

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 9 月 22 日 (22.09.2016)



(10) 国际公布号
WO 2016/145850 A1

- (51) 国际专利分类号:
G10L 15/02 (2006.01) G10L 15/06 (2013.01)
- (21) 国际申请号: PCT/CN2015/092381
- (22) 国际申请日: 2015 年 10 月 21 日 (21.10.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510122982.6 2015 年 3 月 19 日 (19.03.2015) CN
- (71) 申请人: 清华大学 (TSINGHUA UNIVERSITY) [CN/CN]; 中国北京市海淀区清华大学, Beijing 100084 (CN)。
- (72) 发明人: 杨毅 (YANG, Yi); 中国北京市海淀区清华大学, Beijing 100084 (CN)。 孙甲松 (SUN, JiaSong); 中国北京市海淀区清华大学, Beijing 100084 (CN)。
- (74) 代理人: 西安智大知识产权代理事务所 (XI'AN CYDA INTELLECTUAL PROPERTY AGENCY); 中国陕西省西安市雁塔区科技路 195 号世纪颐园 A707, Shaanxi 710075 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

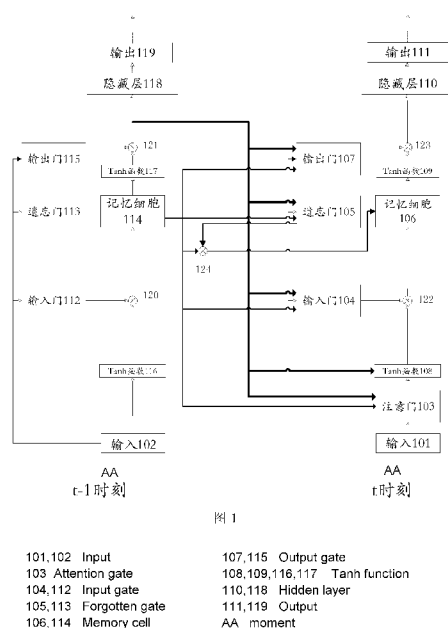
根据细则 4.17 的声明:

- 关于发明人身份(细则 4.17(i))
- 关于申请人有权申请并被授予专利(细则 4.17(ii))

[见续页]

(54) Title: CONSTRUCTION METHOD FOR DEEP LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK ACOUSTIC MODEL BASED ON SELECTIVE ATTENTION PRINCIPLE

(54) 发明名称: 基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法



(57) Abstract: A construction method for a deep long short-term memory recurrent neural network acoustic model based on a selective attention principle. Change of an instant function of neurons of an auditory cortex is represented by adding an attention gate (103) unit in the deep long short-term memory recurrent neural network acoustic model, and the attention gate (103) unit is different from other gate units in that: the other gate units correspond to a time sequence on a one-to-one basis, but the attention gate (103) unit shows a short-term plasticity effect, thereby having intervals on the time sequence; extraction of robust features about Cross-talk noise and construction of a robust acoustic model are realized via the recurrent neural network acoustic model obtained by training a large amount of voice data containing the Cross-talk noise, and the purpose of increasing the robustness about the acoustic model can be achieved by restraining the influence of a non-target stream against the extraction of the features; the method can be extensively applied to the field of a plurality of machine learning related to speaker recognition and keyword recognition in voice recognition, human-machine interaction and the like.

(57) 摘要: 一种基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法, 通过在深度长短期记忆循环神经网络声学模型中增加注意门 (103) 单元, 来表征听觉皮层神经元的瞬时功能改变, 注意门 (103) 单元与其他门单元不同之处在于, 其他门单元与时间序列一一对应, 而注意门 (103) 单元体现的是短期可塑性效应, 因此在时间序列上存在间隔; 通过对包含 Cross-talk 噪声的大量语音数据进行训练获得的上述神经网络声学模型, 可以实现对 Cross-talk 噪声的鲁棒特征提取和鲁棒声学模型的构建, 通过抑制非目标流对特征提取的影响可以达到提高声学模型的鲁棒性的目的; 该方法可广泛应用于涉及语音识别的说话人识别、关键词识别、人机交互等多种机器学习领域。



-
- 关于申请人有权要求在先申请的优先权(细则 4.17(iii))
- 发明人资格(细则 4.17(iv))
- 本国际公布:**
- 包括国际检索报告(条约第 21 条(3))。

