东兴证券 DONGXING SECURITIES

让AI照进投资现实：基于深度学习的市场风险偏好指数及应用

——非银行金融行业专题报告

2018 年 10 月 26 日

看好/维持

非银行金融 专题报告

郑闵钢

010-66554031 zhengmgdxs@hotmail.com

执业证书编号: S1480510120012

研究助理：安嘉晨

010-66554014 anjc@dxzq.net.cn

投资摘要：

我们的研究方法是将基本面分析框架纳入深度学习领域。目前人工智能在互联网和制造业有着越来越广泛的应用，不管是应用领域的不断扩展还是深度学习算法的不断优化，都极大地改善了传统工作模式和思维方式。但金融领域尤其是证券研究行业目前还鲜有落地的应用案例。针对现存的种种问题，我们在此提出了一种将业界较为领先的深度学习算法应用到证券研究领域的切入方式。

- **我们将着眼点缩小至行业研究领域**，并以券商行业为代表。将行业研究的基本面分析框架纳入深度学习领域。
- **我们使用券商行业基本面的分析框架**，对收集到的新闻和文章内容，根据其所表达的全文内容信息，综合考量该文章内容对于券商行业的基本面是利好还是利空，并标记出相应五点式评分，包括非常利空、利空、中性、利好和非常利好。
- **使用网络上从各新闻源爬虫获取的文章内容，形成 17500 篇文章样本**，并对其逐一进行阅读和打分，形成深度学习模型的训练集和测试集。
- **我们采用 Hierarchical Attention Networks 深度学习模型**，并对其进行训练，使其具备依靠基本面逻辑判断文章传达信息属于利好还是利空的能力。并将其结果进行输出。给予输出结果，我们进一步了探讨其应用的方式。

我们将编制的深度学习-风险偏好指数与上证指数联合进行回测分析，目前发现该指数有两方面特点：

- **在回测的各历史阶段，都有比较好的市场背景解释能力。**证明该风险偏好指数有一定的合理性。
- **在市场上涨末期，该风险偏好指数具有较好的预警能力。**当大盘指数上涨但风险偏好指数下降时，提示后市回撤风险，效果比较明显。

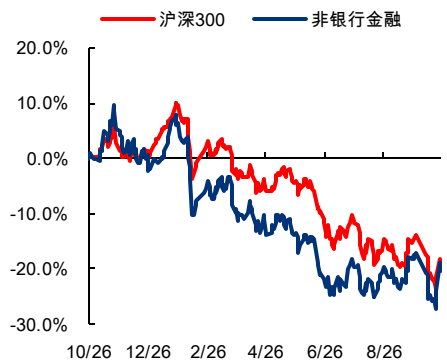
本次专题研究首次尝试将行业研究的基本面框架纳入深度学习的技术领域，未来有着更为广阔的扩展空间：

- 本专题使用券商行业作为代理行业，但若有更为垂直的新闻源，理论上可以纳入任何板块的基本面分析框架从而输出针对不同行业的基本面舆情指数。
- 多行业综合后的风险偏好指数将可能提供更为客观的市场风险偏好测量。从而为风险偏好的观测开拓一条全新的路线。
- 随着模型学习到更多的信息，该舆情指数和风险偏好测度也将越发准确。
- 未来不断扩大新闻源，可以预期当我们每天用来计算舆情指数的新闻源进一步扩充之后，包括纳入类似微信公众号等众多自媒体平台内容以及社区内容后，将具备更好的市场代表性。

行业基本资料

		占比%
股票家数	60	1.69%
重点公司家数	-	-
行业市值	44945.83 亿元	8.95%
流通市值	30031.11 亿元	8.37%
行业平均市盈率	16.06	/
市场平均市盈率	13.70	/

行业指数走势图



资料来源：东兴证券研究所

相关研究报告

1、《证券行业深度报告：券商资管业务：行业整体转型的轴心》2018-09-26

2、《证券行业 2018 年半年报综述：结构性集中！“非好即坏”并非证券行业分化本质》2018-08-31

3、《证券行业深度报告：CDR 打开券商盈利空间及新业务分析》2018-06-19

4、《非银行金融行业报告：破净也可重圆券商估值历史新低》2018-06-01

5、《证券行业 2017 年及 2018 年一季报综述：先行者底部的守望》2018-05-16

6、《非银行金融行业报告：乘着周期的翅膀证券板块贝塔值的再思考》2018-05-09

7、《非银行金融行业报告：叫停私募场外期权无碍券商业绩表现》2018-04-12

目 录

表格目录

插图目录

深度学习在证券研究领域的应用价值潜力巨大

1.1 我们的研究方法是将基本面分析框架纳入深度学习领域

目前人工智能在互联网和制造业有着越来越广泛的应用，不管是应用领域的不断扩展还是深度学习算法的不断优化，都极大地改善了传统工作模式和思维方式。但金融领域尤其是证券研究行业目前还鲜有落地的应用案例，我们认为主要由于以下几方面原因：

- 行业壁垒的存在。最新的技术成果尚未有效的被证券研究领域所吸收。
- 应用领域的限制。大多数有限的应用往往限制在了追求绝对收益的量化投资领域，并且以股票价格信息作为学习基础的模型解释性有限，风险管理难度高，容易被证伪。而以新闻资讯为学习基础的模型又涉及舆情的定义方式，关键词频率等指标又缺少立场和方向性。
- 立场的限制。在深度学习的应用里，一个很重要的问题是模型对标了什么（立场的人）。目前很多基于新闻资讯的深度学习训练集是不存在立场的。而同样的新闻，对于不同立场的受众，所判定正负面方向和程度是不一样的，这也导致了很多案例的训练集内立场复杂冲突，对受众指示价值有限。
- 影响市场变量的复杂。通常人们倾向于用各种手段描述整个市场的起伏与趋势，而影响市场的变量往往十分复杂，而传导逻辑难以用技术手段给与简单的描述和量化。因此用于预测大盘走势等的应用往往效果较差。
- 证券研究的专业性难以输出。在行业研究领域，对于基本面的研究框架具有一定的专业性，在技术领域往往不容易落地，尤其对于需要庞大训练集的深度学习领域更是形成了一种天然屏障，如何将专业的证券研究方法与技术需求相结合成为一种阻碍。

