**大数据时代下的社会信用**

熊晓红：尊敬的各位领导，各位嘉宾，大家下午好！我是鹏元征信的熊晓红，我很荣幸来参加今天这个主题分享，也非常感谢主办方的邀请。

我今天和大家分享的题目是“大数据时代下的社会信用”。大数据和信用这两个话题都是当今非常热门的话题，我更喜欢大数据和人工智能，现在大数据和人工智能也成为了国家战略。我接下来给大家分享的是鹏元公司基于大数据和大数据技术下的信用社会事件。

2012年，我们接到第一个关于互联网应用场景的大数据和大数据应用决策的需求。鹏元做征信服务是从2002年开始，当时的征信服务是传统行业下的征信服务，随着互联网时代的开启，在互联网应用场景的金融产品越来越多，互联网所带来的一些问题也引起了我们公司的大量资源的投入。经过多年的发展，我们在互联网应用场景方面、风控防控，已经研发了出了很多基于大数据和大数据技术，有关人工智能方面的信用决策产品。我认为，这些信用产品将会会我们互联网应用场景的信用决策变得更加智能、高效、更精准。

我今天给大家分享的主要有三个方面。一是普遍关注的问题，二是基于大数据的信用服务，三是信息主体的权益保护。

第一，大家对于社会信用普遍关注的一个问题，在互联网场景当中，主要有伪冒身份、虚假材料、信用欺诈、授信问题、客户的监控。对于这五个方面，鹏元征信这些年投入了大量的资源进行研究，包括在数据层面和技术层面上做了很多产品。在这些产品当中有两个核心，第一是帮助大家解决这些问题，第二是帮助大家提高决策效率。也就是说，从人工的决策转向机器决策。我们在选择信用产品的时候，非常关心信用产品的质量，从我们从业的角度来说，影响我们大数据信用产品的质量主要有三方面。第一方面是大数据的采集能力，包括采集数据的范围、内容和时效性。第二是大数据处理技术，你的处理能力，你是否能够从众多的结构化和非结构化的数据当中去提炼出来比较精准、高效的数据。第三是我们能否有对应用场景能力的把控，能够输出解决应用场景当中实际风险控制问题的产品。

我们经过这么多年，发现有一种技术，叫做语音识别，其实在我们互联网的应用场景当中非常多。鹏元征信从2012年就开始着手研究这方面了，当时我们有一家信用卡中心，提出在互联网上的场景出现了很多有关数据不规范人工识别等等方面的问题。我们通过这些年的研究发现，语音识别对于解决互联网应用场景当中实现人工转化成智能决策是有一定帮助的。它可以实现复杂信息之间的比对，回头我会有详细的介绍。二是智能搜索，三是仿人类记忆模糊的匹配。第四是信息服务，第五是敏感信息过滤。我们对于个人隐私保护也采用了这个技术。人机对话当中，海量数据处理的分类、舆情监控、三情分析。我们在这几个领域当中都有一些，在信用场景当中都在使用。

我们看一下语音识别的具体事例。第一是机构名称的搜索，大家都有印象，例如我们在互联网上去申请一旦业务要填写工作单位，需要填写非常全，有的非常智能友好，例如我输入鹏元征信，只要输入“鹏元”，征信就出来了，它把你最想用的机构名称选择出来了。第二是地址规范化处理，这方面在我们大数据处理过程当中，对于库里面同样一些不规范的地址或者无效地址进行剔除或者补全。第三是机构名称规范化，也是后台处理，例如在信用卡行业，信用卡使用者是大量的个人，个人的工作单位都存在库里面，很多工作单位名称是不规范化的，信用卡精准的风险管理是基于机构名称下进行管理。例如有一家企业已经倒闭了，我针对这家企业名下的持卡人做一个风险的谨慎决策。大家看一下事例，这是我们公司的前台，可以用手机扫一下，就把鹏元征信的名称扫进去。最重要的是后面出来的行业代码，这个代码怎么用，怎么出来的？在一些特定的行业当中，例如对企业的风险控制、对持卡人的风险把控，都要知道这个人所在的企业，这家企业所在的行业是否有系统腥风险。在现在很多很多数据获取渠道当中，企业行业是企业自己填写的，是不准确的，我们这个行业是怎么出来的？大家都知道，现在正常经营活动的企业都有一个官网，上面都有自己公司的介绍，并且还有从事产品业务的介绍。我们通过从这些文字描述当中体验出这个真实的企业究竟从事什么样的行业，这是非常有效的，并且一直在应用。这里的核心技术是语音识别。

前面我讲了一下大家比较关注的信用的问题，包括在信用的行业当中一些大数据的技术。第二，基于大数据的信用服务。因为今天时间有限，我只是针对互联网的一些场景，着重在企业环节当中去分享。我们在个人领域已经做了将近20年的业务，今天不在这里分享了

