



QUBE

QUantitative Bitcoin Exchange

数字货币投资智能量化分析引擎

白皮书（中文版）

目录:

1. 项目背景及定位	3
2. 解决方案及模型	4
2.1 解决方案概述	4
2.2 引擎核心模型	5
2.3 引擎核心算法	7
3. 产品结构及计划	8
3.1 产品结构介绍	8
3.2 产品功能举例	9
3.3 产品研发计划	11
4. 早期投资及团队	12
4.1 早期投资机构	12
4.2 顾问团队成员	12
4.3 创始团队成员	12
4.4 技术运营顾问	14
5. 资产及分发计划	14
5.1 分发计划介绍	15
5.2 回购机制介绍	15
5.3 上线交易规则	16
6. 附录：可公开的基础算法	16
6.1 舆情量化模型	16
6.2 特征提取算法	17
6.3 时序预测算法	17
6.4 机器学习算法	19

传统金融行业过去 20 年在 IT 基础设置、中心化数据存储及互联网生态推动下，脱离初期人工记账和电话交易，成为影响全球经济发展的最重要行业之一，催生 Bloomberg 等基础数据服务提供商；数字货币金融时代，由区块链、大数据和人工智能相关技术共同推动：如何从大数据中实时发现投资机会点、通过人工智能量化分析决策作为商业决策辅助、利用神经网络和深度学习进行市场预测及风险预警，将会成为未来行业趋势。

1. 项目背景及定位

目前全球数字货币超 3781 种，每天都形成新的区块链生态组成部分。数字货币的区块链本身包含 17 种类型数据，如：发行总量、挖矿难度、损耗量、流通锁定量、区块交易数据、矿工费等。

目前全球数字交易平台有超 467 个，数字货币超 3781 种，其中超 1765 种可在交易平台中交易，形成 3183 交易对，数字仍不断快速增加。全球交易平台产生实时交易价格，实时挂单与交易量等，每秒数据产生大量变化。不同交易所间交易价格平均最大差值 9.73%，二八法则长尾交易量平均分布平台 382 个，但仍约有 63.9% 的交易在 OTC 或其他方式进行。除此之外，还有 OTC 数据、非交易数据、期货衍生品数据等，相关标的信息 69 类。

在区块链完全去中心化和匿名技术模型下，数字货币在众多独立分布式交易中心交易，与传统金融行业相比，98.9% 的常规分析数据无法获取。最关键的价和量数据，仅有 12.9% 能实时覆盖，且分布在 73 个数据源中。并且，目前 98% 的交易所只提供交易实时数据（76.9% 的数据仅为当前时间戳数据），完全没有历史数据作为投资决策参考。

过去一年统计，区块链和数字货币日均产生 17875 条新闻资讯、537819 条 UGC 信息，日均 555694 条舆情信息中，包括市场的利空利好以及投资者的看多看空，这些不确定性信息一定程度致使数字货币价格平均每小时波动 5.5%，且 24 小时不停盘。

如何从众多区块、独立分布式交易所、繁杂舆情信息中抓取海量数据，并对日均 5.6017T（目前）结构化化处理、积累，形成目前全球唯一区块链市场结构式数据中心；在此基

础上分析规律、洞察本质、捕捉机会点、智能化进行市场预测和风险预警、给商业决策带来价值，最终形成行业基础和行业标准，全面公开共享，并提供智能量化分析引擎，是 QUBE 的定位。

据 2018 年 1 月 1 日数据测算，全球证券交易市场市值 100 万亿美元，区块链数字货币总市值 7552.7 亿美元，仅占传统证券市场市值的 0.76%。传统证券交易二级市场中量化辅助交易占比已达 19.6%，QUBE 率先定位细分数字货币市场，未来有极强的发展空间和预期。

QUBE 服务于：

- 3000 万数字货币投资者 (10%为收费 VIP 用户)
 - 计划为个人投资者推出以币种为单位的量化策略机器人
- 10000 家专业金融机构 (全部为收费用户)
- 500 家专业数字资产基金 (全部为收费用户)

2. 解决方案及模型

QUBE 智能量化分析引擎，解决方案基于数据大规模处理和深度学习算法的数据降维、噪点去除、特征提取、模型训练等过程，在神经网络量化因子优化归一后，形成最终结果输出。其核心是对日均 5.6017T (目前) 海量数据进行处理，包括区块链数据、各类交易平台数据，网络信息舆情数据等，通过人工智能技术体系下的一系列算法模型训练，最终对数字货币市场各类博弈行为进行模型预测和数据量化。