基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法

技术领域

本发明属于音频技术领域，特别涉及一种基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法。

背景技术

随着信息技术的迅速发展，语音识别技术已经具备大规模商业化的条件。目前语音识别主要采用基于统计模型的连续语音识别技术，其主要目标是通过给定的语音序列寻找其所代表的概率最大的词序列。基于统计模型的连续语音识别系统的任务是根据给定的语音序列寻找其所代表的概率最大的词序列，通常包括构建声学模型和语言模型及其对应的搜索解码方法。随着声学模型和语言模型的快速发展，语音识别系统的性能在理想声学环境下已经大为改善，现有的深度神经网络-隐马尔科夫模型（Deep Neural Network-Hidden Markov Model, DNN-HMM）初步成熟，通过机器学习的方法可以自动提取有效特征，并能对多帧语音对应的上下文信息建模，但是此类模型每一层都有百万量级的参数，且下一层的输入是上一次的输出，因此需要使用 GPU 设备来训练 DNN 声学模型，训练时间长；高度非线性以及参数共享的特性也使得 DNN 难以进行参数自适应。

循环神经网络（Recurrent Neural Network, RNN）是一种单元之间存在有向循环来表达网络内部动态时间特性的神经网络，在手写体识别和语言模型等方面得到广泛应用。语音信号是复杂的时变信号，在不同时间尺度上具有复杂的相关性，因此相比于深度神经网络而言，循环神经网络具有的循环连接功能更适合处理这类复杂时序数据。

作为循环神经网络的一种，长短期记忆（Long Short-Term Memory, LSTM）模型比循环神经网络更适合处理和预测事件滞后且时间不定的长时

序列。多伦多大学提出的增加了记忆模块(memory block)的深度 LSTM-RNN 声学模型则将深度神经网络的多层次表征能力与循环神经网络灵活利用长跨度上下文的能力结合,使得基于 TIMIT 库的音素识别错误率降至 17.1%。

但是循环神经网络中使用的梯度下降法存在梯度消散(vanishing gradient)问题,也就是在对网络的权重进行调整的过程中,随着网络层数增加,梯度逐层消散,致使其对权重调整的作用越来越小。谷歌提出的两层深度 LSTM-RNN 声学模型,在以前的深度 LSTM-RNN 模型中增加了线性循环投影层(Recurrent Projection Layer),用于解决梯度消散问题。对比实验表明,RNN 的帧正确率(Frame Accuracy)及其收敛速度明显逊于 LSTM-RNN 和 DNN;在词错误率及其收敛速度方面,最好的 DNN 在训练数周后的词错误率为 11.3%;而两层深度 LSTM-RNN 模型在训练 48 小时后词错误率降低至 10.9%,训练 100/200 小时后,词错误率降低至 10.7/10.5(%)。

慕尼黑大学提出的深度双向长短期记忆循环神经网络(Deep Bidirectional Long Short-Term Memory Recurrent Neural Networks, DBLSTM-RNN)声学模型,在神经网络的每个循环层中定义了相互独立的前向层和后向层,并使用多隐藏层对输入的声学特征进行更高层表征,同时对噪声和混响进行有监督学习实现特征投影和增强。此方法在 2013 PASCAL CHiME 数据集上,在信噪比[-6dB, 9dB]范围内实现了词错误率从基线的 55%降低到 22%。

但实际声学环境的复杂性仍然严重影响和干扰连续语音识别系统的性能,即使利用目前主流的 DNN 声学模型方法,在包括噪声、音乐、口语、重复等复杂环境条件下的连续语音识别数据集上也只能获得 70%左右的识别率,连续语音识别系统中声学模型的抗噪性和鲁棒性仍有待改进。

随着声学模型和语言模型的快速发展,语音识别系统的性能在理想声学环境下已经大为改善,现有的 DNN-HMM 模型初步成熟,通过机器学习的方法可以自动提取有效特征,并能对多帧语音对应的上下文信息建模。然而大多数识别系统对于声学环境的改变仍然十分敏感,特别是在 cross-talk 噪声

（两人或多人同时说话）干扰下不能满足实用性能的要求。与深度神经网络声学模型相比，循环神经网络声学模型中的单元之间存在有向循环，可以有效地描述神经网络内部的动态时间特性，更适合处理具有复杂时序的语音数据。而长短期记忆神经网络比循环神经网络更适合处理和预测事件滞后且时间不定的长时序列，因此用于构建语音识别的声学模型能够取得更好的效果。

人脑在处理复杂场景的语音时存在选择性注意的现象，其主要原理为：人脑具有听觉选择性注意的能力，在听觉皮层区域通过自上而下的控制机制，来实现抑制非目标流和增强目标流的目的。研究表明，在选择性注意的过程中，听觉皮层的短期可塑性（Short-Term Plasticity）效应增加了对声音的区分能力。在注意力非常集中时，在初级听觉皮层可以在 50 毫秒内开始对声音目标进行增强处理。