针对以上现存的种种问题，我们在此提出了一种将业界较为领先的深度学习算法应用到证券研究领域的切入方式。

- 我们将着眼点缩小至行业研究领域，并以券商行业为代表。将行业研究的基本面分析框架纳入深度学习领域。
- 我们使用券商行业基本面的分析框架，对收集到的新闻和文章内容，根据其所表达的全文内容信息，综合考量该文章内容对于券商行业的基本面是利好还是利空，并标记出相应五点式评分，包括非常利空、利空、中性、利好和非常利好。
- 使用网络上从各新闻源爬虫获取的文章内容，形成 17500 条文章样本，并对其逐一进行阅读和打分，形成深度学习模型的训练集和测试集。
- 我们采用 Hierarchical Attention Networks 深度学习模型，并对其进行训练，使其具备依靠基本面逻辑判断文章传达信息属于利好还是利空的能力。并将其结果进行输出。给予输出结果，我们进一步了探讨其应用的方式。

1.2 深度学习模型 Hierarchical Attention Networks

1.2.1 模型亮点

Hierarchical Attention Networks（HAN）是用来进行文档分类的深度学习模型，其具备两个鲜明的特点：

- 模型本身具有明显的分层结构，这种分层的结构实际上切合了通常文章所体现出的层次结构，有比较好的全文识别能力。
- 模型具备双层的注意力机制，包括词级别和句子级别。当模型试图建立整个文档的表示时可以区分对待不同的句子或者词，给予不同的权重。也就是建立了一种注意力的机制，对某些重要句子或者词可以比较好的进行捕捉。而这也在某种程度上暗合了通常我们理解文章内容的方式。
- 当融合了分层结构和注意力机制了以后，这个模型的阅读方式已经非常贴合人自然的阅读习惯了——人在阅读句子的时候是逐词阅读，但是读者不会给予每一个词等量的注意力。同样，在阅读全文的时候，也会更关注某些重要关键句。

所以我们采用 Hierarchical Attention Networks（HAN）来与行业研究基本面判断框架相结合，在技术上具有一定的前瞻性。

1.2.2 模型基本思路符合人们阅读理解习惯

文档分类是自然语言处理领域非常基础性的一类问题。其任务目标就是对给定的文档进行分类或者打标签。在目前较为前沿的领域，采用深度学习的方式已经开始占据主流，并不断取得较好的效果。

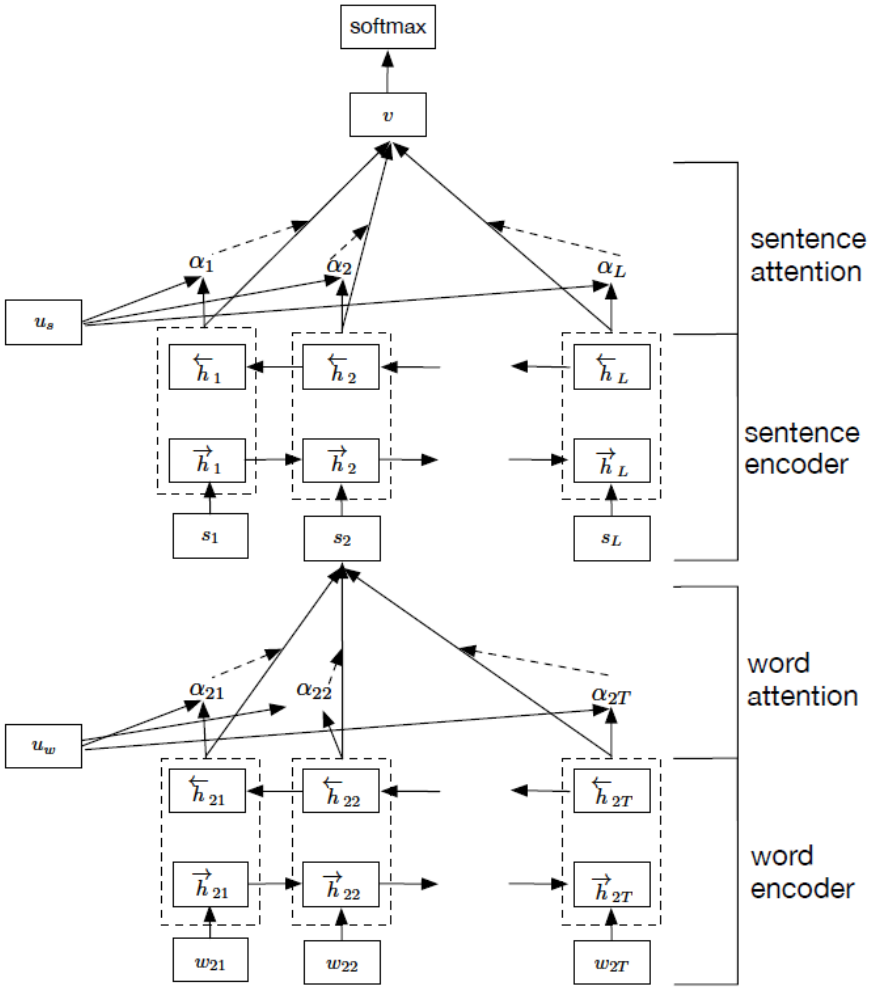
尽管传统深度学习模型在一定范围内取得了不错的效果，但是如果在模型的构建中将文章本身结构方面的知识加以考量（词与词之间的逻辑、句子与句子之间的逻辑），那么将会得到更好的结果。