2.1 解决方案概述

1) 数据获取

我们通过全球 276 个节点来获取数据，通过 WebSocket、API 接口，配合爬虫实时抓取 467 个交易平台数据中 1765 个数字货币的 3183 交易对价格、实时挂单与交易量等

信息，覆盖 95%以上币种，90%以上交易平台，平均 3.9 秒钟刷新更新数据。

通过爬虫技术抓取影响区块链市场投资的文字信息，分类为资讯类信息和 UGC 信息，共 6786 个信息源，日均 876 万未结构化数据；通过对区块链节点数据的监听挖掘，获取区块链隐藏信息，并结构化存储 17 类数据。QUBE 引擎目前每分钟平均获取原始数据 3.91GB，处理后存储数据仓库。

2) 数据量化

全球 26 个数据中心进行数据处理，去除噪点，输入模型量化，形成交易数据的中心结构化数据库；通过语义分析对舆情原始数据进行处理，切词标签化，深度学习量化成市场多空影响因子与情绪因子一系列指标；对区块链数据进行技术处理后，形成实时数据源。

通过 AI 人工智能对数据进行量化处理，在 983 个维度上进行深度建模分析，形成 28919 中间因子机器学习，最终形成 397 类直接量化数据的价值输出结果。众所周知，数据建模处理和人工智能训练需要大规模计算量，我们利用区块链技术招募合作伙伴，进行分布式计算来解决算力问题，推动相关技术落地时间比传统模式 ROI 提高 87%。

3) 智能策略

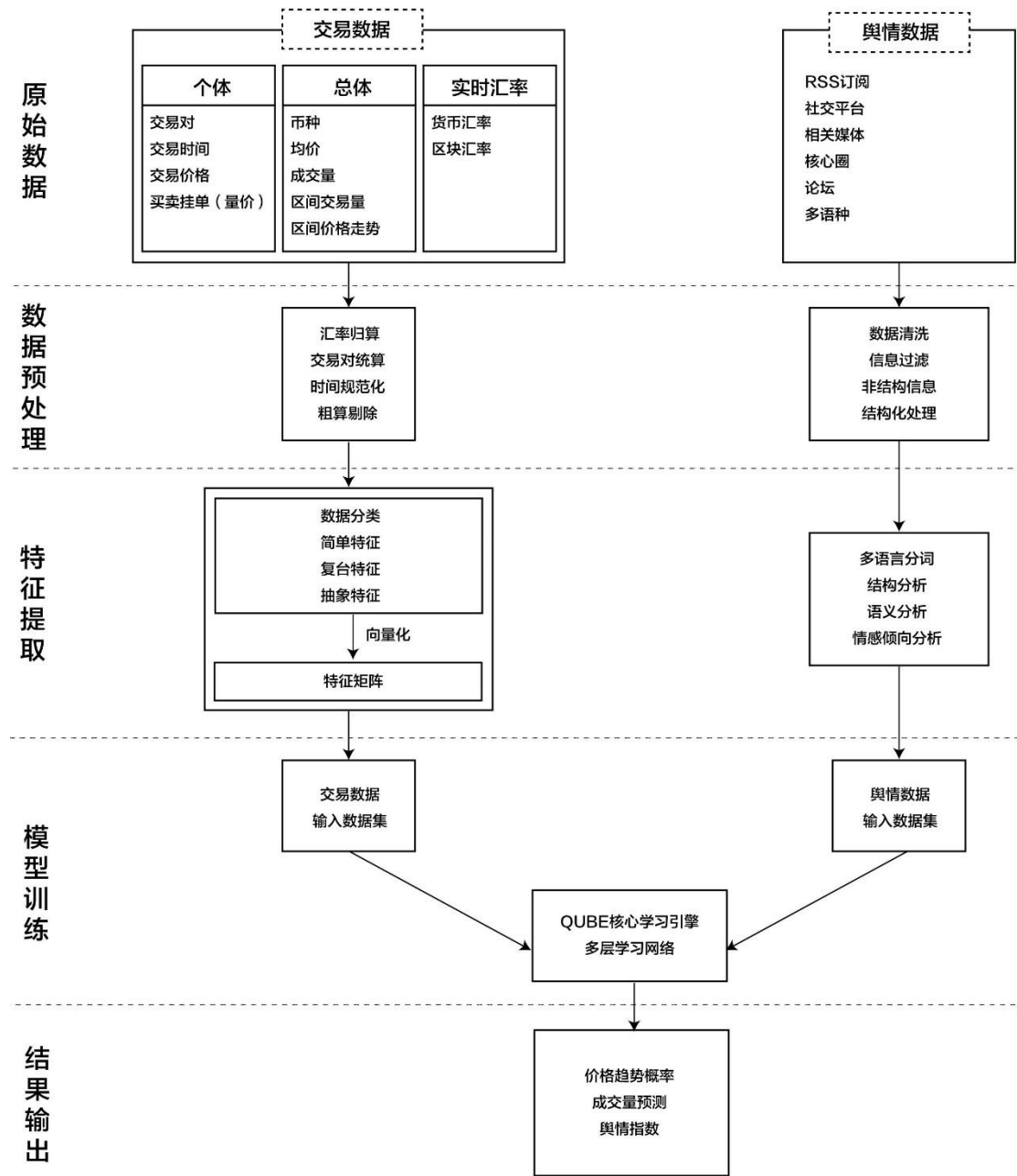
日均产生的 5.6017T 数据中，包括大量有价值信息，通过机器学习和深度学习算法，在最小时间戳为 1 分钟维度上进行迭代结果导入训练，优选最佳的数字货币投资策略。