发明内容

为了克服上述现有技术的缺点，本发明的目的在于提供一种基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法，建立了基于选择性注意原理的深度长短期记忆循环神经网络声学模型，通过在深度长短期记忆循环神经网络声学模型中增加注意门单元，来表征听觉皮层神经元的瞬时功能改变，注意门单元与其他门单元不同之处在于，其他门单元与时间序列一一对应，而注意门单元体现的是短期可塑性效应，因此在时间序列上存在间隔。通过对包含 Cross-talk 噪声的大量语音数据进行训练获得的上述神经网络声学模型，可以实现对 Cross-talk 噪声的鲁棒特征提取和鲁棒声学模型的构建，通过抑制非目标流对特征提取的影响可以达到提高声学模型的鲁棒性的目的。

为了实现上述目的，本发明采用的技术方案是：

一种基于选择性注意原理的连续语音识别方法，包括如下步骤：

第一步，构建基于选择性注意原理的深度长短期记忆循环神经网络

从输入到隐藏层定义为一个长短期记忆循环神经网络，深度指的是每个长短期记忆循环神经网络的输出为下一个长短期记忆循环神经网络的输入，

如此重复，最后一个长短期记忆循环神经网络的输出作为整个系统的输出；在每一个长短期记忆循环神经网络中，语音信号 x_t 为 t 时刻的输入， x_{t-1} 为 $t-1$ 时刻的输入，以此类推，总时间长度上的输入为 $x=[x_1,...,x_T]$ 其中 $t \in [1,T]$ ， T 为语音信号的总时间长度； t 时刻的长短期记忆循环神经网络由注意门、输入门、输出门、遗忘门、记忆细胞、 $\tanh$ 函数、隐藏层、乘法器组成， $t-1$ 时刻的长短期记忆循环神经网络由输入门、输出门、遗忘门、记忆细胞、 $\tanh$ 函数、隐藏层、乘法器组成；总时间长度上的隐藏层输出为 $y=[y_1,...,y_T]$ ；

第二步，构建基于选择性注意原理的深度长短期记忆循环神经网络声学模型

在第一步的基础上，每间隔 s 时刻对应的深度长短期记忆循环神经网络存在注意门，其他时刻的深度长短期记忆循环神经网络不存在注意门，即，基于选择性注意原理的深度长短期记忆循环神经网络声学模型由间隔存在注意门的深度长短期记忆循环神经网络组成。

如何在复杂环境干扰，特别是在 cross-talk 噪声干扰下进行识别，一直是语音识别的难点之一，阻碍了语音识别的大规模应用。与现有技术相比，本发明借鉴人脑在处理复杂场景的语音时存在选择性注意的现象来实现抑制非目标流和增强目标流，通过在深度长短期记忆递归神经网络声学模型中增加注意门单元，来表征听觉皮层神经元的瞬时功能改变，注意门单元与其他门单元不同之处在于，其他门单元与时间序列一一对应，而注意门单元体现的是短期可塑性效应，因此在时间序列上存在间隔。在一些包含 Cross-talk 噪声的连续语音识别数据集上采用这种方法，可以获得比深度神经网络方法更好的性能。

附图说明

图 1 是本发明的基于选择性注意原理的深度长短期记忆循环神经网络流程图。

图 2 是本发明的基于选择性注意原理的深度长短期记忆神经网络声学模

型流程图。

具体实施方式

下面结合附图和实施例详细说明本发明的实施方式。

本发明利用基于选择性注意原理的深度长短期记忆循环神经网络声学模型，实现了连续语音识别。但本发明提供的模型及方法不局限于连续语音识别，也可以是任何与语音识别有关的方法和装置。

本发明主要包括如下步骤：

第一步，构建基于选择性注意原理的深度长短期记忆循环神经网络

如图 1 所示，输入 **101** 和输入 **102** 为 t 时刻和 $t-1$ 时刻语音信号输入 x_t 和 x_{t-1} ($t \in [1, T]$, T 为语音信号的总时间长度)； t 时刻的长短期记忆循环神经网络由注意门 **103**、输入门 **104**、遗忘门 **105**、记忆细胞 **106**、输出门 **107**、tanh 函数 **108**、tanh 函数 **109**、隐藏层 **110**、乘法器 **122** 以及乘法器 **123** 组成； $t-1$ 时刻的长短期记忆循环神经网络由输入门 **112**、遗忘门 **113**、记忆细胞 **114**、输出门 **115**、tanh 函数 **116**、tanh 函数 **117**、隐藏层 **118**、乘法器 **120** 以及乘法器 **121** 组成。 t 时刻和 $t-1$ 时刻隐藏层输出分别为输出 **111** 和输出 **119**。