而 HAN 在以上这些共识的基础上，进一步优化并加入注意力机制，即认为回答某类问题时，并不是文章所有部分的内容都是平等的。而为了确认哪些才是真正相关内容时，就需要将词语之间、句子之间的交互作用加以考虑，而不是独立的看待词语和句子——这一思想，与我们通常的阅读理解方式尤其一致。我们认为，在经济金融领域的文章和资讯，通常相对逻辑较普通文章复杂，修辞手法多样，该模型更具有适用性。

总结下来，HAN 的建模思路可以分为以下两步：

- 根据文章本身的层次结构（由此组成句子、由句子组成文章），建立一个类似的模型。首先对句子进行建模表示，随后再聚合句子形成文章的建模表示。
- 评估不同词和句子在文章中的信息重要性，使得其与上下文背景具有相关性。引入两层的注意力机制，一层针对词，另一层针对句子。从而使得模型可以识别不同词与句子对于最后分类结果的重要性。

图 1: HAN 的模型结构



资料来源：Hierarchical Attention Networks for Document Classification(Zhang et al. 2016)，东兴证券研究所

模型包括以下四个部分：

- **Word Encoder：**首先将给定的句子在分词之后，转换为词向量。并输入到一个双向 GRU 神经元中。（GRU 可以较好解决传统 RNN 递归神经网络长期记忆能力不佳的问题）其数学表达为：

$$x_{it} = W_e w_{it}, t \in [1, T],$$
$$\overrightarrow{h}_{it} = \overrightarrow{\mathrm{GRU}}(x_{it}), t \in [1, T],$$
$$\overleftarrow{h}_{it} = \overleftarrow{\mathrm{GRU}}(x_{it}), t \in [T, 1].$$

其中，ω_{it}代表第i个句子中的第t个词。通过正向顺序读取句子中的词获得隐层表示

$\overrightarrow{h}_{it}$，再通过反向顺序读取句子中的词获得隐层表示$\overleftarrow{h}_{it}$。从而$h_{it}=[\overrightarrow{h}_{it},\overleftarrow{h}_{it}]$即表示了整个句子在词$\omega_{it}$上所聚合的信息（这里[]表示单纯的向量拼接）。

- **Word Attention：**并非所有词对于句子的意思表示都是同样重要的，因此该机制用来识别具有关键意义的词汇。

$$u_{it}=\tanh(W_w h_{it}+b_w)$$
$$\alpha_{it}=\frac{\exp(u_{it}^{\top}u_w)}{\sum_t \exp(u_{it}^{\top}u_w)}$$
$$s_i=\sum_t \alpha_{it}h_{it}.$$

首先将h_{it}导入单层感知器中进行训练得到其隐层表示u_{it}，加入权重向量u_w并对其进行 softmax 标准化得到权重α_{it}。最后对h_{it}进行加权得到整个句子的表示s_i。

- **Sentence Encoder：**与处理词的逻辑相同，将s_i同样导入双向 GRU 中，并得到句子的隐层表示$h_i=[\overrightarrow{h}_i,\overleftarrow{h}_i]$。其代表了周围的句子与$s_i$的交互信息。

$$\overrightarrow{h}_i=\overrightarrow{\mathrm{GRU}}(s_i), i\in[1,L],$$
$$\overleftarrow{h}_i=\overleftarrow{\mathrm{GRU}}(s_i), t\in[L,1].$$

- **Sentence Attention：**为了能够给对文章分类提供关键信息的句子以更高的关注度，最后再次加入一个 attention 机制，重复在词层面所做的处理。加入权重向量u_s等，然后得到整篇文章的向量表示v。其涵盖了文章中所有句子的信息。

$$u_i=\tanh(W_s h_i+b_s),$$
$$\alpha_i=\frac{\exp(u_i^{\top}u_s)}{\sum_i \exp(u_i^{\top}u_s)},$$
$$v=\sum_i \alpha_i h_i,$$

最终的产出物v就是我们所关注文章的一个高度抽象的代表。而它就可以作为模型对给定文章进行打分（分类）的评价基础。同样使用 softmax 函数$p=softmax(W_c v+b_c)$，将其转化为概率形式，并对其所属分类进行判断。

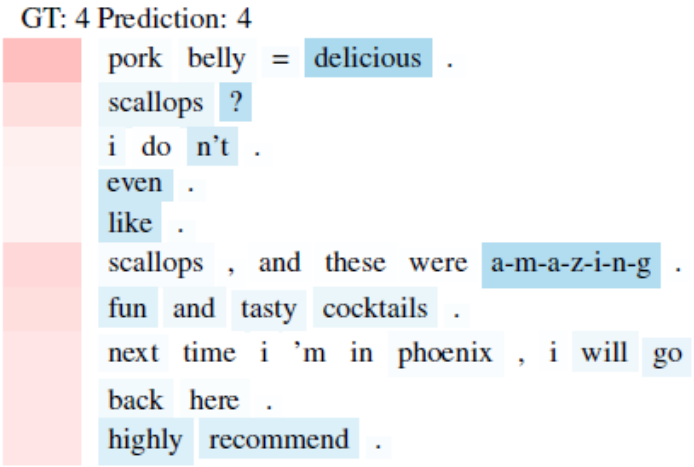
1.2.3 模型特点适用于证券研究领域

HAN 模型所具备的特点包括以下两方面：

- 首先，我们不需要定义关键词，而模型自身的 self-attention 机制会自行学习具有较高信息量的词汇。这一点对于千变万化的资本市场新闻资讯来说，非常适合。考虑到特定行业研究经常出现新生的概念与词汇，该模型将具有很好的泛化能力。
- 其次，该模型可以有效的处理跨句子的复杂逻辑，使得每个句子都将获得在文章背