(相关基础算法见「附录：引擎可公开核心算法文档」)

QUBE 引擎从 2017 年 7 月份开始启动运行，到目前共 5 个月算法和模型训练时间，AI 人工智能投资策略启动后，与比特币单一币种的同期收益相比提高 27.9%。且月环比引擎本身提高 11.6%，呈现环比率增加。QUBE 引擎的目标是在 2018 年推出 60 类智能量化策略并产品化。

2.2引擎核心模型

QUBE 引擎的核心为以深度学习算法为基础模型，旨在利用各类区块链货币的海量历史数据，结合自动获取的每日交易数据和区块链生态相关数据（如交易各类信息量化、舆情数据、区块数据、货币汇率等），作为模型输入样本数据，通过大量运算训练并迭代优化 AI 模型，最终实现能够发现一定时间范围内或特定场景下区块价格的动态变化风险规律。



模型的原始数据主要分交易数据和舆情数据等。其中，交易数据为定量数据，包含各大交易平台的交易数据、行情数据、OTC 场外交易数据记录、以及各类货币的实时交易汇率。舆情数据为定性数据，包括与区块链相关的各类新闻、资讯、公告、用户 UGC 信

息等等，舆情数据最终被量化为对某个时间段的影响因子，与交易数据一并作为 AI 模型的数据源。

通过定义市场趋势的指标，将市场背景按照多、空、震荡等 27 个因子划分成若干情形，按币种发行时间，将币种划分为新币、短期、长期等 36 个类型；按币种价格，将币种划分为微额、小额、中额、大额等 45 种类型；按照舆情指数，将市场情绪划分为买入、卖出、观望等 13 个类型。按上述方法，将整个区块链市场的 N 个特征进行提取和分类，即 N 维市场，若每个维度有 M 个分类，则市场被量化成 $N \times M$ 的矩阵，将 $N \times M$ 个组合情形下的价格、短期走势、中长期走势进行抽象，即特征向量化。

按照上述分类，将任意时间点历史数据以及每日更新的数据快速分解为各类特征的描述，可快速作为输入数据纳入最终模型中，并利用机器学习算法训练出能够快速对市场进行判读和走势预测的 AI 模型。这是引擎的核心模型。

模型为监督学习和无监督学习的结合，初始特征的分类和输入，帮助模型快速解读市场，随着数据量的增多和特征的明显，模型可自动发现市场中存在的隐含特征，同时也可将持续超出临界置信的特征自动剔除。AI 模型的最终输出结果为对当前时段价格、成交量、市场情绪的短期趋势和长期趋势的概率分布，最终形成 397 类直接量化数据的价值输出结果。

2.3 引擎核心算法

QUBE 引擎的核心算法包括舆情分析、时序预测、回归分析和机器学习模型等，我们团队成员不仅在人工智能方向有充分的技术储备，同时具备丰富的量化金融建模经验。人工智能的关键是算法和对应的巨型数据量训练，我们公开部分引擎基础算法，供行业监督反馈和验证。（详见白皮书第六部分附录）