其中，输入 **102** 同时作为输入门 **112**、遗忘门 **113**、输出门 **115** 以及 tanh 函数 **116** 的输入，输入门 **112** 的输出与 tanh 函数 **116** 的输出送入乘法器 **120**，运算后的输出作为记忆细胞 **114** 的输入，记忆细胞 **114** 的输出作为 tanh 函数 **117** 的输入，tanh 函数 **117** 的输出和输出门 **115** 的输出送入乘法器 **121**，运算后的输出作为隐藏层 **118** 的输入，隐藏层 **118** 的输出即为输出 **119**。

输入 **101**、记忆细胞 **114** 的输出以及乘法器 **121** 的输出共同作为注意门 **103** 的输入，注意门 **103** 的输出和乘法器 **121** 的输出共同作为 tanh 函数 **108** 的输入，注意门 **103** 的输出、记忆细胞 **114** 的输出和乘法器 **121** 的输出分别共同作为输入门 **104**、遗忘门 **105** 以及输出门 **107** 的输入，遗忘门 **105** 的输出和记忆细胞 **114** 的输出送入乘法器 **124**，输入门 **104** 的输出与 tanh 函数 **108** 的输出送入乘法器 **122**，乘法器 **124** 的输出和乘法器 **122** 的输出作为记忆细胞 **106** 的输入，记忆

细胞 **106** 的输出作为 tanh 函数 **109** 的输入, tanh 函数 **109** 的输出和输出门 **107** 的输出送入乘法器 **123**, 乘法器 **123** 的输出作为隐藏层 **110** 的输入, 隐藏层 **110** 的输出即为输出 **111**。

即: 在 $t \in [1, T]$ 时刻的参数按照如下公式计算:

$$\begin{aligned}
 G_{\text{atten_}t} &= \text{sigmoid}(W_{ax} x_t + W_{am} m_{t-1} + W_{ac} \text{Cell}_{t-1} + b_a) \\
 G_{\text{input_}t} &= \text{sigmoid}(W_{ia} G_{\text{atten_}t} + W_{im} m_{t-1} + W_{ic} \text{Cell}_{t-1} + b_i) \\
 G_{\text{forget_}t} &= \text{sigmoid}(W_{fa} G_{\text{atten_}t} + W_{fm} m_{t-1} + W_{fc} \text{Cell}_{t-1} + b_f) \\
 \text{Cell}_t &= G_{\text{forget_}t} \odot \text{Cell}_{t-1} + G_{\text{input_}t} \odot \tanh(W_{ca} G_{\text{atten_}t} + W_{cm} m_{t-1} + b_c) \\
 G_{\text{output_}t} &= \text{sigmoid}(W_{oa} G_{\text{atten_}t} + W_{om} m_{t-1} + W_{oc} \text{Cell}_{t-1} + b_o) \\
 m_t &= G_{\text{output_}t} \odot \tanh(\text{Cell}_t) \\
 y_t &= \text{soft max}_k (W_{ym} m_t + b_y)
 \end{aligned}$$

其中 $G_{\text{atten_}t}$ 为 t 时刻注意门 **103** 的输出, $G_{\text{input_}t}$ 为 t 时刻输入门 **104** 的输出, $G_{\text{forget_}t}$ 为 t 时刻遗忘门 **105** 的输出, Cell_t 为 t 时刻记忆细胞 **106** 的输出, $G_{\text{output_}t}$ 为 t 时刻输出门 **107** 的输出, m_t 为 t 时刻隐藏层 **110** 的输入, y_t 为 t 时刻的输出 **111**; x_t 为 t 时刻的输入 **101**, m_{t-1} 为 $t-1$ 时刻隐藏层 **118** 的输入, Cell_{t-1} 为 $t-1$ 时刻记忆细胞 **114** 的输出; w_{ax} 为 t 时刻注意门 a 与 t 时刻输入 x 之间的权重, w_{am} 为 t 时刻注意门 a 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{ac} 为 t 时刻注意门 a 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{ia} 为 t 时刻输入门 i 与 t 时刻注意门 a 之间的权重, w_{im} 为 t 时刻输入门 i 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{ic} 为 t 时刻输入门 i 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{fa} 为 t 时刻遗忘门 f 与 t 时刻注意门 a 之间的权重, w_{fm} 为 t 时刻遗忘门 f 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{fc} 为 t 时刻遗忘门 f 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{ca} 为 t 时刻记忆细胞 c 与 t 时刻注意门 a 之间的权重, w_{cm} 为 t 时刻记忆细胞 c 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{oa} 为 t 时刻输出门 o 与 t 时刻注意门 a 之间的权重, w_{om} 为 t 时刻输出门 o 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{oc} 为 t 时刻输出门 o 与 $t-1$ 时刻