景下的理解，而非单独孤立的理解。这一点可以避免文章资讯使用了修辞手法而导致分类异常。例如下图所示：

图 2: HAN 的跨句子理解能力



资料来源：Hierarchical Attention Networks for Document Classification(Zhang et al. 2016)，东兴证券研究所

上图为模型作者对点评网站 Yelp 的测试。左侧的红色深浅代表了深度学习模型对于每个句子的关注程度，而词汇的蓝色深浅则代表了对单词的关注程度。我们可以看到，如果单独理解句子“i don't even like”将偏向负面，但模型通过对全文的逻辑判断，判定该评论为正面，并选择忽视该句子在全文中的权重。

当然，模型的逻辑判断能力是完全由训练集所学习而来。良好的判断能力也取决于训练集所暗含的判断思维。这一点在点评网站上，训练集的质量似乎并未体现出突出的作用，因为训练集来源于已有的众多用户点评和评分，模型就可以学习到“大众”的点评和评分逻辑。

但是在行业研究方面，资讯本身所携带的信息与最终对于行业的正负面判断，就无法简单的获取到足够多的训练集，而非专业人士也较难做出相对准确的判断。因此在我们所关注的证券研究领域，训练集本身的作用就非常突出了。

2. 我们提出基于 HAN 的市场风险偏好间接测度方法

2.1 深度学习模型与行业研究的结合思路

深度学习技术与行业研究分为两条路径：

- 技术方面，不断涌现新的更好的深度学习模型，大家基于公开的数据集，比拼更好的准确率和效率；
- 行业研究方面，通过观察行业数据，理解商业模式，判断当前行业的景气程度和预判未来行业的演变方向。

我们认为深度学习技术与行业研究的两条路径有着一个潜在的结合点，即使用由行业分析师所提供的训练样本来大规模训练深度学习模型，从而使得模型学习行研的分析框架，进而具备对市场上每天产生的海量资讯进行专业判断的能力。

- 我们将仅针对特定行业，由深度学习模型去复制该行业的基本面分析框架，从而可以对未来任何新资讯判断其对该行业基本面是利好还是利空。我们的关注点聚焦在特定行业上，而非市场整体，因此将影响因素大大简化，并且使得行业基本面分析框架（行业知识）成为可以被学习和利用的对象。
- 理论上，由谁提供大量的训练样本，深度学习模型就会模仿谁的判断思维。那么因此，提供训练样本的行业研究员本身的水平将影响模型的优劣。而模型往往也都无法做到100%接近人的判断能力，在此基础上会进一步打折扣。如果仅仅针对单独的一篇资讯，模型的判断结果将不具备参考意义，但如果使用每天全网的资讯内容，进行大数据判断，其结果和趋势我们认为将具备应用价值。

2.2 使用券商行业作为风险偏好测度的代理行业

在本次专题研究中，我们使用深度学习模型，利用券商行业作为代理，来间接测度市场风险偏好的水平。以期能够提供一个除了风险溢价、心理学测量以外新的可参考市场指标。

- 券商行业比较适宜作为风险偏好测量的代理。我们统计了具有10年完整数据的申万二级行业的贝塔值，券商行业排名第一。说明其对于风险偏好水平十分敏感。
- 由于我们使用市场新闻资讯来判断其对于某个行业是利好还是利空信息，券商行业的各业务条线不管是轻资产还是重资产业务，都是整个资本市场的反映，因此每天所生成的具有相关性的资讯量会更大。
- 我们从券商行业研究员的视角，准备了17500个已完成打标的新闻训练样本。打标或者分类即为我们通过阅读全文后，针对其所表达信息对于券商行业的基本面是利好还是利空予以的判断。分类包括：不相关、很利空、利空、中性、利好、很利好。
- 对于训练样本的标记原则，我们秉承券商基本面的分析框架如下：

表 1：券商基本面分析框架

业务分类	收入端因素	成本端因素
轻资产业务		
经纪	佣金率、股债交易额、开户数、代销金额、交易席位租赁	代理买卖证券交易成本、投保基金上缴比例
资管	AUM（券商资管、控股基金）、管理费率、业绩提成	风险准备金、交易成本、非标违约率
投行	股债融资规模、承销保荐费率、过会率、兼并重组规模	管理费用
重资产业务		
信用	两融规模、股票质押规模、融资利率、净资本	质押回购利率、债券利率、流动性、违约坏账
自营	A股收益率、债券收益率、衍生品名义本金规模、净资本	交易成本

资料来源：东兴证券研究所

对于新闻资讯所表达信息对于券商行业基本面的影响，我们使用以上基本面的分析框架，判断其内容是利好或者利空券商行业。（包括但不限于以上基本面因素）

2.3 深度学习-市场风险偏好指数的构建方法

2.3.1 基于深度学习的券商行业舆情指数贡献了新的增量信息

当 HAN 模型使用我们提供的数据集完成训练之后（也就是学习了券商行业基本面的分析框架），即可对于每篇新输入的资讯进行针对券商行业利好利空的判断，我们采用以下方式进行数据输出：