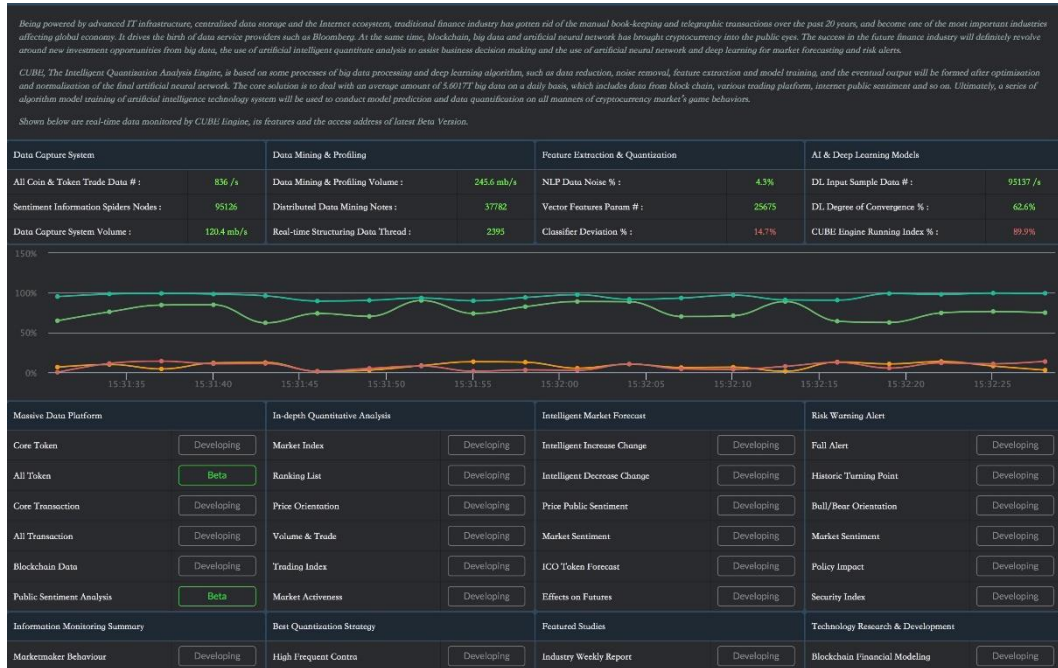
3. 产品结构及计划

3.1 产品结构介绍



3.2 产品功能举例

1) 主页全部功能模块:



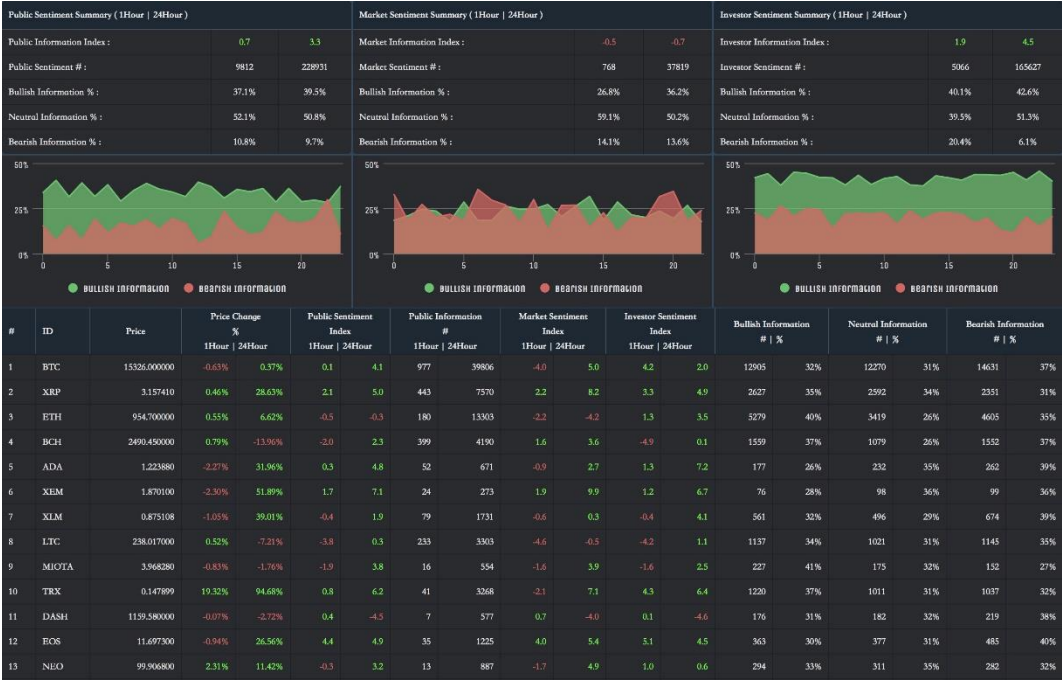
QUBE 引擎实时更新四大关键模块（原始数据获取系统、数据处理结构化系统、特征提取与量化系统、AI 深度学习模型系统）的运营状态指标。QUBE 引擎的 8 大类别、48 个功能模块，将按计划逐步上线。

2) 功能模块示例：(海量数据平台/全部币种)