记忆细胞 c 之间的权重； b_a 为注意门 a 的偏差量， b_i 为输入门 i 的偏差量， b_f 为遗忘门 f 的偏差量， b_c 为记忆细胞 c 的偏差量， b_o 为输出门 o 的偏差量， b_y 为输出 y 的偏差量，不同的 b 代表不同的偏差量；且有 $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$ ，

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \text{soft max}_k(x) = \frac{e^{x_k}}{\sum_{l=1}^K e^{x_l}}, \quad \text{其中 } x_k \text{ 表示第 } k \in [1, K] \text{ 个 soft max 函数的输入, } l \in [1, K] \text{ 用于对全部 } e^{x_k} \text{ 求和; } \odot \text{ 代表矩阵元素相乘。}$$

第二步，构建基于选择性注意原理的深度长短期记忆循环神经网络声学模型

在第一步的基础上，每间隔 s ($s=5$) 时刻对应的深度长短期记忆循环神经网络存在注意门，其他时刻的深度长短期记忆循环神经网络不存在注意门，即，基于选择性注意原理的深度长短期记忆循环神经网络声学模型由间隔存在注意门的深度长短期记忆循环神经网络组成。如图 2 所示为所建立的基于选择性注意原理的深度长短期记忆循环神经网络声学模型， t 时刻的深度长短期记忆循环神经网络存在注意门 **201**， $t-s$ 时刻的深度长短期记忆循环神经网络存在注意门 **202**，如此循环。

1、一种基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法，包括如下步骤：

第一步，构建基于选择性注意原理的深度长短期记忆循环神经网络

从输入到隐藏层定义为一个长短期记忆循环神经网络，深度指的是每个长短期记忆循环神经网络的输出为下一个长短期记忆循环神经网络的输入，如此重复，最后一个长短期记忆循环神经网络的输出作为整个系统的输出；在每一个长短期记忆循环神经网络中，语音信号 x_t 为 t 时刻的输入， x_{t-1} 为 $t-1$ 时刻的输入，以此类推，总时间长度上的输入为 $x = [x_1, \dots, x_T]$ 其中 $t \in [1, T]$ ， T 为语音信号的总时间长度； t 时刻的长短期记忆循环神经网络由注意门、输入门、输出门、遗忘门、记忆细胞、 $\tanh$ 函数、隐藏层、乘法器组成， $t-1$ 时刻的长短期记忆循环神经网络由输入门、输出门、遗忘门、记忆细胞、 $\tanh$ 函数、隐藏层、乘法器组成；总时间长度上的隐藏层输出为 $y = [y_1, \dots, y_T]$ ；

在 $t \in [1, T]$ 时刻的参数按照如下公式计算：

$$G_{\text{atten}_t} = \text{sigmoid}(W_{ax} x_t + W_{am} m_{t-1} + W_{ac} \text{Cell}_{t-1} + b_a)$$

$$G_{\text{input}_t} = \text{sigmoid}(W_{ia} G_{\text{atten}_t} + W_{im} m_{t-1} + W_{ic} \text{Cell}_{t-1} + b_i)$$

$$G_{\text{forget}_t} = \text{sigmoid}(W_{fa} G_{\text{atten}_t} + W_{fm} m_{t-1} + W_{fc} \text{Cell}_{t-1} + b_f)$$

$$\text{Cell}_t = G_{\text{forget}_t} \odot \text{Cell}_{t-1} + G_{\text{input}_t} \odot \tanh(W_{ca} G_{\text{atten}_t} + W_{cm} m_{t-1} + b_c)$$

$$G_{\text{output}_t} = \text{sigmoid}(W_{oa} G_{\text{atten}_t} + W_{om} m_{t-1} + W_{oc} \text{Cell}_{t-1} + b_o)$$

$$m_t = G_{\text{output}_t} \odot \tanh(\text{Cell}_t)$$

$$y_t = \text{soft max}_k(W_{ym} m_t + b_y)$$

其中 G_{atten_t} 为 t 时刻注意门的输出， G_{input_t} 为 t 时刻输入门的输出， G_{forget_t} 为 t 时刻遗忘门的输出， Cell_t 为 t 时刻记忆细胞的输出， G_{output_t} 为 t 时刻输出门的输出， m_t 为 t 时刻隐藏层的输入， y_t 为 t 时刻的输出； x_t 为 t 时刻的输入， m_{t-1} 为 $t-1$ 时刻隐藏层的输入， Cell_{t-1} 为 $t-1$ 时刻记忆细胞的输出； w_{ax} 为 t 时刻注意门