- 完成训练的 HAN 模型使用 softmax 进行分类，也就是可以输出程序判定的某篇新闻资讯可能被划分为利好、利空和中性的三个概率。（由于目前训练样本依然尚且不足，我们在输出的时候将之前的 5 点式评分简化为 3 点式。“很利好”和“很利空”与“利好”和“利空”分别进行合并）。

- 假设某天爬取到的新闻数量为N，深度学习模型判断第i篇新闻的利好概率为p_i，利空概率为n_i，则当天全部新闻资讯利好、利空券商板块概率分别为pos, neg：

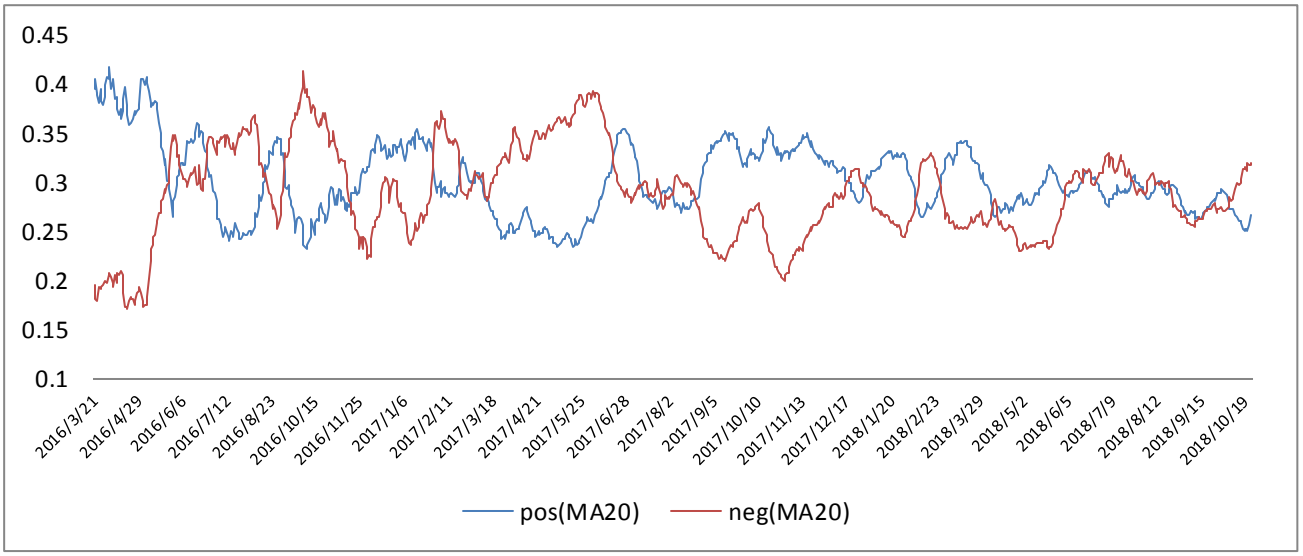
$$pos = \frac{1}{N}\sum_{i=1}^{N} p_i, neg = \frac{1}{N}\sum_{i=1}^{N} n_i$$

从而pos和neg就代表了通过深度学习模型所判断的当天全部新闻资讯整体对于券商板块基本面利好和利空的一种程度表达，也就是当日券商基本面的舆情指数。

- 注意，此处我们从网络爬取新闻资讯时，定义了第t日用于计算pos和neg的新闻为$t-1$日下午 3:00 收盘到t日下午 3:00 收盘的时间区间内所获取的网络新闻内容。（可以理解为我们认为影响当日股票价格的新闻资讯为上一日的盘后和当日的盘中所产生）

- 由于pos和neg每日的波动幅度较大，我们使用 MA20 进行输出如下：

图 3：基于深度学习模型的券商行业基本面舆情指数（2016/3/21-2018/10/24）



资料来源：深圳市有鱼智能科技，东兴证券研究所

综合起来考虑，我们计算$pos-neg$指标，其可以代表信息的**相对正面倾向**。更重要的是，我们认为深度学习的市场舆情结果是对于预测未来市场收益的一种“增量信息”而非噪音：

- 首先，我们知道券商指数的收益率有自己的自相关性质，也就是历史的收益率与未来收益率有相关性，如下图：

图 4：券商行业指数的自相关检验

天数	1	2	3	4	5	10	15	20
1	-0.9%	2.6%	4.3%	0.5%	0.4%	-1.4%	-2.6%	-1.7%
2	2.2%	5.6%	4.6%	1.5%	-1.9%	-1.6%	-5.6%	-4.3%
3	5.0%	5.0%	3.5%	-1.8%	-4.8%	-3.1%	-8.4%	-5.6%
4	2.2%	2.3%	-1.2%	-5.7%	-8.3%	-3.9%	-10.4%	-7.2%
5	1.2%	-1.8%	-4.4%	-8.5%	-10.3%	-5.5%	-12.1%	-9.2%
10	-2.9%	-4.3%	-6.9%	-7.6%	-8.4%	-11.0%	-13.8%	-13.6%
15	-6.4%	-10.3%	-12.9%	-14.8%	-15.4%	-15.6%	-18.8%	-19.0%
20	-3.6%	-7.0%	-8.9%	-10.6%	-12.5%	-15.3%	-19.6%	-21.4%

行和列意义为过去N天指数涨跌幅和未来M天券商指数涨跌幅的相关系数

资料来源：深圳市有鱼智能科技，东兴证券研究所

- 其次，我们认为$pos-neg$与未来券商指数收益率也具有相关性，并且该相关性的来源是由于提供了增量信息，而非仅仅来源于券商指数的自相关性质（也就是舆情指数并不是借由历史收益率才和未来收益建立的联系）。