[illegible]

目前 QUBE Beta V0.1 版已开通 “All Token” 功能模块，支持 1385 个币种的 376 个交易平台价格、实时挂单、交易量等交易数据，每 3.6 分钟之内轮回更新周期，其中 176 个币种数据在 3.9 秒内更新。

3) 功能模块示例：(舆情信息量化指标：[-10，+10])



QUBE Beta V0.1 版已开通舆情量化模块的舆情分析功能，QUBE 抓取全球共 6786 个信息源，日均 876 万未结构化数据。通过语义分析智能识别量化，最终产生 3 大量化指数 (网络舆情指数、市场多空指数和投资情绪指数)，从-10 到+10，已支持 1 小时 /24 小时维度，数据 5 分钟内更新。QUBE Beta V0.1 版已支持中、英文信息量化指数，以 5 个月为周期上线新语言的信息量化指数。

3.3 产品研发计划



全球拓展计划：

- 2018 年 1 月 30 日， 上线中、英语言版本，包含中、英语言脚本舆情量化分析。
- 2018 年 2 月 30 日， 上线日、韩语言版本，5 个月后支持日韩语言舆情量化分析。
- 2018 年 5 月 30 日前， 支持全球 30 个以上语言版本。
- 2018 年 1 月 30 日以后， 每 5 个月支持一种语言脚本的舆情量化分析。

QUBE 官方唯一信息发布渠道：<http://www.qube.vip>， 欢迎体验测试 QUBE 引擎 Beta 版本。

4. 早期投资及团队

4.1 早期投资机构

INBlockchain Inc. (Li Xiao Lai)

Alpha Key Capital Inc. (Travis Chaw)

4.2 顾问团队成员

李笑来

EOS, Sia, ZCash 和 yunbi.com 早期投资者。

赵东

墨迹天气联合创始人, 知名区块链、数字资产交易所投资人, 互联网创业天使投资人

帅初

量子链创始人, 中国科学院博士辍学, 曾就职于阿里巴巴集团, 具有丰富的区块链技术开发和管理经验

黄敏强

公信宝创始人兼 CEO, 香港财经学院 MBA, 在数据交换、互联网金融、区块链领域工作和研究十余年

David Vorick

SIA 数据存储平台联合创始人, Bitcoin Core 贡献者

4.3 创始团队成员

Michael Chen 创始人兼 CEO

精于金融数据和量化分析, 北美 10 年数据挖掘分析经验, 2013 年进入区块链行业, 深度研究数字资产的发展趋势。善于技术成果产品化, 多款大型互联网项目孵化运营经验