a 与 t 时刻输入 x 之间的权重, w_{am} 为 t 时刻注意门 a 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{ac} 为 t 时刻注意门 a 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{ia} 为 t 时刻输入门 i 与 t 时刻注意门 a 之间的权重, w_{im} 为 t 时刻输入门 i 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{ic} 为 t 时刻输入门 i 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{fa} 为 t 时刻遗忘门 f 与 t 时刻注意门 a 之间的权重, w_{fm} 为 t 时刻遗忘门 f 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{fc} 为 t 时刻遗忘门 f 与 $t-1$ 时刻记忆细胞 c 之间的权重, w_{ca} 为 t 时刻记忆细胞 c 与 t 时刻注意门 a 之间的权重, w_{cm} 为 t 时刻记忆细胞 c 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{oa} 为 t 时刻输出门 o 与 t 时刻注意门 a 之间的权重, w_{om} 为 t 时刻输出门 o 与 $t-1$ 时刻隐藏层输入 m 之间的权重, w_{oc} 为 t 时刻输出门 o 与 $t-1$ 时刻记忆细胞 c 之间的权重; b_a 为注意门 a 的偏差量, b_i 为输入门 i 的偏差量, b_f 为遗忘门 f 的偏差量, b_c 为记忆细胞 c 的偏差量, b_o 为输出门 o 的偏差量, b_y 为输出 y 的偏差量, 不同的 b 代表不同的偏差量; 且有 $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$, $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, $\text{soft max}_k(x) = \frac{e^{x_k}}{\sum_{l=1}^K e^{x_l}}$, 其中 x_k 表示第 $k \in [1, K]$ 个 soft max 函数的输入, $l \in [1, K]$ 用于对全部 e^{x_k} 求和; $\odot$ 代表矩阵元素相乘;

第二步, 构建基于选择性注意原理的深度长短期记忆循环神经网络声学模型

在第一步的基础上, 每间隔 s 时刻对应的深度长短期记忆循环神经网络存在注意门, 其他时刻的深度长短期记忆循环神经网络不存在注意门, 即, 基于选择性注意原理的深度长短期记忆循环神经网络声学模型由间隔存在注意门的深度长短期记忆循环神经网络组成。

