1. 一种基于注意力机制神经网络与知识图谱的医疗文本分级方法，包括四个模块：直接关系抽取模块（1）、直接关系统一模块（2）、间接语义关系推理模块（3）、神经网络分级模块（4）；

（1）直接关系抽取模块

从知识图谱中得到实体-关系-实体之间的直接关系三元组数据，从文本数据中基于规则得到体-关系-实体之间直接关系三元组数据；

（2）直接关系统一模块：该模块将上一模块中得到的所有直接关系三元组数据基于标准化的点互信息将从知识图谱与文本数据中的实体-关系-实体直接关系三元组统一起来，得到相同的实体与关系节点。

（3）该模块主要作用是将上一模块中得到的所有的直接关系三元组通过TransE算法推导，得到两个实体之间存在间接语义关系，从而得到文本数据中的所有数据。

（4）该模块应用基于Attention机制的增强型LSTM神经网络模型得到文本分级结果。附图2为具体的神经网络模型，通过增强型LSTM模型作为Encoder模型，基于Attention机制计算得到实体-关系-实体三元组对于文本分级的重要程度，并且通过普通的LSTM作为Decoder模型通过softmax变换得到文本分级结果。

2.根据权利要求1所述的一种基于注意力机制神经网络与知识图谱的医疗文本分级方法，其特征在于步骤（2），具体如下：

基于标准化点互信息的直接关系统一算法（21）：由于文本数据中的实体与关系的描述的不准确性和不唯一性，需要将从文本数据中抽取的实体之间直接关系三元组与从知识图谱中抽取的实体之间直接关系三元组合并为统一的直接关系三元组。

具体内容是将从文本数据中抽取出来的实体i与从知识图谱中抽取出来的实体c以此计算关联度，当关联度大于阈值时，我们认为i、c这两个实体之间是等价，可以将其合并为同一个实体来表示，建立统一的知识库。其中阈值的最佳取值经过交叉验证之后为0.85.

具体公式如下：

其中P(c|i) 表示对于一个实体 i它所对应的实体为c 的概率。P(i|c) 则表示对于一个实体c它所对应的实体为i 的概率。n(c,i) 表示实体 i 和实体 c同时出现的次数。SETC为全部实体c的集合，SETI为全部实体i的集合。表示实体 i 与集合SETC中每一个实体c共现次数之和。表示实体c与集合SETI中每一个实体i共现次数之和。

计算得到两个实体之间的条件概率， 用标准化的点互信息（NPMI）来计算两个实体i、c 之间的关联度评分，其具体公式如下

其中为实体i出现次数占所有实体出现次数的概率值；为实体i与实体c共同出现的次数占所有实体对出现次数的概率值；其中PMI为衡量两个实体之间相关性的点互信息，其具体公式如下：

3.根据权利要求1所述的一种基于注意力机制神经网络与知识图谱的医疗文本分级方法，其特征在于步骤（3），具体如下：

TransE（Translation Embedding）是基于实体和关系的分布式向量表示，将三元组(head，relation，tail)看成向量h通过r翻译到t的过程，通过不断的调整向量h、r和t，找到一个使得从实体h 到实体t之间造成损失最小的关系r。通过定义势能函数，使知识库中定义的势能大于不在知识库中的三元组的势能，即最小化整体势能，其具体公式如下：

其中为平滑系数在此设为1，为在知识库中的三元组数据，为不在知识库中的三元组数据，势能函数的计算方法为计算三元组数据线性组合的第二范式。

基于TransE算法计算两个实体之间的整体势能，得到最小的势能值所对应的关系，即是两个实体间的间接语义关系。

4.根据权利要求1所述的一种基于注意力机制神经网络与知识图谱的医疗文本分级方法，其特征在于步骤（4），具体如下：

附图3为具体的增强型LSTM模型，在增强型LSTM模型中细胞向量有两个，分别作为实体细胞向量与关系细胞向量，两个细胞向量分别计算实体与关系的长短信息并分别保存，在输出门采用将两个细胞向量通过一个全连接层得到最后的输出向量。

在t时刻的增强型LSTM模型的具体公式如下：

在公式（6）中计算为实体遗忘门系数，其中、分别为实体1与实体2在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（7）中计算为关系遗忘门系数，其中为关系在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（8）中计算为实体输入门系数，其中、分别为实体1与实体2在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（9）中计算为关系输入门系数，其中为关系在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（10）中计算为实体输入门备选状态，其中、分别为实体1与实体2在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（11）中计算为关系输入门备选状态，其中为关系在该公式中的权重矩阵，为t-1时刻输出向量在该公式中的权重矩阵，为在该公式中的偏置。

在公式（12）中计算为实体细胞状态，为t-1时刻的实体细胞状态，为遗忘门系数，为输入门系数，为实体输入门备选状态。

在公式（13）中计算为关系细胞状态，为t-1时刻的关系细胞状态，为遗忘门系数，为输入门系数，为关系输入门备选状态。

在公式（14）中计算为输出门系数，分别为实体1、实体2在该公式中的权重矩阵，为关系在该公式中的权重矩阵，为t-1时刻的输出向量在该公式中的权重矩阵，为该公式中的偏置。

在公式（15）中计算为输出门结果，为输出门系数，为实体细胞状态在该公式中的权重矩阵，为实体细胞状态，为关系细胞状态在该公式中的权重矩阵，为关系细胞状态。

在增强型LSTM模型，实体与关系向量采用word2vec初始化，其向量长度为200；所有的权重矩阵与偏置采用截断的正态分布做初始化，当参数收敛时或者当达到最大迭代次数10次时训练结束；模型中的为sigmoid函数，tanh为tanh函数。

采用Attention机制得到每个实体-关系-实体三元组对于文本分级的重要程度，在t时刻得到每一个三元组的重要性程度。记decoder模型时刻 t 的target hidden为，encoder的输出隐藏向量hidden state为encoder的全部输出向量集合为HS，对于其中任意,其权重的具体公式为：

其中：

为评分函数中的重要性系数，采用截断的正态分布做初始化；为在该公式中的权重矩阵，当参数收敛时或者当达到最大迭代次数10次时训练结束。

在Decoder层采用标准LSTM模型，将基于Attention机制之后的中间隐藏向量作为Decoder层的输入，在LSTM层之后的隐藏向量通过softmax得到最后的文本分级结果。