周诸明 工程项目 CTO

曾任微软总部数据产品顾问，数据产品运营总监，15 年以上数据挖掘分析经验，曾主导开发多款商业数据产品研发

Ethan Loh 首席量化分析科学家

美国 MIT 金融工程学硕士，金融量化分析专家，曾就职于华尔街渣打银行和马来西亚 N2N 公司，参与 N2N Connect 项目研发

徐骞 量化算法科学家

新加坡国立大学博士，金融工程专业，绿色能源以及经济学数据分析科学家，在瑞士瑞信银行参与多个金融数据项目研发与运营

曹彬 产品技术负责人

同济大学博士，大数据挖掘分析方向，曾任霍尼韦尔数据架构师，约 18 年的数据量化经验，曾负责多个大型数据项目工程研发

梁朋 舆情分析负责人

清华大学本硕毕业，光电子专业，曾任谷歌语义分析团队核心技术人，舆情分析技术专家，负责研发多个舆情分析量化项目

王伟 人工智能技术负责人

清华大学本硕毕业，计算机专业，曾任完美世界游戏 AI 核心工程师，在硅谷多家 AI 团队担任技术顾问，开源社区 AI 项目资深贡献者

Shawn Xu 量化投资负责人

斯坦福大学硕士毕业，计算机专业，量化投资分析，区块链技术钻研者，毕业后在硅谷多家区块链公司负责相关数据技术工作

刘嘉宁 区块链负责人

南京大学毕业，国内区块链技术资深专家，曾在多家网络安全公司担当负责人，精于分布式系统，加密数字货币，加密算法

Vicky Tian 项目运营负责人

新加坡管理大学/英国曼彻斯特商学院，金融信息管理、商业创新学、区块链研究、股票研究与投资组合，曾任多家跨国金融机构商业化产品运营

4.4 技术运营顾问

Azmi Suhaimi 数据分析顾问

美国密歇根大学人工智能方向博士，数学模型分析，Bloomberg 数据研究和深度学习

Johnny Goh Chia Min 风险控制顾问

新加坡管理大学毕业，商业管理学专业，曾在瑞士瑞信银行亚太地区负责人，深度研究金融流程以及，亚洲交易委员会的场外交易和服务结算

David Chau 市场营销顾问（国际市场）

在东南亚及欧洲市场的金融行业及产品的市场营销领域有深厚造诣，有多年丰富经验

Gary Lim 法律顾问

新加坡 Lee & Lee 法律顾问团队核心成员，主要负责金融业的公司法律风险咨询服务

5. 资产及分发计划

QUBE 是基于以太坊 Ethereum 发行的去中心化区块链数字资产，基于 ERC 20 标准。QUBE 专用于购买 QUBE 数字货币投资智能量化分析引擎相关服务，具有全球流通价值。

QUBE 基金会计划建立第三方数字资产私募基金信息服务平台，为全球数字货币投资者基金方提供定制量化投资分析系统服务。QUBE 基金将向第三方收取一定比例的平台专项服务费，其收益将发放给 QUBE 的持有者。

5.1 分发计划介绍

QUBE 的总量为 10 亿个，且永不增发。共分为定向份额 (30%)、QUBE 基金会份额 (30%)，市场拓展份额 (20%)、团队持有 (15%) 以及社区奖励 (5%)。

	第一轮	第二轮	第三轮
	定向基石投资人	定向机构投资人	定向投资人
份额	30%		
描述	锁定 6 个月	锁定 3 个月，此后每月解锁 1/3	无锁定 150ETH 起
价值	1 ETH = 10000 QUBE		
开始/结束	开始：2018 年 1 月 9 日；结束：2018 年 1 月 16 日		
合计	30000 ETH		

- 定向：300,000,000 个 QUBE，占 30%。未完成定向的 QUBE 将归于基金会。
- QUBE 基金会：持有 300,000,000 个 QUBE，占 30%，锁定期半年，之后逐年释放，QUBE 基金会账号对全社会公开，并在其官网公布资金使用计划、用途等，接受社会监督。
- 市场拓展与合作：200,000,000 个 QUBE，占 20%，用于 QUBE 市场拓展及合作。
- 团队持有：150,000,000 个 QUBE，占 15%，前半年解锁 5%，半年后解锁 10%。
- 社区奖励：50,000,000 个 QUBE，占 5%。QUBE 将分发不超过 5%总量的 QUBE 给社区粉丝，可用于购买 QUBE 的高级收费服务。

5.2 回购机制介绍

QUBE 将用平台数据销售的部分盈利回购二级市场上流通的 QUBE。回购的 QUBE 将被转移至销毁账户进行销毁，直至发行总量的 20% 为止。我们将确保整个过程的公开透明，届时用户也可通过 QUBE 区块链浏览器查询。

5.3 上线交易规则

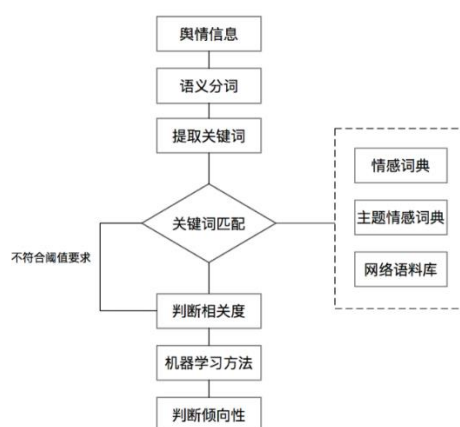
QUBE 将在 2018 年 1 月 30 日前上线交易所。

6. 附录：可公开的基础算法

QUBE 引擎的核心算法包括舆情分析、时序预测、回归分析和机器学习模型等，我们团队成员不仅在人工智能方向有充分的技术储备，同时具备丰富的量化金融建模经验。人工智能的关键是算法和对应的巨型数据量训练，我们公开少量不牵扯核心的基础算法。