我们看一下电商分期购物的整体流程。在这个流程当中，我们哪些场景是需要信用去做支撑的？实名认证环节，然后填写资料，在填写资料的时候是否非常友好，不需要个人填写那么多文字，直接辅助录入即可。第三是自动化核实资料，个人填写的资料是否正确，工作单位、地址是否正确，别看这是两个简单的比对，因为它是机器，例如鹏元征信和鹏元征信有限公司是一家公司。第四是决策，自动化审批，其中有一些评估的模型，包括申请人的行为情况进行自动化的决策，可以完全取代人工作业。最后是贷后监控，那么多客户怎么去管理，我们鹏元提供贷后监控的产品，直接对你所关注的申请人或者客户进行实时监控，如果一旦有新的变化或者变动，我们会把这些信息及时传输给你。

在互联网，大家都谈到了互联网当中最主要是实名认证，现在实名认证非常多，包括我们公司做的私密性验证。每一种验证都有缺陷，通过率和精准率都不能完全解决问题，可能通过率达到80%，但是精准率就比较低了。有的精准率比较高，通过率比较低，也不能去使用。在这么多不太完善的实名认证的情况下，我们应该怎么去处理？鹏元征信有一套逻辑支持，通过不同的路径去完成实名认证。

我们现在看一下电商法人身份的核查。我们给一些电商已经提供了这个产品的服务，录入这些只是简单的演示，我们会对这个人的身份和行为进行验证，除此之外，还对这个人在企业的真实身份，是企业股东、法定代表还是职工进行验证。这是辅助录入，征信服务不仅仅是是在企业后台的工控决策当中，已经延伸到了前端，在客户填写资料的时候，我们已经对他信用情况进行把控。这是一个填写，把你最想要的机构选择出来，直接选择就可以了。在填写地址的时候，我们也是尽量让客户自己去选择，而非手工录入。例如打一个西直门，西直门大街所有的地址都出来了。如果不在表里面，我们会让他填写进去。如果这个库里的地址不存在，也有可能存在地址的欺诈。

除了信用决策的流程当中对信息的核查比对，我们基本采取机器比对，不需要人工去看，这不是一个非常简单的工作，在技术层面要求非常高。我这里举的例子比较简单，例如交通银行信用卡中心，我们写的是太平洋信用卡中心和交通银行信用卡中心是一个机构，要做到机器能够像人一样智能的识别。包括地址的识别，如果不在福田区，可能各位不知道阳光高尔夫大厦就是7008号，但有些直接填写了阳光大道7008号，机器要进行识别和学习，看它们是否一致。这就是机构名称的检索。

下面介绍一下关于企业征集到的相关信息，给大家做企业决策的时候使用。一个是起用负债情况，有清算、股权出资、股权冻结等五项信息。还有关于一个企业怎么能够多维度的把控它的风险，企业的信用的基本状况和企业股东或者被投资的企业，更深层次的去挖掘，才能更能认识或者了解和你做生意的这家机构。刚才说了是实时监控的产品，这个产品做得非常灵活，可以根据你的条件，还有选择你所要监控的事项，把监控的需求定下来，然后我们系统自动进行推送。

我们可以监控，大家用的比较多是信息变更的预计，重要的信息变更、不良信息的增加，还有关联的变动提醒。例如三家以上的贷款申请，有些机构需要推送。还有联系方式被标记成了黑名单，这在行业当中使用非常多，同业标记大片用户，发生了其他行为还有出现了法院诉讼、地址的异常变动。

我们还做了企业垂直搜索，主要是在关注企业在互联网上的正负面的新闻，我们可以做到15分钟推送。也就是说公司有一条新闻在网上出来了，我们在15分钟内反馈给你，关注公司实时舆情情况，或者自己的用户、供应商是否发生了重大的事件。

第二是从企业目标客户获客方面也进行了新奇推送。我们可以根据不同的条件进行检索，然后把你所需要的相关信息给你。这是企业的筛选。我上面讲的那些信息，包括产品，都是一小部分，如果大家感兴趣，我们可以接下来沟通。

第三是分享一下信息主体权益保护。大家都非常关注，包括我自己。我们常听到有人说，现在感觉到我自己在互联网上就是在裸奔了。从我们做征信来说，怎么去保护信息主体？在三个层面。包括采集、应用、安全保障、资源投入。采集和应用方面首先需要授权，不能侵权企业商业秘密。在自己内部，不管是做一个数据源单位还是征信机构、使用机构，一定要管控好自己的员工，不能把信息泄漏出去。

刚才有朋友问我，鹏元征信是一家什么样的架构？我简单说一下。鹏元征信做征信非常长，已经差不多16年，我们在数据资源、产品丰富都非常多，得到了行业的认可。我们是从1999年，深圳市开始征信系统建设，我们就开始着手。2002年成为了北航的征信系统之一。

这里值得说一下，我们在2009年的时候，我们的系统是国家信息化试点工程。2012年，我们系统在电子商务行业的应用被评为电商试点项目。这是我们公司在安全方面的投入，我们的系统是等保三级，是ISO20001的认证。到今天，我们服务了400多家银行和2900多家机构，包括京东也是我们的客户，互联网上很多出名机构都是我们的客户。年信用公报达到了7份。鹏元在政府和多个行业建立了征信合作圈，我们现在非常开放。

由于时间的关系，以上就是我的分享。我自己预测，将来人工智能会在信用决策这个行业发挥更大的作用，让我们的信用决策场景效率越来越高，并且还非常精准。

谢谢大家。