2、根据权利要求1所述基于选择性注意原理的深度长短期记忆循环神经网络声学模型的构建方法, 其特征在于, 所述 $s=5$ 。

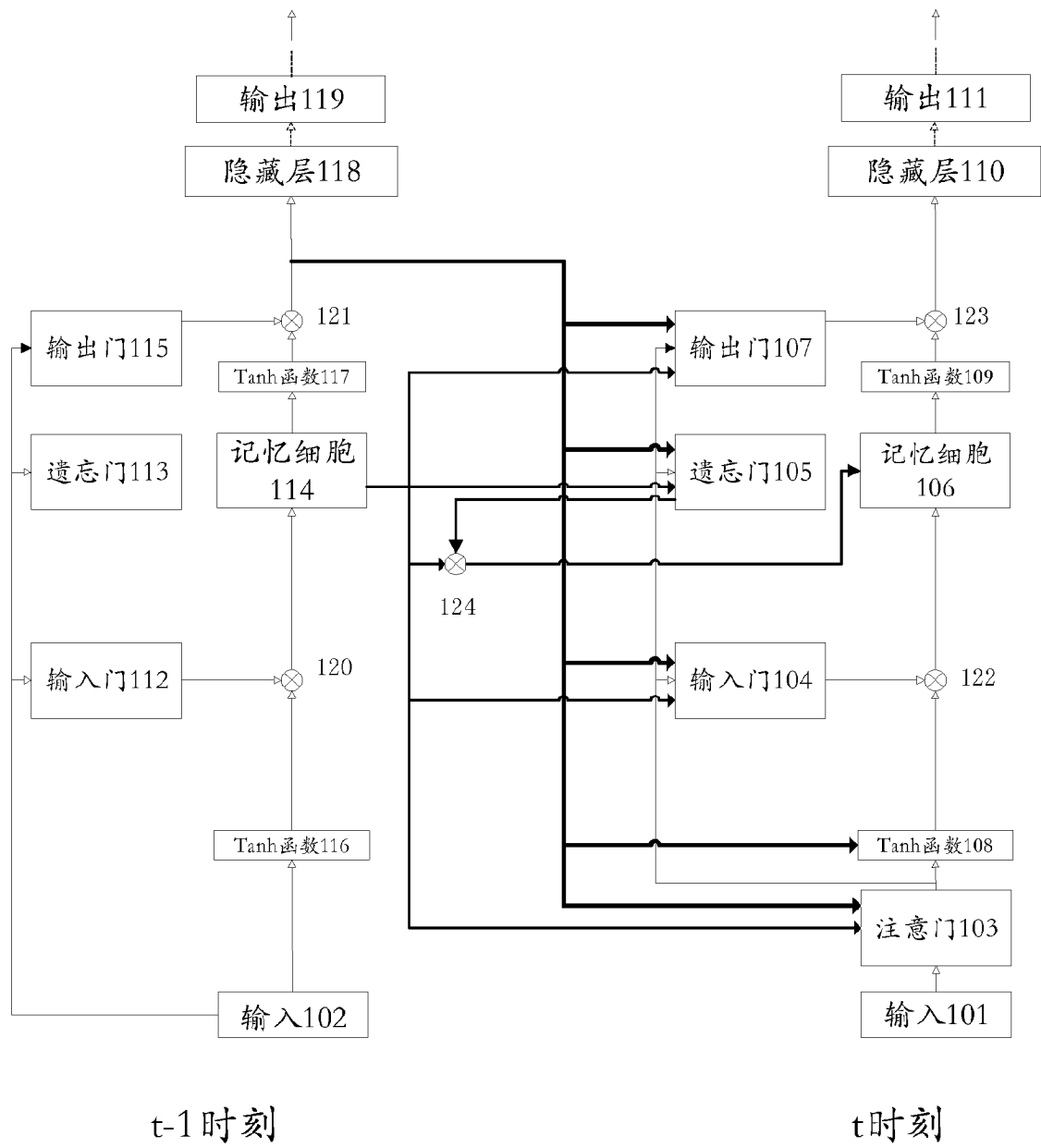


图 1

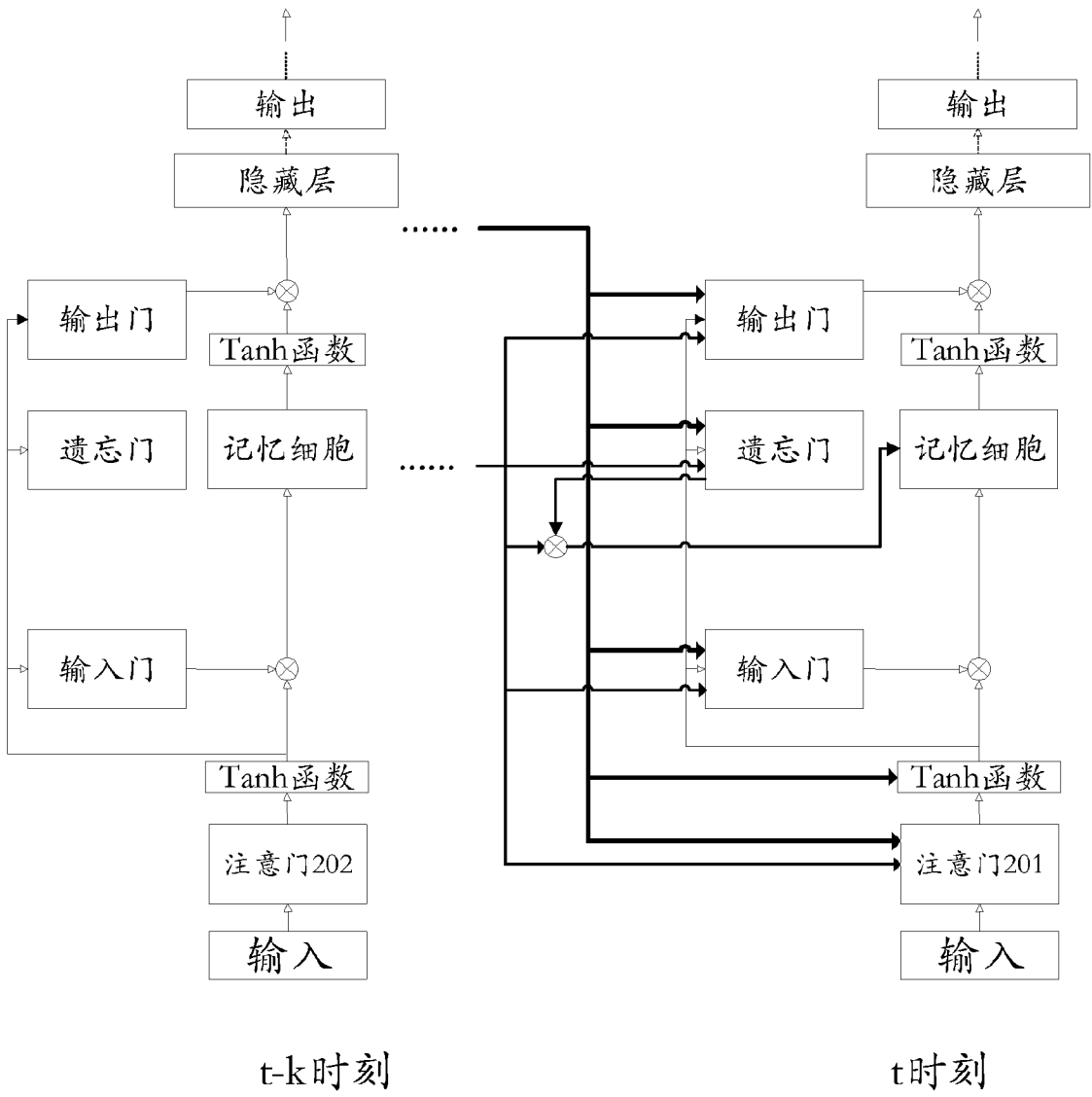


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/092381

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02 (2006.01) i; G10L 15/06 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: DNN, depth nerve network, memory, circulation, recurrent neural network, long short term, LSTM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104700828 A (TSINGHUA UNIVERSITY), 10 June 2015 (10.06.2015), claims 1-2	1-2
A	CN 103049792 A (MICROSOFT CORPORATION), 17 April 2013 (17.04.2013), description, paragraphs 15-51, and figures 1-3	1-2
A	CN 104217226 A (TIANJIN UNIVERSITY), 17 December 2014 (17.12.2014), the whole document	1-2
A	CN 103680496 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 26 March 2014 (26.03.2014), the whole document	1-2
A	US 2008172349 A1 (TOYOTA ENGINEERING & MANUFACTURING NORTH AMERICA, INC.), 17 July 2008 (17.07.2008), the whole document	1-2
A	CN 102983819 A (NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS), 20 March 2013 (20.03.2013), the whole document	1-2

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 30 December 2015 (30.12.2015)	Date of mailing of the international search report 01 March 2016 (01.03.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer ZHANG, Feixian Telephone No.: (86-10) 010-62413601

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/092381

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104700828 A	10 June 2015	None	
CN 103049792 A	17 April 2013	US 2013138436 A1	30 May 2013
CN 104217226 A	17 December 2014	None	
CN 103680496 A	26 March 2014	None	
US 2008172349 A1	17 July 2008	JP 2008204437 A	04 September 2008