图 5：舆情指数与券商指数未来收益率相关关系

天数	1	2	3	4	5	10	15	20
1	4.2%	-0.3%	-0.8%	-2.4%	-0.4%	-8.1%	-11.6%	-13.4%
2	-0.1%	-2.6%	-3.9%	-3.2%	-3.8%	-12.1%	-17.6%	-17.8%
3	-0.5%	-3.5%	-3.3%	-4.5%	-6.4%	-13.8%	-19.9%	-20.4%
4	-1.9%	-3.0%	-4.4%	-6.7%	-8.1%	-15.8%	-21.9%	-22.5%
5	-0.6%	-3.5%	-5.9%	-7.7%	-9.6%	-18.0%	-22.7%	-23.9%
10	-5.5%	-8.9%	-11.3%	-13.4%	-16.1%	-25.6%	-28.1%	-31.6%
15	-8.6%	-13.7%	-16.5%	-18.3%	-19.7%	-26.6%	-30.5%	-34.7%
20	-7.6%	-11.1%	-14.1%	-15.9%	-17.7%	-25.7%	-30.4%	-34.2%

行和列意义为过去N天pos-neg的移动平均和未来M天券商指数涨跌幅的相关系数

资料来源：深圳市有鱼智能科技，东兴证券研究所

图 6：舆情指数与券商指数过去收益率相关关系

1	2	3	4	5	10	15	20
1.5%	15.8%	18.5%	19.4%	19.0%	16.9%	12.7%	6.7%

单元格为过去N天pos-neg移动平均和过去N天券商指数涨跌幅的相关系数

资料来源：深圳市有鱼智能科技，东兴证券研究所

- 我们看到过去 20 天$pos-neg$移动平均与未来 20 天券商指数收益率相关系数为-34.2%(强度也高于券商指数的自相关)，说明$pos-neg$指标具有一定的预测能力。由于下一个交易日的新闻难免会有对于上一个交易日发生事件的总结与复盘，所以舆情指数与券商指数过去的值也存在相关性。

- 我们计算了过去 20 天$pos-neg$移动平均，并考察其与过去 20 天内不同时间窗口的券商指数收益率的相关系数，发现皆小于 20%。也就是舆情指数与券商指数历史收益率相关性较弱。

以上几点说明了该舆情指数对预测券商指数未来收益是提供了增量信息，而非依靠历史价格信息。印证了该舆情指数未来在交易策略方面的应用潜力。

2.3.2 间接法：从券商行业舆情指数到风险偏好测度

最后我们来编制一种间接衡量市场风险偏好的方式：

- 首先，我们假设券商板块指数的变动由三部分构成：基本面驱动+风险偏好驱动+其他因素驱动。

- 我们计算每日的$pos-neg$，该差值可以代表利好和利空的一种博弈，当$pos-neg$扩大时，可以认为券商行业基本面的舆情在变好。我们简单剔除券商板块指数变动中基本面舆情所贡献部分，数据归一化后将证券行业指数减去$pos-neg$基本面舆情指数，即得到市场风险偏好的一种间接测度（此处假设其他因素驱动较少，暂时忽略）。

- 券商板块指数以申万证券 II 替代，归一化使用 z-score 方法（基于我们的测试时间区间，2016/3/25-2018/10/24），最终的风险偏好测度依然使用 MA20 方式输出：

