;

[0055] 具体地,所述处理模块11用于:

[0056] 按照预设的节点比例随机对所述神经网络的隐层节点进行筛选。

[0057] 在一个实施例中,所述处理模块11用于:

[0058] 根据预设的节点隐藏比例,将所述神经网络中每层隐层节点中与所述节点隐藏比例对应数量的部分节点进行隐藏处理,以使每层隐层节点中没有隐藏的部分节点作为用于训练每个单词的保留节点。

[0059] 在另一个实施例中,所述处理模块11用于:

[0060] 根据预设的节点保留比例,将所述神经网络中每层隐层节点中与所述节点保留比例对应数量的部分节点进行保留处理,作为用于训练每个单词的保留节点。

[0061] 训练模块12,用于应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;

[0062] 生成模块13,用于对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。

[0063] 基于上述实施例,进一步地,在另一个实施例中,所述处理模块11还用于:

[0064] 根据所述训练数据中的单词数量调整所述神经网络的层数、每层节点数、以及所述预设的节点比例。

[0065] 需要说明的是,前述对基于人工智能的字母音素转换模型生成方法实施例的解释说明也适用于该实施例的基于人工智能的字母音素转换模型生成装置,此处不再赘述。

[0066] 本申请实施例的基于人工智能的字母音素转换模型生成装置,通过在应用神经网络对训练数据中的每个单词进行字母音素转换训练的过程中,按照预设的节点比例对所述神经网络的隐层节点进行随机筛选,获取用于训练每个单词的保留节点;应用与所述保留节点对应的子神经网络训练对应的单词,更新所述子神经网络各保留节点的相关权重;对所有子神经网络各保留节点的相关权重进行均值处理,生成字母音素转换模型。由此,实现了应用节点动态隐藏的深度学习神经网络训练数据,生成字母音素转换模型,提高了字母音素转换模型的学习能力和泛化能力,并且避免了过拟合现象。

[0067] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征或者特点包含于本申请的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0068] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本申请的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0069] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本申请的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本申请的实施例所属技术领域的技术人员所理解。