6.1 舆情量化模型

通过对舆情核心内容（标题、发布时间、作者、详细信息、评论跟帖）进行特征提取，对其内容进行语义分词后。将待分类舆情内容中出现的所有特征词（词组）组成特征集。



$$T = (\text{关键词 } 1, \text{关键词 } 2, \text{关键词 } n \dots, t_n)$$

对于舆情内容，归化形成特征向量，从而计算舆情信息间的相似度。计算函数：相似函数，采用向量间的夹角大小表示相似度，通常采用夹角余弦表示：

$$\cos(A, B) = \frac{\sum A \times B}{\sqrt{\sum A^2} \sqrt{\sum B^2}}$$

距离函数，采用向量间的距离大小表示相似度，数值小则相似度高：

$$\text{dist}(A, B) = \sqrt{\sum |A - B|^2}$$

A, B 为两篇舆情内容的特征向量。最终根据特征分布对舆情信息分类形成舆情因子。

6.2 特征提取算法

运用支持向量机算法，泛化误差界：

$$R(w) \leq R_{emp}(w) + \Phi(n/h)$$

其中， $R(w)$ 代表真实风险， $R_{emp}(w)$ 代表经验风险， $\Phi(n/h)$ 代表置信风险。

通过 Logistic 算法回归寻找因子最优集合，映射后的值被认为是属于 $y=1$ 的概率：

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}} \quad , \quad g(z) = \frac{1}{1+e^{-z}} \begin{cases} P(y=1 | x; \theta) = h_{\theta}(x) \\ P(y=0 | x; \theta) = 1 - h_{\theta}(x) \end{cases}$$

最终目标为使模型训练数据中的特征 $\theta^T x \gg 0$ 。超平面公式 $f(x) = w^T x + b$ 表示

$$\begin{cases} f(x) > 0 & y = 1 \\ f(x) < 0 & y = -1 \end{cases}$$

对于数据点 x ，将其带入分类函数判断其类别，分类问题转化为寻找最佳分割直线。

给定一个训练样本 $(x^{(i)}, y^{(i)})$ ， x 是特征， y 是结果标签。 i 表示第 i 个样本：

$$\hat{y}^{(i)} = y^{(i)}(w^T x^{(i)} + b) \quad \hat{y} = \min_{i=1, \dots, m} \hat{y}^{(i)}$$

对 w 和 b 增加约束条件，确定唯一的 w 和 b ，几何间隔描述归一化距离。

$$x = x^{(i)} - \gamma^{(i)} \frac{w}{\|w\|} \quad w^T x + b = 0 \text{ 且 } w^T w = \|w\|^2$$

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{\|w\|} = \left(\frac{w}{\|w\|} \right)^T x^{(i)} + \frac{b}{\|w\|}$$

$$\gamma = \frac{w^T x + b}{\|w\|} = \frac{f(x)}{\|w\|}$$

$$y(w^T x + b) = 1,$$

$$y(w^T x + b) > 1.$$

6.3 时序预测算法

自回归移动平均模型 (ARMA)：以自回归模型 (AR) 和移动平均模型 (MA) 为基础，使用统计学手段对时间序列上的价格进行预测。

$$x_t = \phi x_{t-1} + \epsilon_t$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t.$$

$$E_t(x_{t+1}) = E_t(\phi x_t + \epsilon_{t+1}) = \phi x_t$$

$$E_t(x_{t+2}) = E_t(\phi^2 x_t + \phi \epsilon_{t+1} + \epsilon_{t+2}) = \phi^2 x_t$$

$$\dots = \dots$$

$$E_t(x_{t+k}) = E_t(\phi^k x_t + \phi^{k-1} \epsilon_{t+1} + \dots + \epsilon_{t+k}) = \phi^k x_t$$

$$\text{Var}_t(x_{t+1}) = \text{Var}_t(\phi x_t + \epsilon_{t+1}) = \sigma_\epsilon^2$$

$$\text{Var}_t(x_{t+2}) = \text{Var}_t(\phi^2 x_t + \phi \epsilon_{t+1} + \epsilon_{t+2}) = (1 + \phi^2) \sigma_\epsilon^2$$

$$\dots = \dots$$

$$\text{Var}_t(x_{t+k}) = \text{Var}_t(\phi^k x_t + \phi^{k-1} \epsilon_{t+1} + \dots + \epsilon_{t+k}) = \sum_{j=0}^{k-1} \phi^{2j} \sigma_\epsilon^2$$

$$E_t(x_{t+k}) \rightarrow 0$$

$$\text{Var}_t(x_{t+k}) \rightarrow \sigma_\epsilon^2 / (1 - \phi^2)$$

$$x_t = \epsilon_t + \theta \epsilon_{t-1}$$

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}.$$

$$E_t(x_{t+1}) = E_t(\epsilon_{t+1} + \theta \epsilon_t) = \theta \epsilon_t$$

$$E_t(x_{t+2}) = E_t(\epsilon_{t+2} + \theta \epsilon_{t+1}) = 0$$

$$\dots = \dots$$

$$E_t(x_{t+k}) = E_t(\epsilon_{t+k} + \theta \epsilon_{t+k-1}) = 0$$

$$\text{Var}_t(x_{t+1}) = \text{Var}_t(\epsilon_{t+1} + \theta \epsilon_t) = \sigma_\epsilon^2$$