图 7：深度学习-风险偏好测度（2016/3/25-2018/10/24）

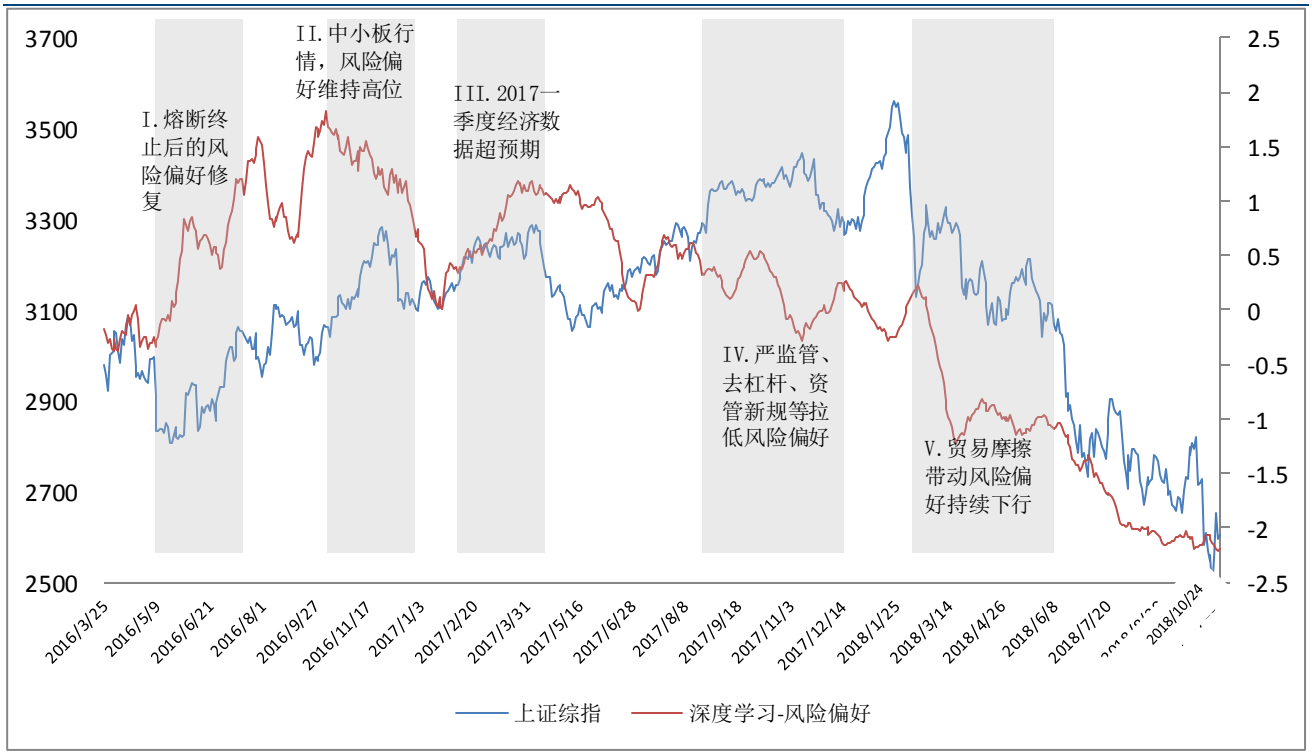
资料来源：深圳市有鱼智能科技，东兴证券研究所

3. 指数的历史回测与应用

我们将编制的深度学习-风险偏好指数与上证指数联合进行回测分析，目前发现该指数有两方面特点：

- 在回测的各历史阶段，都有比较好的市场背景解释能力，证明该风险偏好指数有一定的合理性。
- 在市场上涨末期，该风险偏好指数具有较好的预警能力。当大盘指数上涨但风险偏好指数下降时，提示后市回撤风险，效果比较明显。

图 8：深度学习-风险偏好指数的历史解释能力



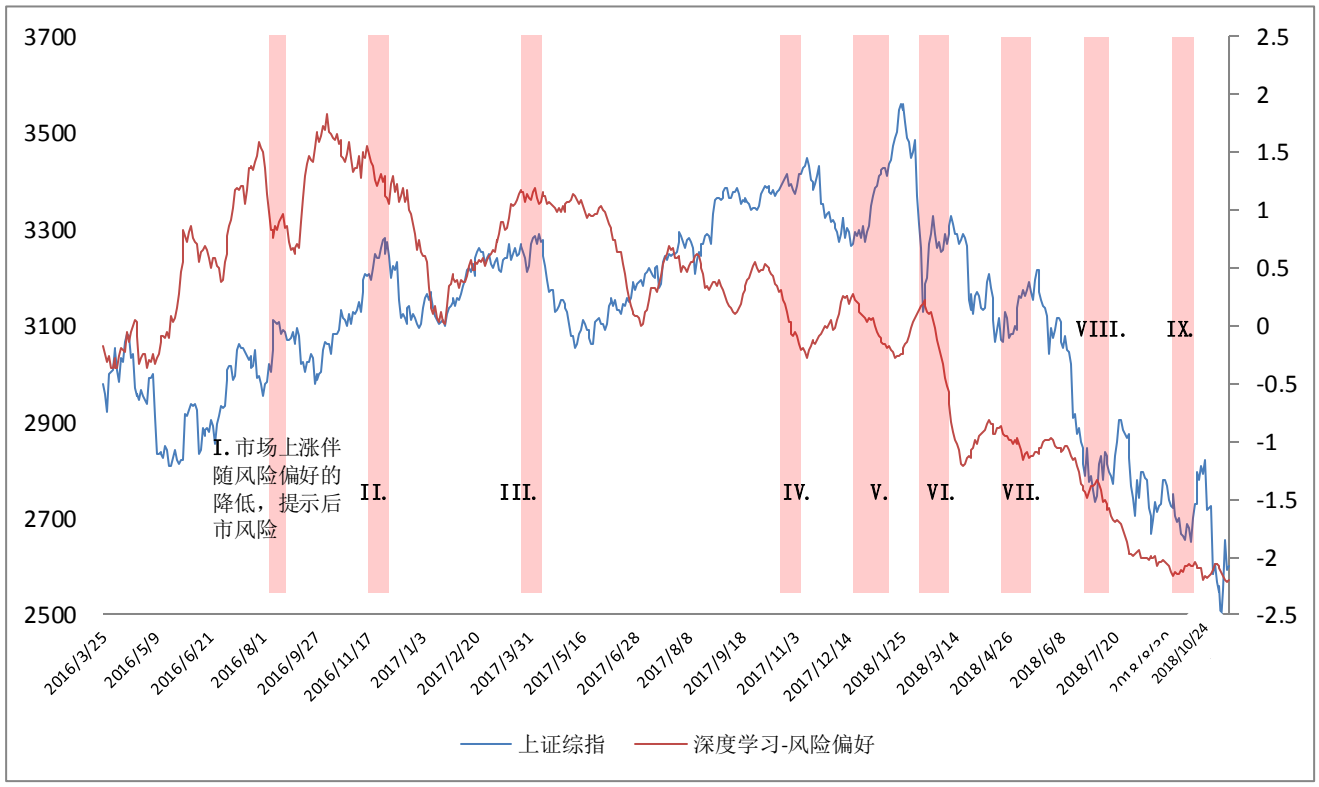
资料来源：深圳市有鱼智能科技，东兴证券研究所

从上图可见，在关键的历史时期，该风险偏好指数都有较为不错的市场解释能力：

- I.2016 年二季度：经历熔断流动性危机之后的市场风险偏好修复。
- II.2016 年四季度：中小板市场行情，中小板企业盈利增速创新高，市场风险偏好持续走高，并维持高位。
- III.2017 年一季度：经济数据超市场预期，市场风险偏好随之走高。
- IV.2017 年四季度：严监管、去杠杆、资管新规等拉低市场风险偏好水平。
- V.2018 年二季度：贸易摩擦、新兴市场回撤等外部因素带动风险偏好持续下行。