		US 7647284 B2	12 January 2010
		JP 5102046 B2	19 December 2012
CN 102983819 A	20 March 2013	CN 102983819 B	04 February 2015

国际检索报告

国际申请号

PCT/CN2015/092381

A. 主题的分类		
G10L 15/02 (2006.01)i; G10L 15/06 (2013.01)i		
按照国际专利分类 (IPC) 或者同时按照国家分类和IPC两种分类		
B. 检索领域		
检索的最低限度文献 (标明分类系统和分类号)		
G10L 15/-		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))		
CNPAT, WPI, EPODOC, CNKI: 深度神经网络, 记忆, 循环神经网络, 长短期, DNN, depth nerve network, memory, circulation, recurrent neural network, long short term, LSTM		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104700828 A (清华大学) 2015年 6月 10日 (2015 - 06 - 10) 权利要求1-2	1-2
A	CN 103049792 A (微软公司) 2013年 4月 17日 (2013 - 04 - 17) 说明书第15-51段, 图1-3	1-2
A	CN 104217226 A (天津大学) 2014年 12月 17日 (2014 - 12 - 17) 全文	1-2
A	CN 103680496 A (百度在线网络技术北京有限公司) 2014年 3月 26日 (2014 - 03 - 26) 全文	1-2
A	US 2008172349 A1 (TOYOTA ENGINEERING & MANUFACTURING NORTH AMERICA, INC.) 2008年 7月 17日 (2008 - 07 - 17) 全文	1-2
A	CN 102983819 A (南京航空航天大学) 2013年 3月 20日 (2013 - 03 - 20) 全文	1-2
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期		国际检索报告邮寄日期
2015年 12月 30日		2016年 3月 1日
ISA/CN的名称和邮寄地址		受权官员
中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088		张飞弦
传真号 (86-10) 62019451		电话号码 (86-10) 010-62413601

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/092381

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104700828	A	2015年 6月 10日	无			
CN	103049792	A	2013年 4月 17日	US	2013138436	A1	2013年 5月 30日
CN	104217226	A	2014年 12月 17日	无			
CN	103680496	A	2014年 3月 26日	无			
US	2008172349	A1	2008年 7月 17日	JP	2008204437	A	2008年 9月 4日
				US	7647284	B2	2010年 1月 12日
				JP	5102046	B2	2012年 12月 19日
CN	102983819	A	2013年 3月 20日	CN	102983819	B	2015年 2月 4日

表 PCT/ISA/210 (同族专利附件) (2009年7月)