$$\text{Var}_t(x_{t+2}) = \text{Var}_t(\epsilon_{t+2} + \theta \epsilon_{t+1}) = (1 + \theta^2) \sigma_\epsilon^2$$

$$\dots = \dots$$

$$\text{Var}_t(x_{t+k}) = \text{Var}_t(\epsilon_{t+k} + \theta \epsilon_{t+k-1}) = (1 + \theta^2) \sigma_\epsilon^2$$

$$E_t(x_{t+1}) = E_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+1-j}\right) = \sum_{j=1}^q \theta_j \epsilon_{t+1-j}$$

$$E_t(x_{t+2}) = E_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+2-j}\right) = \sum_{j=2}^q \theta_j \epsilon_{t+2-j}$$

$$\dots = \dots$$

$$E_t(x_{t+k}) = E_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+k-j}\right) = \sum_{j=k}^q \theta_j \epsilon_{t+k-j} \quad \text{for } k \leq q$$

$$E_t(x_{t+k}) = E_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+k-j}\right) = 0 \quad \text{for } k > q$$

$$\begin{aligned}
\text{Var}_t(x_{t+1}) &= \text{Var}_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+1-j}\right) = \sigma_\epsilon^2 \\
\text{Var}_t(x_{t+2}) &= \text{Var}_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+2-j}\right) = 1 + \theta_1^2 \sigma_\epsilon^2 \\
&\dots = \dots \\
\text{Var}_t(x_{t+k}) &= \text{Var}_t\left(\sum_{j=0}^q \theta_j \epsilon_{t+k-j}\right) = \sum_{j=0}^k \theta_j^2 \sigma_\epsilon^2 \quad \forall \quad k > 0
\end{aligned}$$

$$\mathfrak{X}^{\mathfrak{f}} = \phi^{\mathfrak{I}} \mathfrak{X}^{\mathfrak{f}-\mathfrak{I}} + \phi^{\mathfrak{J}} \mathfrak{X}^{\mathfrak{f}-\mathfrak{J}} + \cdots + \phi^{\mathfrak{b}} \mathfrak{X}^{\mathfrak{f}-\mathfrak{b}} + \epsilon^{\mathfrak{f}} + \theta^{\mathfrak{I}} \epsilon^{\mathfrak{f}-\mathfrak{I}} + \cdots + \theta^{\mathfrak{d}} \epsilon^{\mathfrak{f}-\mathfrak{d}}.$$

6.4机器学习算法

$$F(P'_{j+1}) = \sum_{v \in L(P'_{j+1})} \hat{p}_v F(\hat{q}_v) = \sum_{i=1}^{w_{j+1}} \hat{p}_i F_i,$$

$$F(P_{j+1}) = \sum_{i=1}^{w_{j+1}} \hat{p}_i F(\hat{q}_i).$$

$$F(P_{j+1}) - F(P'_{j+1}) = \sum_{i=1}^{w_{j+1}} \hat{p}_i (F(\hat{q}_i) - F_i).$$

$$\begin{aligned}
F(P_{j+1}) - F(P'_{j+1}) &\leq \frac{\gamma_{j+1}}{3} \sum_{i=2}^{w_{j+1}-1} \hat{p}_i \inf_{x \in I_i} F(x) + (\hat{p}_1 + \hat{p}_{w_{j+1}}) \frac{\gamma_{j+1}}{6} F(P_j) \\
&\leq \left(\frac{\gamma_{j+1}}{3} + \frac{\gamma_{j+1}}{6} \right) F(P_j).
\end{aligned}$$

$$\begin{aligned}
D_{P_0}(P'_0||P_1) &= \sum_x P_0(x) \log \frac{P'_0(x)}{P_1(x)} = \sum_{x^1,x^2,...,x^n} P_0(x^1,x^2,...,x^n) \sum_i \log \frac{P_0(x^i)}{P_1(x^i)} \\
&= \sum_i P_0(x^i) \log \frac{P_0(x^i)}{P_1(x^i)} = D(P'_0||P_1) = D(P_0||P_1) - D(P_0||P'_0)
\end{aligned}$$

$$P(z|u)=\frac{\sum_{(y,v_{u,y})}P(z|u,y,v_{u,y})}{\sum_{z'}\sum_{(y,v_{u,y})}P(z'|u,y,v_{u,y})}=\frac{\sum_{(y,v_{u,y})}P(z|u,y,v_{u,y})}{|\{(u,y,v_{u,y})\}|}.$$