除了市场解释能力以外，我们通过对该指数的初步研究，发现其在市场上涨的尾段有较为理想的预警功能，比单独使用券商指数更为有效，能更多的捕捉到预警信号的出现。

图 9：深度学习-风险偏好指数的市场预警功能



资料来源：深圳市有鱼智能科技，东兴证券研究所

当大盘指数上涨时，风险偏好指数若未见明显上行，则提示后市调整风险较大。**我们可以看到 2018 年 10 月 22 日以来的市场反弹，依然尚未见到明显的市场风险偏好抬升迹象，提示目前市场对于反弹的谨慎态度。**

4. 未来进一步待讨论内容

4.1 本次专题研究仅是人工智能与证券研究相结合的起点

本次专题研究首次尝试将行业研究的基本面框架纳入深度学习的技术领域，未来有着更为广阔的扩展空间：

- 本次专题使用券商行业作为代理行业，但若有更为垂直的新闻源，理论上可以纳入任何板块的基本面分析框架从而输出针对不同行业的基本面舆情指数。只需要由特定行业研究员准备足够的训练样本和分析框架。未来该结合方式可以面向全板块的证券研究人员。
- 当有足够多的板块舆情数据时，也可以使用同样的风险偏好间接测度方式，输出不同代理行业所计算得出的风险偏好指数，多行业综合后的风险偏好指数将可能提供更为客观的市场风险偏好测量。从而为风险偏好的观测开拓一条全新的路线。
- 由于我们使用的训练集也是在一定历史时期内，基于该训练集的深度学习模型在时

间上有效性还有待进一步验证。我们认为应该至少应该在每年提供更多的训练样本升级深度学习模型。否则有可能会影响未来输出指数的泛化能力。但我们相信随着模型学习到更多的信息，该舆情指数和风险测度也将越发准确。

- 未来不断扩大新闻源，可以预期当我们每天用来计算舆情指数的新闻源进一步扩充之后，包括纳入类似微信公众号等众多自媒体平台内容以及社区内容后，将具备更好的市场代表性。
- 基于行业基本面的舆情指数 pos 和 neg 未来可以进一步开发量化策略或者因子，开拓交易算法方面的研究。

4.2 合作机构鸣谢

本次专题研究内容由东兴证券非银行金融研究组与深圳有鱼智能科技有限公司合作开发。深圳市有鱼智能科技有限公司为本次研究所用深度学习算法的模型与技术后台开发商。深圳市有鱼智能科技有限公司是香港上市公司云锋金融集团旗下的子公司。

5. 风险提示

深度学习算法的黑盒问题导致难以解释结果输出的逻辑、新闻源的有限信息不能代表市场资讯、假设不合理导致指数失真

分析师简介

郑闵钢

房地产行业首席研究员，房地产、传媒、计算机、家电、农业、非银金融、钢铁、煤炭等小组组长。央视财金嘉宾。2007 年加盟东兴证券研究所从事房地产行业研究工作至今。获得“证券通-中国金牌分析师排行榜”2011 年最强十大金牌分析师（第六名）。“证券通-中国金牌分析师排行榜”2011 年度分析师综合实力榜-房地产行业第四名。朝阳永续 2012 年度“中国证券行业伯乐奖”优秀组合奖十强（第七名）。朝阳永续 2012 年度“中国证券行业伯乐奖”行业研究领先奖十强（第八名）。2013 年度房地产行业研究“金牛奖”最佳分析师第五名。2014 万得资讯年度“卖方机构盈利预测准确度房地产行业第三名”。2016 年度今日投资天眼房地产行业最佳选股分析师第三名。

研究助理简介

安嘉晨

本科毕业于北京航空航天大学，硕士毕业于北京大学。曾经在腾讯、香港云锋金融工作多年。在互联网金融与金融科技等领域有丰富工作经验和深入研究。2018 年加入东兴证券，从事非银金融行业研究。

分析师承诺

负责本研究报告全部或部分内容的每一位证券分析师，在此申明，本报告的观点、逻辑和论据均为分析师本人研究成果，引用的相关信息和文字均已注明出处。本报告依据公开的信息来源，力求清晰、准确地反映分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

风险提示

本证券研究报告所载的信息、观点、结论等内容仅供投资者决策参考。在任何情况下，本公司证券研究报告均不构成对任何机构和个人的投资建议，市场有风险，投资者在决定投资前，务必要审慎。投资者应自主作出投资决策，自行承担投资风险。

免责声明

本研究报告由东兴证券股份有限公司研究所撰写，东兴证券股份有限公司是具有合法证券投资咨询业务资格的机构。本研究报告中所引用信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

我公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。本报告版权仅为我公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为东兴证券研究所，且不得对本报告进行有悖原意的引用、删节和修改。

本研究报告仅供东兴证券股份有限公司客户和经本公司授权刊载机构的客户使用，未经授权私自刊载研究报告的机构以及其阅读和使用者应慎重使用报告、防止被误导，本公司不承担由于非授权机构私自刊发和非授权客户使用该报告所产生的相关风险和责任。

行业评级体系

公司投资评级（以沪深 300 指数为基准指数）：

以报告日后的 6 个月内，公司股价相对于同期市场基准指数的表现为标准定义：

强烈推荐：相对强于市场基准指数收益率 15%以上；

推荐：相对强于市场基准指数收益率 5%～15%之间；

中性：相对于市场基准指数收益率介于-5%～+5%之间；

回避：相对弱于市场基准指数收益率 5%以上。

行业投资评级（以沪深 300 指数为基准指数）：

以报告日后的 6 个月内，行业指数相对于同期市场基准指数的表现为标准定义：

看好：相对强于市场基准指数收益率 5%以上；

中性：相对于市场基准指数收益率介于-5%～+5%之间；

看淡：相对弱于市场基准指数收益率 5%以上。