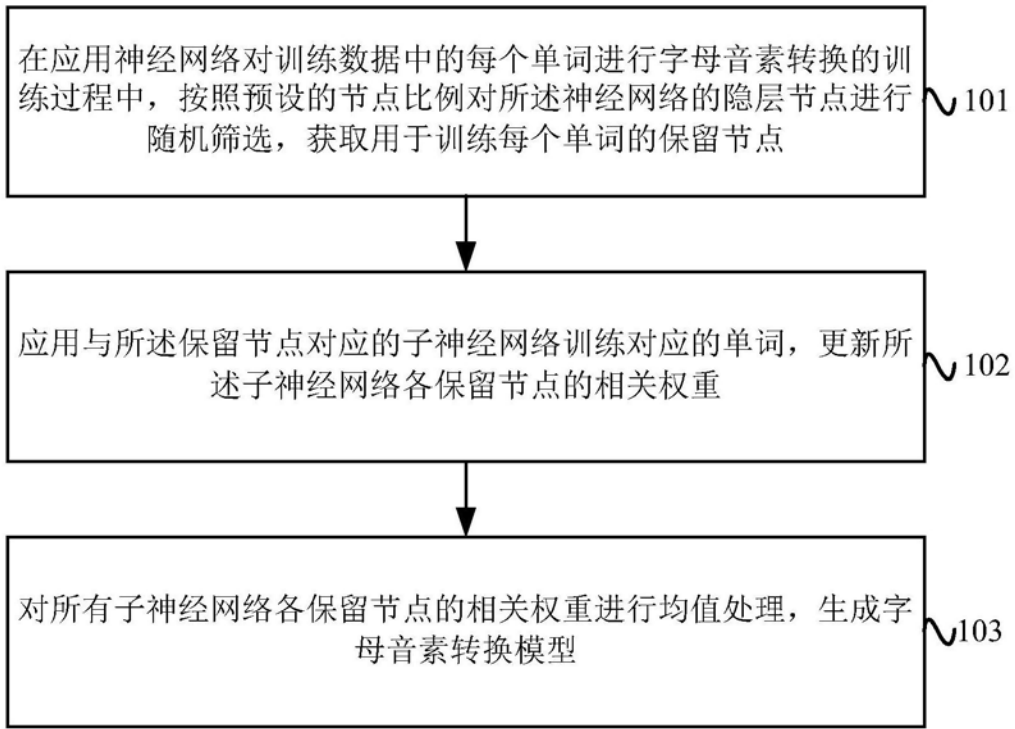


图1

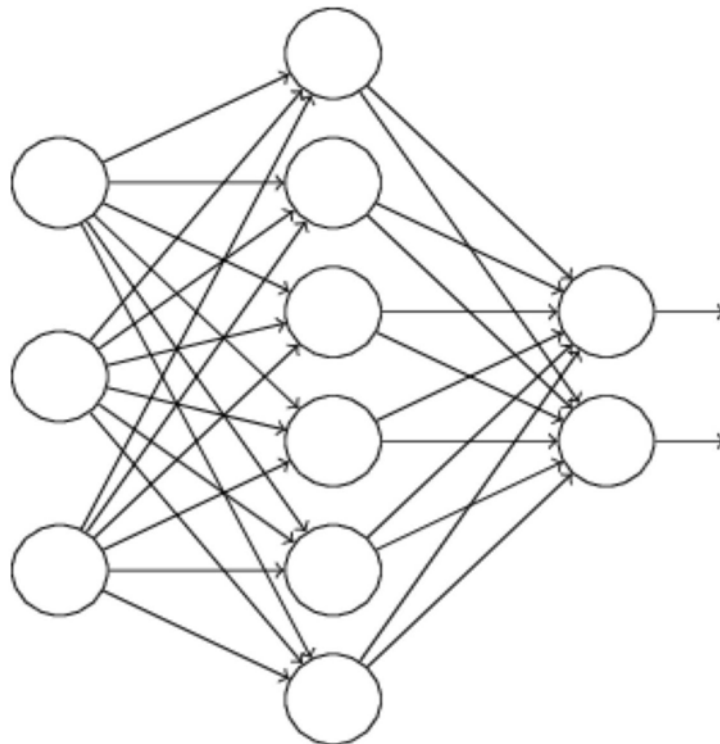


图2

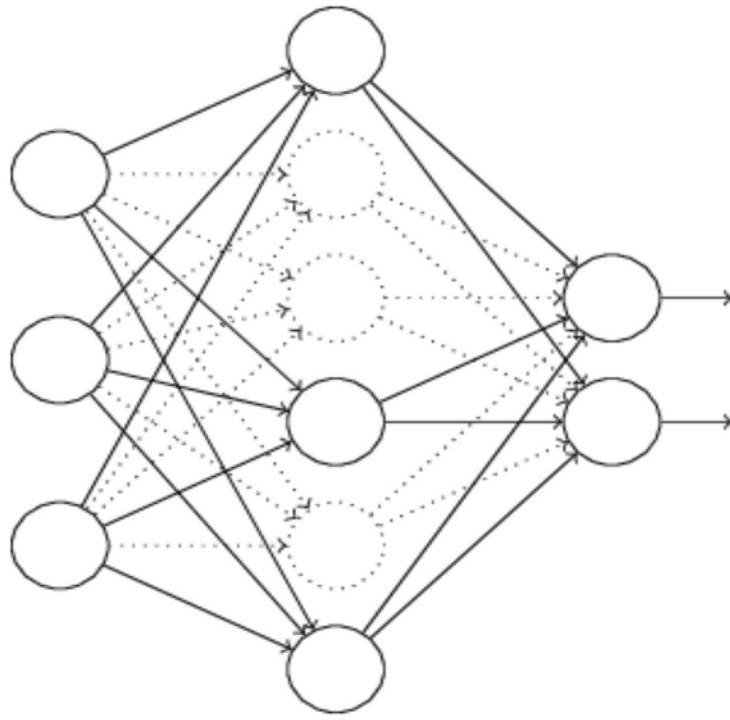


图3



图4